



ASHG 2022 Poster Abstracts

As of September 30, 2022

Instructions:

1. View the Table of Contents on the next page.
 - a. The Table of Contents is sorted chronologically by Poster Board Number.
2. Click on the Poster Board Number/Abstract Title of the abstract you would like to view and you will be taken to the page the abstract is on.
3. Details included with each abstract are: the poster session name, session date, abstract title, author block, and abstract body.

Table of Contents

PB1001. 3D genomics with Arima Hi-C sequencing enables detection of clinically relevant gene fusions in pediatric cancer samples.....	117
PB1002*. 4C guided cis- and trans-interaction networks associated with the RCCD1 gene at the 15p26.1 breast and ovarian cancer risk locus.....	118
PB1003. A B-ALL pediatric patient with IGH rearrangement	119
PB1004. A Case of A Hyperdiploid Complex Karyotype in a Patient with a Myelodysplastic Syndrome	120
PB1005. A cross-ancestry meta-analysis identifies 451 susceptibility loci for prostate cancer and yields a genetic risk score effective in multiple populations.....	121
PB1006. A genome-wide association study to identify predictive markers for the risk of nivolumab-induced immune-related adverse events	122
PB1007. A genome-wide gene expression signature of Breast cancer in the Moroccan women	123
PB1008. A novel immunohistochemical marker helps in the diagnosis of Lynch syndrome: detection of deficient DNA mismatch repair pathway proteins crypts within non-tumor colonic mucosa.	124
PB1009. A polygenic risk score for prostate cancer risk prediction.....	125
PB1010. A ten-year-long experience of <i>BAP1</i> -Tumor Predisposition Syndrome diagnostics	126
PB1011. Accelerated Tumor Only Variant Analysis Utilizing a GPU Framework.....	127
PB1012. AmpliconClassifier detects the mechanisms of focal genome amplifications in cancer.....	128
PB1013. Amplification of <i>RUNX1</i> in a patient with AML.....	129
PB1014. An active learning framework improves tumor variant interpretation.....	130
PB1015. An AML patient showing an abnormal hyperdiploid karyotype.	131
PB1016. An integrative genomic transcriptomic and proteomic investigation to characterize differences between isogenic radioresistant prostate cancer cell lines.	132
PB1017. Analysis of second-hit events in Tuberous Sclerosis Complex normal tissues and lesions...	133
PB1018. Analysis of the spatial transcriptomic profiles in tumor and tumor microenvironment of colorectal cancer in Alaska Native patients.	134
PB1019. ArCCH: Improving the Performance of Clonal Hematopoiesis Variant Calling and Interpretation using a Consensus Based Approach.....	135
PB1020. Artemis disease-associated variants and their effect on VDJ recombination and genome instability.....	136
PB1021. Assessing the relationship between coding variation and cancer risk in known cancer GWAS susceptibility regions	137
PB1022*. Assessment of genomic instability in aggressive forms of breast cancer in Mali.....	138
PB1023. Association of the rs12587 variant in the KRAS gene in patients with breast cancer and control group of the Mexican population.....	139

ASHG 2022 Annual Meeting Poster Abstracts

PB1024*. Associations of genetic ancestry to the somatic mutational landscape from tumor profiling data of 100,000 cancer patients.....	140
PB1025. B cell repertoire analysis in prostate cancer identifies gene expression pattern differences in FC gamma receptors.	141
PB1026. Breast cancer genetics in non-Ashkenazi breast cancer patients - carrier rates and variant classification in a genetically diverse population.....	142
PB1027*. Breast cancer polygenic risk score as a tool to risk stratify participants in the Healthy Nevada Project.	143
PB1028. Can impairment of kidney function cause kidney cancer? A bidirectional Mendelian randomization study.....	144
PB1030. Cancer subclone detection based on DNA copy number in single cell and spatial omic sequencing data.....	145
PB1031. Cancer-driving mutations are enriched in genic regions intolerant to germline variation.....	146
PB1032. Case report: Olaparib use in metastatic lung adenocarcinoma with <i>BRCA2</i> pathogenic variant	147
PB1033. Case-control likelihood ratio calculation for clinical classification of <i>BRCA1</i> and <i>BRCA2</i> variants of uncertain significance	148
PB1034. Cell-free DNA fragmentation patterns predict the response to immunotherapy in head and neck squamous cell carcinomas	149
PB1035. Cell-intrinsic metabolism and cofactor balancing define the metabolomic impact of NAD ⁺ supplementation in pancreatic cancer cell metabolism	150
PB1036. Characterization of molecular differences between normal weight and overweight/obese cancer patients.....	151
PB1037. Characterization of nucleosome positions and chromatin interactions of regulatory elements comparing Hi-C, Micro-C, and promoter capture Micro-C.....	152
PB1038. Characterization of the <i>2q22</i> renal cancer susceptibility locus	153
PB1039*. Chromatin accessibility of primary human cancers ties regional mutational processes and signatures with tissues of origin.....	154
PB1040. Chromosomal alteration profiling of lymph node metastases in non-small cell lung cancer	155
PB1041. Chromosome Arm Copy Number Profiling using a Clinical, Targeted NGS Panel.....	156
PB1042. Circulating miRNA signature for diagnostic prediction in breast cancer.	157
PB1043*. <i>Cis</i> - and <i>trans</i> -eQTL TWAS of breast and ovarian cancer identify more than 100 risk genes in the BCAC and OCAC consortia.	158
PB1044. Classification of Central Nervous System Tumors with DNA methylation - The Brazilian Experience.....	159
PB1045. Clinical and genetic features of patients with multiple endocrine neoplasia type 1 in Korea	160
PB1046. Clinical utility of genomic sequencing for hereditary cancer syndromes: A retrospective chart review.....	161

ASHG 2022 Annual Meeting Poster Abstracts

PB1047*. Clinicogenomic Neural Networks with Real World Evidence: Predicting Outcomes and Targeting Interventions at a Comprehensive Cancer Center.	162
PB1048. CML with a typical BCR-ABL1 fusion in a 15-year-old male patient presenting with marked leukocytosis, thrombocytosis, left-shifted granulocyte maturation in the peripheral blood	163
PB1049*. Co-delineating genomic and transcriptomic modes of resistance to MEK inhibition in individual triple negative breast cancer cells.	164
PB1050. Combined FISH and cytogenetic testing increases the accuracy of classification and risk stratification of multiple myeloma samples.	165
PB1051. Comprehensive Characterization of Functional Cancer Susceptibility at the 5p15.33 TERT/CLPTM1L Pan-Cancer Risk Locus	166
PB1052. Comprehensive CRISPR-based dissection of a complex common and familial melanoma risk locus on chromosome band 9p21.....	168
PB1053. Comprehensive study of gene expression outliers and their regulation mechanisms in pan-cancer.	169
PB1054. Computational method for the detection of microsatellite instability in tumor tissue samples	170
PB1055. Concurrent Core-binding factor beta subunit (CBFB) gene rearrangement/inv(16)(p13.1q22) in p210 BCR-ABL1 positive chronic myelogenous leukemia.....	171
PB1056. Confirmation of pathogenic germline variants identified by tumour testing in British Columbia.....	172
PB1057. Congenital uterine abnormalities and the risk of gynecological cancers	173
PB1058*. Copy number variation at <i>ATF7IP</i> is associated with increased risk for testicular germ cell tumors.	174
PB1059. Cost-utility of universal screening for common BRCA1 and BRCA2 variants among Ashkenazi Jewish women: a real-life analysis.....	175
PB1060. <i>CRLF2</i> Rearrangements in a Patient with B-ALL.	176
PB1061. Cross disorder genetic analysis identifies autoimmune disease loci inversely associated with diverse cancer types	177
PB1062. Deciphering a prognostic DNA methylation signature in tubo-ovarian high-grade serous cancer.	178
PB1063. Deep and error corrected sequencing via the low-cost Ultima Genomics platform enables ultra-sensitive circulating tumor DNA cancer monitoring.....	179
PB1064. Deep learning identifies gene expression and DNA methylation as highly predictive of lung cancer histologic subtype.....	180
PB1065. Deep learning morphology profiling identifies and enriches carcinoma cells from effusion samples in real-time for cytological and molecular analysis	181
PB1066. Detecting pleiotropic breast cancer susceptibility variants from genome-wide association studies.	182
PB1067. Detection of androgen receptor splice variants from clinical sequencing.....	183

ASHG 2022 Annual Meeting Poster Abstracts

PB1068. Detection of APC and MUTYH germline mutations in two south Indian families with hereditary colorectal cancer.	184
PB1069*. Detection of circulating tumor DNA in resectable pancreatic ductal adenocarcinoma	185
PB1070. Detection of <i>EGFR</i> exon skipping variants using clinical sequencing and its application....	186
PB1071. Detection of somatic copy number alterations using paired DNA and RNA sequencing.....	187
PB1072. Determining the mechanisms of radioresistance in breast cancer.....	188
PB1073. Developing a high-quality, sample-to-result Hereditary Breast and Ovarian Cancer Panel Assay Pipeline for using a novel sequencing platform.	189
PB1074. Development and Clinical Validation of a Bioinformatics Pipeline for Cancer Transcriptome Sequencing.....	190
PB1075*. Development and Clinical Validation of a Liquid Biopsy Assay for Neurofibromatosis 1.	191
PB1076. Development of a breast cancer risk prediction model with carrier status, a polygenic risk score, and epidemiologic risk score.	192
PB1077. DHGAN, Digital Histology Generative Adversarial Networks Used in Creating Synthetic Genitourinary Tissue Slides	193
PB1078. Diagnostic gene fusion detection in cancer by adaptation of short-read CHOP Fusion Panel to Oxford Nanopore long-read sequencing.	194
PB1079. Dietary factors and endometrial cancer risk: A Mendelian randomization study.....	195
PB1080*. Differentially expressed serum RNAs as potential early markers for primary sclerosing cholangitis-associated cholangiocarcinoma.	196
PB1081. Discovery of novel predisposing coding and noncoding variants in familial Hodgkin lymphoma	197
PB1082. Discovery of transcriptomic predictors of colorectal cancer mortality by race and ethnicity	198
PB1083. Dissecting admixture effects on a 313-variant polygenic risk score model for breast cancer in a census-based cohort of Brazilians.	199
PB1084. Does non-coding variation in known breast cancer susceptibility genes contribute to breast cancer risk?	200
PB1085. Dosing-specific modes of resistance to quizartinib in individual acute myeloid leukemia cells exposed by ResolveOME combined genomics and transcriptomics chemistry.....	201
PB1086. Dynamic play between human N a-acetyltransferase D and H4-mutant histones: Molecular dynamic study	202
PB1087. EagleC: A deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps.	203
PB1088. Efficacy of universal testing for hereditary cancer syndromes: A community based cancer center (CBCC) experience.	204
PB1089. EG VEGF in ovarian cancer and preeclampsia as an early diagnostic marker: toward a targeted therapy in the treatment of tumors, especially those resistant to chemotherapy.	205
PB1090. Eligibility, uptake and response to germline genetic testing in women with DCIS.....	206
PB1091. Elongin C (<i>ELOC/TCEB1</i>) gene: a novel cause of von Hippel-Lindau disease	207

ASHG 2022 Annual Meeting Poster Abstracts

PB1092*. Enhanced expression of mutant alleles of cancer driving genes in pan-adenocarcinoma....	208
PB1093. Estimating total tumor-specific microRNA content in human tissues using computational deconvolution.....	209
PB1094. Evaluating biomarker potential of germline genomic factors for predicting clinical outcomes in prostate cancer.	210
PB1095. Evaluation of miR-21 and miR-938 expression in the progression of gastric cancer in the Magellanic Chilean cohort - MAGIC.	211
PB1096. Evaluation of non-coding regulatory variants in susceptibility of Chordoma	212
PB1097. Evidences for polygenic inheritance in familial cancer	213
PB1098. Exome-wide discovery of the contribution of coding variants identifies novel genes associated with breast cancer risk.	214
PB1099. Expanding the clinical and molecular spectrum of hereditary <i>TINF2</i> cancer predisposition syndrome reporting a family with Hodgkin Lymphoma.....	216
PB1100. Exploring the non-invasive potential of miRNA 145 and miRNA 363 in Prostate Cancer...	217
PB1101. Family History-Based Risk of BRCA Mutations in the Malian Populations.....	218
PB1102. Filling gaps in whole genome analysis in hematology: A chance for optical mapping and long-read NGS.	219
PB1103. Finding Therapeutic Targets in Cancer Cells Using Essential Genes.....	220
PB1104. Frequently mutated TOP2B binding sites reveal candidate structural and regulatory cancer drivers of hepatocellular carcinoma.....	221
PB1105. Functional characterization of the 1p36.33 pancreatic cancer GWAS locus	222
PB1106*. Functional screens implicate candidate breast cancer risk genes in T cell-mediated immune-surveillance	223
PB1107. Functionalizing methylation-derived gene panels in the UK Biobank, TCGA, and HCA datasets.....	224
PB1108. Gene fusion detection and characterization in long-read cancer transcriptomes with FusionSeeker.....	225
PB1109. Gene-environment interaction of antioxidant genes and a healthy lifestyle index (HeLiX) on breast cancer risk reduction.....	226
PB1110. Genetic ancestry correlates with somatic mutations in endometrial cancer.....	227
PB1111*. Genetic ancestry differences in tumor mutation between early and average onset colorectal cancer	228
PB1112. Genetic Modifiers of <i>KRAS</i> -Mutant Colorectal Cancer.....	229
PB1113. Genetically Predicted Telomere Length and Risk of Multiple Primary Cancers in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.	230
PB1114. Genome and transcriptome landscape of familial pancreatic cancer and implications for altered glucose metabolism in moderate-penetrance cancer susceptibility.....	231
PB1115*. Genome-wide Analysis of Rare Haplotypes Associated with Breast Cancer Risk: Discovery, Replication, and Generalizability Evaluation	232

ASHG 2022 Annual Meeting Poster Abstracts

PB1116. Genome-wide association study identifies 4 novel risk loci including a missense variant in <i>LGR5</i> for small intestinal neuroendocrine tumors.....	233
PB1117*. Genome-wide association study of intracranial germ cell tumors: a common deletion at <i>BAK1</i> attenuates the enhancer activity and confers risk for the rare disease.	234
PB1118. Genome-wide profiling of DNA N6methylation from a breast cancer and a matched normal cell lines	236
PB1119*. Genomic landscape of malignant peripheral nerve sheath tumor (MPNST) reveals novel pathways of tumor evolution that correlate with tumor behavior.	237
PB1120. Germline functional non-coding variants associated with ovarian cancer predisposition	239
PB1121. Germline genetics of venous thromboembolism (VTE): Non-cancer polygenic risk scores improve risk stratification in cancer patients	240
PB1122. Germline sequencing in Brazilian pancreatic carcinoma patients.....	241
PB1123. Germline sequencing of DNA damage repair genes in two hereditary prostate cancer cohorts reveals rare risk-associated variants.....	242
PB1124. Germline variants in the Ah Receptor in Colombian smoker patients with squamous cell carcinoma of the head and neck.....	243
PB1125. GLUT10 is a biomarker for poor prognosis and a predictor of vitamin C combination therapy in breast cancer	244
PB1126. Going beyond the atlas: Mapping the tumor landscape with complete genomic signatures..	245
PB1127*. HDM2 differentially modulates the molecular phenotype of lung tumor cells of varying p53 genotypes	246
PB1128. Hemophagocytosis related to acute promyelocytic leukemia is also positive for the <i>PML::RARA</i> gene fusion.....	247
PB1129. Hereditary Cancer Variants & Homologous Recombination Repair Deficiency in Biliary Tract Cancer	248
PB1130*. Heterozygous loss of <i>AURKB</i> in the germline associates with protection from metastasis and death from cancer.....	249
PB1131. High tumor mutational burden by whole-genome sequencing in resected non-small cell lung cancer in never smokers is associated with worse prognosis.....	250
PB1132*. High-throughput functional screening of candidate causal variants from lung cancer susceptibility loci	251
PB1133. Identification of key molecular mechanisms in IDH mutant brain tumors to enable precise risk stratification	252
PB1134. Identification of target genes and pathways involved in triple-negative breast cancer proliferation and stemness using genome-wide CRISPR screens.....	253
PB1135. Identification of <i>USP9X</i> as a leukemia susceptibility gene.....	254
PB1136*. Identifying molecular markers of early stage ovarian cancer pathogenesis through whole genome CRISPR screening of mutant <i>BRCA1</i> fallopian tube cells.	255

ASHG 2022 Annual Meeting Poster Abstracts

PB1137. Identifying strong instrumental variables for pleiotropic assessment in Mendelian randomization: CALGB/SWOG 80405 (Alliance)..... 256

PB1138. Impact of polygenic risk, pathogenic variants, and family history on detection of breast cancers within and outside the national mammography screening programme on 117,252 women over 1992-2019 258

PB1139. Implementation of new ClinGen/CGC/VICC recommendations for classification of oncogenicity of somatic variants using AI..... 259

PB1140. Implementation, utilization, and diagnostic outcomes of a technology-enabled telehealth-based program for individuals to access multi-cancer early detection testing 260

PB1141. Inference on the genetic architecture of breast cancer risk 261

PB1142*. Integrating GWAS and 3D chromatin interactome data to identify multi-cancer risk genes in hormone-related cancers. 262

PB1143. Integrative data analysis to uncover genes and pathways underlying the aggressiveness of lung cancer brain metastasis and the overall prognosis 263

PB1144. Investigating the link between cholesterol-lowering genes and risk of skin cancers 264

PB1145. Isolated multinodular goiter in a family with DICER1 heterozygous germline variant 265

PB1146. Large-scale whole genome sequencing reveals that APOBEC mutagenesis is a common process in normal human small intestine 266

PB1147. Leveraging network architecture and transfer learning for precision oncology..... 268

PB1148. Leveraging somatic cancer mutation data to predict the pathogenicity of germline missense variants..... 269

PB1149. Liquid biopsy for the detection of actionable human variants in the blood of cancer-diagnosed dogs: Opening the door to comparative oncology and therapeutic development. 270

PB1150. Loss of function of *neurofibromin* affects *afadin* -6 in neurofibromatosis type-1..... 271

PB1151. Low CD2 expression in sentinel lymph nodes of early breast cancer patients is associated with postoperative tumor recurrence or metastasis 272

PB1152. Massively parallel sequencing-based panel strategy for microsatellite instability testing..... 273

PB1153. Mechanistic insight into HIF1 α -Notch1 axis in chronic myeloid leukemia patients and its diagnostic implications 274

PB1154. Metabolic Profiling of Primary and Metastatic Breast Cancer Cells and New Treatment Approaches. 275

PB1155. Metagenomics and metabolomics of persistent high-risk Human Papillomavirus (hrHPV) infections and cervical cancer 276

PB1156*. Methylation-sensitive restriction enzyme sequencing (MRE-seq)-based early detection of cancer using deep neural network..... 277

PB1157. Mitonuclear genotype remodels the metabolic and microenvironmental landscape of Hürthle cell carcinoma 278

PB1158. Modulating the immune system not to reject an allograft..... 279

ASHG 2022 Annual Meeting Poster Abstracts

PB1160. Molecular signature predicts immunotherapy resistance in gynecologic cancer patients treated with immune checkpoint inhibition.	280
PB1161. Mosaic chromosomal alterations detected in HIV infected men and its relationship to non-Hodgkin lymphoma.	281
PB1162. Multigene deregulation 12q12 renal cancer-susceptibility locus.	282
PB1163. Multi-modal characterization of ultra-rare germline genetic variants driving breast cancer risk in the indigenous Arab population.	283
PB1164. Multiomic sequencing analysis with whole genome association study at chromosome 16 of wilms tumor patients.	284
PB1165. Multi-omic single-cell profiling of endogenous and engineered T cells in patients undergoing CAR T cell therapy for high-grade glioma.	285
PB1166. Multiplex ligation dependent probe amplification versus fluorescent in situ hybridization for screening RB1 copy number variations in Egyptian patients with retinoblastoma.	286
PB1167. Multi-population GWAS analysis identified functional variants in ever- and never-smoking lung cancer.	287
PB1168*. Multi-stage germline exome sequencing study of 17,546 men with aggressive or non-aggressive prostate cancer identifies genes that may inform clinical gene panel testing.	288
PB1169. Next Generation SP DNA Sample Prep and DLS Labeling Readies Optical Genome Mapping Workflows for Adoption at Scale.	290
PB1170. No evidence of association between clonal hematopoiesis and risk of prostate cancer in large samples of European ancestry men.	291
PB1171. Novel Germline Pathogenic Variants of <i>APC</i> and <i>BMPRIA</i> genes in Algerian patients with Hereditary Polyposis Syndromes.	292
PB1172. Nucleotide resolution of large tandem duplications in cancer genomes.	293
PB1173. OM2BFB: Detecting and elucidating Breakage Fusion Bridge structures in cancer genomes using Optical Mapping data.	294
PB1174. Oncological biomarkers identified by NGS: characterization of the colombian population tumors from 2019 to 2022.	295
PB1175. Online implementation of combined polygenetic protein prediction algorithm for oncology (C3PO): Tumor mutation contributions to cancer hallmarks based on proteomics.	296
PB1176. Ovarian sex cord stromal tumor is a likely additional phenotype of <i>BAP1</i> -Tumor Predisposition Syndrome.	297
PB1177. <i>PALB2</i> -mutated human mammary cells display a broad spectrum of morphological and functional abnormalities beyond defective DNA damage response.	298
PB1178. Pan-Cancer Analysis introduces HLF as a tumor suppressor gene and a promising biomarker for prognosis and targeted therapy in human tumors.	299
PB1179. Pan-cancer polygenic risk distinctly contributes to prediction of heterogeneous subsequent malignancies across treatment profiles in survivors of childhood cancer.	300
PB1180. Parent-of-origin-aware genomic analysis in the management of patients and families with genetic disease.	301

ASHG 2022 Annual Meeting Poster Abstracts

PB1181. Pathogenic genetic variants from highly connected cancer susceptibility genes confer the loss of structural stability and further link with diabetes	302
PB1182. Pathogenic Variants in Adult-Onset Cancer Predisposition Genes Identified by Pediatric Tumor Genomic Testing: Prevalence, Tumor Features, and Germline Testing Outcomes	303
PB1183. Phylogeny-aware detection of single-nucleotide variants and mode of evolution from cancer single-cell DNA sequencing data.....	304
PB1184. Potential misrepresentation of inherited breast cancer risk by common germline alleles.....	305
PB1185. Potential subtype-informative genetic risk variants and genes identified from a case-case analysis for breast cancer	306
PB1187. Prevalence of mutations in hereditary cancers of Czech patients.	307
PB1188*. Prevalence of suspected pathogenic or likely pathogenic germline <i>TP53</i> variants in population sequencing databases: A genotype-first analysis.	308
PB1189. Primary myelofibrosis and progression to acute myeloid leukemia in a young patient with variant in <i>CARL</i>	309
PB1190. Prognostic Value of Bile Acid Transporter SLC10A1 Expression in Hepatocellular Carcinoma.....	310
PB1191*. Proteome studies of breast, prostate, ovarian, and endometrial cancers implicate plasma protein regulation in cancer susceptibility.	311
PB1192. Rare germline variants in Deubiquitylating enzymes (DUBs) in young women with breast cancer	312
PB1193. Rare variant association study in lymphoid cancers	313
PB1194. Relative involvement of human papillomavirus constitutional genomic instability: Cellular sequela syndromes.	314
PB1195. Resolving clone-and haplotype-specific copy number variation and DNA methylation in heterogeneous tumors with nanopore sequencing.....	315
PB1196. Revisiting tumor evolution in multifocal hepatocellular carcinoma with different clonal origins	316
PB1197. Rewired phospho-signaling networks by genetic variants in Clear Cell Renal Cell Carcinoma	317
PB1198. Risk of cancer in persons with a single Fanconi anemia gene variant: A population-based study of MyCode Cohort	318
PB1199. Serum biomarkers are altered in UK Biobank participants with mosaic chromosomal alterations.....	319
PB1200. Sex differences in the Lung Adenocarcinoma transcriptome: Analysis of TCGA LUAD and GTEx lung tissue.....	320
PB1201. Shared predisposition to cancer and metabolic disorders using large-scale genomic data	321
PB1202. Single cell RNA sequencing reveals global cellular landscape of appendiceal cancer.....	322
PB1203. Single cell transcriptomics of Pituitary Neuroendocrine Tumors (PitNETs).....	323

ASHG 2022 Annual Meeting Poster Abstracts

PB1204. Single Sample Gene Set Enrichment Analysis (ssGSEA) to Identify Dysregulated Pathways in a Biomarker Matched Pilot Study of Mantle Cell Lymphoma. 324

PB1205. Single-cell identity and state coupled with genome-wide SNV and CNV in primary breast cancer with ResolveOME profiling. 325

PB1206. Single-cell methylation and 3C sequencing in prostate tumors shows substantial epigenome reprogramming..... 326

PB1207. Single-cell multi-omic characterization of MPAL phenotypes reveals patient heterogeneity associated with poor treatment outcomes 327

PB1208. Single-cell sequencing of genomic DNA: Assessing clonal populations of tumor cells in endometrial cancer 328

PB1209. *SMARCA4* Schwannomatosis: a known cancer susceptibility gene, a new phenotype..... 329

PB1210. Splicing analysis of endometrial cancer GWAS risk loci: splicing-associated variants reveal *NFI* and *SKAPI* as candidate susceptibility genes. 330

PB1211. t(6;9)(p22;q34) translocation with DEK-NUP214 fusion identified in one patient with chronic myeloid leukemia..... 331

PB1212. Targeting mutant HLA-bound cancer peptide demonstrated in vivo efficacy in delaying tumour growth in a low mutational burden aggressive ovarian cancer..... 332

PB1213. Telomeric nuclear organization of urinary circulating cells in bladder cancer (BC): A new method uncovers unreported markers. 333

PB1215. The effect of weight-loss on the colorectal transcriptome and its relation to colorectal cancer risk. 334

PB1217. The role of senescent tumor cells in cancer progression..... 336

PB1218. The Role of Tumor Microenvironment and Tumor-Active Pathways in the Progression of Glioma 337

PB1219. The use of *TNF*-*α* polymorphic forms as a possible risk assessment tool for the management of breast cancer among Nigerian women. 338

PB1220. Transcriptome Wide Association Study identifies novel candidate risk genes for testicular germ cell tumors 339

PB1221. Transcriptome-wide association study identifies new breast cancer susceptibility genes in Latinas..... 341

PB1222*. Transcriptome-wide association study identifies novel susceptibility genes for childhood acute lymphoblastic leukemia in Latinos..... 342

PB1223*. Uncovering gene fusions with 3D genomics: from clinical validation to actionable insights for undiagnosable solid tumors 343

PB1224*. UNISON: a software toolkit for streamlined detection of clonal hematopoiesis of indeterminate potential in large-scale sequencing studies 344

PB1225. Unraveling the mechanisms behind the malignant behavior of mast cell tumors using laser capture microdissection mediated RNA sequencing on canine models..... 345

PB1226. Unravelling RecQ helicase function in genome stability using Strand-seq 346

ASHG 2022 Annual Meeting Poster Abstracts

PB1227. Using human genetics to evaluate the causal role of circulating inflammatory markers in risk of adult cancer..... 347

PB1228. Utility of hereditary cancer panel testing in individuals with a known familial variant. 348

PB1229*. Utilizing Electronic Health Records (EHR) and Tumor Panel Sequencing to Demystify Prognosis of Cancer of Unknown Primary (CUP) patients 349

PB1230. Variants in *FANCM* confer risk to estrogen receptor-negative breast cancer among Latinas 350

PB1231. Whole genome CpG-resolution DNA methylation profiling of HNSCC reveals previously ignored heterogeneity among HPV(+)tumors and clinical implications..... 351

PB1232. A catalog of genetic predictors impacting lifelong medication use patterns in hyperlipidemia, hypertension and type 2 diabetes 352

PB1233. A comparative PRS-PheWA of 14 autoimmune disorders, and their genetic relationship.... 353

PB1234. A comprehensive coronary artery disease risk prediction framework incorporating genetic and nongenetic risk factors. 354

PB1235*. A comprehensive genetic map of cytokine responses to pathogens in 1063 Lyme patients reveals novel regulatory mechanisms underlying diseases. 355

PB1236. A *Drosophila* platform to validate and assess pathogenicity of LMNA variants identified from dilated cardiomyopathy patients. 356

PB1237. A genome wide association study of chronic spontaneous urticaria risk and heterogeneity. 357

PB1238. A genome-wide association study in Peruvians suggests new risk loci for Alzheimer disease 358

PB1239. A genome-wide association study of lifetime estrogen exposure in Korean postmenopausal women..... 359

PB1240. A genome-wide association study of mammographic density phenotypes among pre-menopausal women of European ancestry..... 360

PB1241. A Haptoglobin (HP) Exon Deletion Polymorphism Alters the Effect of *APOE* Alleles on Alzheimer’s Disease in Individuals of European Ancestry with *APOEε4*..... 362

PB1242. A large-scale genome-wide association meta-analysis of polycystic ovary syndrome in over 540,000 women provides insights into biological mechanisms..... 363

PB1243. A Mendelian randomization study of genetic liability to Post-traumatic stress disorder and risk of Ischemic Stroke. 364

PB1244. A multi-ancestry genome-wide association study identifies two novel loci associated with increased risk of diabetic macular edema. 365

PB1245. A multi-ethnic genome-wide association study in type 1 diabetes. 366

PB1246. A nationwide approach to understand the role of health, socioeconomic and genetic information in predicting COVID-19 vaccination uptake 367

PB1247. A novel missense variant in melanopsin associated with delayed sleep phenotype: A whole genome sequencing study. 368

PB1248. A pan cancer immunogenomic atlas and its applications for immune checkpoint blockade immunotherapy 369

ASHG 2022 Annual Meeting Poster Abstracts

PB1249. A polygenic score-based approach improves discrimination between familial, idiopathic, and pathologic short stature in a cohort of pediatric patients presenting for evaluation.....	370
PB1250. A second wave SARS-CoV-2 variant in Quebec is associated with persistent infection.	371
PB1251. A single nucleus transcriptomics study of alcohol use disorder in postmortem brain tissue.	372
PB1252. A Systematic Multi-omics Approach Identifies miRNA and Proteomic Dysregulation in Human Post-traumatic Stress Disorder	373
PB1253. A trans ancestry genomics based approach to study the interplay between the immune system, infectious type, and HLA type, ancestry and sepsis outcome.....	374
PB1254. Accurate identification of causal variants of intracellular LDL uptake via Bayesian modeling of base editor reporter screens	375
PB1255. Activity-dependent transcriptional program in glutamatergic neurons enriched for genetic risk for schizophrenia.....	376
PB1256. Advancing rare disease research through web-based recruitment: the 23andMe Systemic Sclerosis Research Study	377
PB1257. Age differences in preadipocyte proportions and expression profiles link to obesity and insulin resistance.....	378
PB1258. Altered expression levels of <i>TASIR1</i> , <i>TASIR3</i> genes among SARS-CoV-2 variants	379
PB1259. Alzheimer Disease plasma biomarker pTau-181 in individuals of diverse admixed ancestral backgrounds	380
PB1260. Alzheimer's disease genetic risk and plasma biomarker profiles in the Midwestern Amish.	381
PB1261*. Alzheimer's disease pathology drives distinct homeostatic and active microglia phenotypes revealing clues to early pathogenic transcriptional switches	382
PB1262. An emirati cohort of autism spectrum disorders.	383
PB1263. An integrated genomic and phenotypic analysis of sleep traits in autism spectrum disorder	384
PB1264. An unprecedented level of complexity in the schizophrenia-associated 3q29 region of the human genome with unique segments that increase the risk for non-allelic homologous recombination.	385
PB1265. Analysis of 3,273 undiagnosed neurodevelopmental disorder trios reveals over-transmission of polygenic risk and a female protective effect	386
PB1266. Analysis of adipose and liver expression profiles from dual-tissue transcriptomic cohort discovers 10 serum biomarker candidates for NAFLD.....	387
PB1267. Analysis of both shared and specific associations among various autoimmune disorders greatly helps the identification of the causal gene(s) and the functional mechanisms.....	389
PB1268. Analysis of MRI-derived spleen iron in the UK Biobank identifies genetic variation linked to iron homeostasis and hemolysis.....	390
PB1269. Analysis of nascent RNA reveals distinct gene signatures for acute inflammation in cytokine stimulated keratinocytes.....	391
PB1270. Analysis of poison exons caused by splice donor variants in genes associated with developmental brain disorders	392

ASHG 2022 Annual Meeting Poster Abstracts

PB1271. Analysis of Relationship between Metabolic Syndrome and Genome-Wide Association Study (GWAS) of Vitamin D in Koreans: Community-Based Cohort.	393
PB1272. <i>APOE</i> genotype and rare variants of APP processing complex in autopsy-confirmed rapidly progressive Alzheimer disease.	394
PB1273*. Assessing T-Cell Receptor Diversity and Genetic Variation in Complex Phenotypes.	395
PB1274. Association analysis of <i>LHCGR</i> gene polymorphisms with Polycystic Ovary Syndrome in Punjabi population.	396
PB1275. Association analysis of structural variants in whole-genome sequencing of 150K UK Biobank participants.	397
PB1276. Association between a type 2 diabetes polygenic risk score and type 2 diabetes-related risk factors and complications.	398
PB1277. Association between Blood Pressure-related Polygenic Risk Scores and Hypertension among White and Black Individuals Across the Life Course.	399
PB1278. Association between genomic runs of homozygosity and complex traits in the Taiwan Biobank.	400
PB1279. Association between polymorphisms in <i>CD32</i> , <i>CD36</i> , <i>CD40</i> , <i>CD54</i> and Malaria phenotypes among under-fives in Ibadan South-west Nigeria.	401
PB1280. Association of a genetic risk score with incident coronary heart disease in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) cohort.	402
PB1281. Association of African Ancestry with risk for fibroproliferative diseases is consistent with selection for a Th2 favored genome in African derived populations.	403
PB1282. Association of genetic predisposition and physical activity with risk of gestational diabetes mellitus in nulliparous women.	405
PB1283. Association of Genetically-Predicted Placental Gene Expression and Diseases of the Cardiovascular System.	406
PB1285. Association of Structural Variants with Coronary Artery Disease.	407
PB1286. Associations between polygenic risk scores and gene expression identify core genes for bipolar disorder.	408
PB1287. Associations between polygenic risk scores, alcohol use, and differences in brain measures in the UK Biobank.	409
PB1288. Autism and Cognitive Ability CNV- Genome Wide Association study (CNV-GWAS).	410
PB1289. Autism Spectrum Disorder (ASD) in Qatar: Whole genome sequencing of 100 affected families highlights Dominant and Recessive risk genes of ASD.	411
PB1290. Automated identification, GWAS, and ExWAS of enlarged perivascular space burden in human brain MRI in the UK Biobank.	412
PB1291. Burden of functional variants in epilepsy patients using a deep learning approach.	413
PB1292. Causal effects of gut-related metabolites on human psychiatric disorders.	414
PB1293. Cell-type-specific activity-dependent multiomic profiling in single hiPSC-derived neurons for neuropsychiatric disorders.	415

ASHG 2022 Annual Meeting Poster Abstracts

PB1294*. Cellular context-specific gene regulation of neuropsychiatric disorders in single human neurons.....	416
PB1295. Characterization of a diverse Frontotemporal Dementia cohort, enriched for Caribbean Hispanic patients.....	417
PB1296. Characterization of gene expression profile and molecular pathways involved in type 2 diabetes among self-identified Hispanic American individuals.....	418
PB1297. Characterization of sequencing-based HLA alleles in a Quebec COVID-19 biobank	419
PB1298. Characterizing 31 cardiometabolic candidate genes for a role in early vascular inflammation and foam cell formation using CRISPR-Cas9, <i>in vivo</i> imaging and deep learning	420
PB1299. Cholesterol and LDL-C levels are elevated prior to symptom onset in Alzheimer’s disease independent of APOE-4.....	421
PB1300. Choroidal thickness as a biomarker for age-related macular degeneration progression: Genome-wide Association Study in Amish	422
PB1301. Classical HLA Alleles are Associated with Fasting Glucose and Type 2 Diabetes in Multiple Populations.....	423
PB1302. CLEC16A regulates the adipogenic and lipolytic capacity of adipocytes	424
PB1303*. Clinical utility of polygenic scores for cardiometabolic disease in Arabs	425
PB1304. Clonal behavior across multiple timepoints in clonal hematopoiesis	426
PB1305. Collagen gene cluster expressions role on liver fibrosis in biliary atresia patients.....	427
PB1306. Colocalization and variant-to-gene mapping nominates pleiotropic bone mineral density effector genes	428
PB1307. Combining genetic risk with machine learning derived biological age improves prediction of type 2 diabetes.....	429
PB1308. Combining polygenic and monogenic causes for improved osteoporotic fracture risk prediction.	430
PB1309. Common and rare variant associations with spinal stenosis in 29,488 cases.....	431
PB1310. Comparison of clinical characteristics among survivor and non-survivor of covid-19 infected patients: Scoping review and meta-analysis.	432
PB1311. Concordance of directional effects between sexes suggests pervasive epistasis and omnigenic regulation of complex metabolic traits in mice.....	433
PB1312. Considerations in developing a framework for defining clinical actionability of polygenic risk scores	434
PB1313. Construction of clinically significant mutation map and their spatio-temporal single cell transcriptomics trajectories in congenital heart disease	435
PB1314. Construction of mutational landscape and characterization of associated genes in a Bangladeshi cohort of neurodevelopmental disorders	436
PB1315. Copy number variants and mental health in a community pediatric sample.....	438
PB1316. Coronary artery disease risk assessment in Asian Indian Punjabis using a genome wide polygenic risk score	439

ASHG 2022 Annual Meeting Poster Abstracts

PB1317. Cross-tissue single-nucleus RNA-sequencing reveals distinct tissue-specific expression profiles of adipocytes, depending on their residency in human adipose tissue or heart	440
PB1318. Curation burden mitigation: strategies for empowering high throughput gene panel analysis using whole genome sequencing.....	441
PB1319*. Cytokine treatment increases <i>APOEε4</i> expression in reactive astrocytes in European local ancestry but not in African local ancestry.....	442
PB1320. Data-driven Clustering of Human Cardiovascular Disease (CVD) Related Variables	443
PB1321. Data-driven genome-wide association study of longitudinal pain and mobility trajectories among cases of Parkinson’s disease from the Fox Insight Data Exploration Network	444
PB1322. <i>De novo</i> CNVs impacting constraint overlapping genes in neurodevelopmental disorders and congenital anomalies.....	445
PB1323. <i>De novo</i> germline and somatic variants as modifiers of cleft lip/palate severity in mice	446
PB1324. <i>De novo</i> variant analysis of childhood-onset obsessive-compulsive disorder in the French-Canadian population.	447
PB1325. Deciphering the causal relationship between blood pressure and white matter integrity: a Mendelian Randomization study in the UK Biobank	448
PB1326. Deep embedded clustering by phenotype and genome-wide association study in autism from SPARK individuals.....	449
PB1327. Deep learning models of gene regulation in Alzheimer’s disease improve functional fine-mapping and thereby cross-ancestry PRS portability	450
PB1328. Deep venous thrombosis: genotyping and whole-exome sequencing data to improve genetic risk prediction	451
PB1329. Differential gene expression analysis of alcohol use disorder across postmortem human brain tissues.....	452
PB1330. Differential splicing occurs in individuals with disseminated coccidioidomycosis infections.	453
PB1331. Differentiating Mesenchymal Stem Cells from iPSCs for Analysis of Type 2 Diabetes and Related Trait GWAS Loci	454
PB1332. Direct participation of individuals scales the development of cardiometabolic polygenic risk score models.....	455
PB1333. Discovering disease context eQTLs in a patient cohort with active psoriatic arthritis	456
PB1334*. Discovering the early molecular markers of lipid dysregulation using deep phenotyping and adipose tissue gene expression.....	457
PB1335. Discovery of rare variants in 219 genes associated with adult human height via burden analysis in exomes of 785,210 individuals	458
PB1336. Discovery of Type 2 Diabetes genes using an accessible tissue.	459
PB1337. Discrepancy between genetically-predicted and actual body mass index is a significant predictor of incident cardiovascular disease	460

ASHG 2022 Annual Meeting Poster Abstracts

PB1338. Dissecting autism heterogeneity by genotype phenotype analysis among autism risk gene carriers in SPARK.....	461
PB1339. Dissecting single cell splicing isoforms in human hearts using single cell nanopore sequencing.....	462
PB1340. Dissecting the relationship between recurrent pregnancy loss and type 1 and type 2 diabetes in Western Ukrainian population	463
PB1341. Dissecting the role of pleiotropy in the genetic adaptation of Tibetans to high altitudes.	464
PB1342. Dose-responsive mRNA biomarkers of alcohol consumption and placebo response: A transcriptome-wide gene expression analysis.	465
PB1343. Drug repurposing for osteoporosis in men by integrating multi-omics and pharmacogenomics	466
PB1344. Effect of artificial light on metabolic disorders and circadian gene expression.....	467
PB1345*. Effect of cortisol on cortical organoids: Building a "stress in a dish" model system.....	468
PB1346. Effects of epigenetic age acceleration on kidney function: a mendelian randomization study.	469
PB1347. Effects of the Tau Haplotype on Cognition in Childhood	470
PB1348. Elevated levels of CXCL16 in severe COVID-19 patients: Effects on mortality.....	471
PB1349*. Enriching genetic hypotheses of schizophrenia through neuroimaging transcriptomics.	472
PB1350. Enrichment of Patients with Ehlers Danlos Syndrome in Idiopathic Gastroparesis - A Gene Set Enrichment Analysis.....	473
PB1351. Epstein-Barr virus- and genotype- dependent transcriptional regulation in B cells from patients with multiple sclerosis.....	474
PB1352. ERAP1, ERAP2, and two copies of HLA-Aw19 alleles increase the risk for Birdshot Chorioretinopathy in HLA-A29 carriers.....	475
PB1353. Essential hypertension genomic regions from linkage analysis studies.....	476
PB1354*. Estimating the causal influence of body mass index on gut microbiome variation	477
PB1355. Ethnicity- and sex-specific genome-wide association study on Parkinson's disease	478
PB1356. Evaluating the causal effect of tobacco smoking on whitematter brain aging: a Mendelian randomization analysis in UK Biobank	479
PB1357. Evaluating the frequency and impact of structural variation in amyotrophic lateral sclerosis.	480
PB1358. Evaluation of genetic support for targeting the IL-23 pathway: current and future therapeutic opportunities	481
PB1359. Evaluation of mitochondrial DNA variation in autism and neurodevelopmental disorders ..	482
PB1360. Evaluation of polygenic risk scores to differentiate between type 1 and type 2 diabetes.	483
PB1361. Evidence of allelic series with fine-mapped protein quantitative trait loci across 1,470 protein abundance measurements and 30 biochemistry measurements in UK Biobank participants.	484

ASHG 2022 Annual Meeting Poster Abstracts

PB1362. Examination of PTSD symptoms in older veterans as a function of AD risk genes and combat implicates <i>CLU</i> as a stress-response gene.....	485
PB1363. Exome analysis reveals plausible glaucoma causing mutations in novel genes in dominant JOAG families of Pakistani origin	486
PB1364. Exome sequencing of six singlet Qatar families identified nine neurodevelopmental candidate genes	487
PB1365. Exome-wide association study of DSM-5 antisocial personality disorder in a nationally representative sample.....	488
PB1366. Exploring disease-course trajectories using genomic scores in the Finnish SUPER study - a cohort of 10,407 psychotic individuals.	489
PB1367. Exploring genome-wide association results for neuroticism as a proxy for mental health disorders in individuals of African ancestry	490
PB1368. Exploring the complexity of scleroderma etiology by trio whole genome sequencing	491
PB1369. Expression of adipokine genes associated with gestational diabetes mellitus	492
PB1370. Expression profiling of small bowel tissue from Crohn's disease patients reveals alterations of immune functions and dampening of epithelial response.	493
PB1371. Extensive Transcriptional Complexity of Autism Associated <i>SHANK</i> Family Genes by Capture Based and Long Read sequencing Method	494
PB1372. Findings at both known and novel loci in genome wide linkage and association study of a depression endophenotype, major depressive disorder with chronic pain.	495
PB1373. Fine-mapping 110 migraine risk loci using 98,375 migraine cases	496
PB1374. Fine-mapping 56 lipidomics loci identified by univariate and multivariate genome-wide association analyses	497
PB1375. Fine-mapping and genomic analyses identify causal variants and genes for hypertension ...	498
PB1376. Fine-mapping and signal co-localization of asthma and white blood cell traits within the chr17q12-21 locus across diverse ancestries.....	499
PB1377*. Fine-mapping the association of <i>CYP2A6</i> with nicotine metabolism in African ancestry smokers.	500
PB1378. Fine-mapping to find differences in effect sizes across ancestries for skeletal phenotypes...	501
PB1379*. Functional analysis of rare copy number variations across psychiatric disorders.....	502
PB1380. Functional validation of rs7132908 as the causal variant at the childhood obesity locus on chr12q13.	503
PB1381. Gene-based association studies of rare variants identify novel genes for asthma-associated quantitative traits in African-American children.	504
PB1382. Gene-based tests for early and late onset Alzheimer Disease: common and non-overlapping factors.....	505
PB1383. Gene-by-stress interaction in Alzheimer's disease risks.	507
PB1384. Gene-environment interaction of coffee with body mass index in multiple populations.....	508

ASHG 2022 Annual Meeting Poster Abstracts

PB1385. Gene-level GWAS fine-mapping combined with single cell RNA-seq yields insight into neurodegenerative disease biology	509
PB1386. Genes at asthma-associated GWAS loci are induced in activated CD4 tissue resident memory T cell	510
PB1387. Genes causing significant effects on the dynamics of peroxisomes	511
PB1388. Genetic analysis of bruxism and its associations to sleep, psychiatric and behavioural traits.	512
PB1389. Genetic analysis using large biobank controls validates known genes and discover novel genes in Hirschsprung disease.	513
PB1390. Genetic and environmental regulation of caudate nucleus transcriptome in schizophrenia ..	514
PB1391. Genetic architecture of asthma in African American Patients	515
PB1392*. Genetic architecture of insulin-stimulated glucose uptake reveals novel regulators of post-prandial glucose metabolism.....	516
PB1393. Genetic associations of rare variants in alcohol and tobacco use in up-to 526,400 individuals	517
PB1394. Genetic characterization of central serous chorioretinopathy and pleiotropic effects with age-related macular degeneration	518
PB1395. Genetic colocalization between Covid-19 and diseases of the immune system.	519
PB1396. Genetic epidemiology of carotid intima-media thickness in Sub-Saharan African populations contributes to atherosclerosis biology.....	520
PB1397. Genetic Exploration of Military Sudden Cardiac Arrest (GEMini): Genome Sequencing Results from the First 24 Participants.	521
PB1398. Genetic Interactions Implicate Disruption in Ciliogenesis in the Etiology of Non-syndromic Orofacial Clefts.....	522
PB1399. Genetic Mechanisms of Supraventricular Tachycardias	523
PB1400. Genetic Overlap among Alzheimer's Disease and Five Psychiatric Disorders: Genetic Convergence and Specificity.	525
PB1401. Genetic overlap of idiopathic pulmonary fibrosis and hypertension.	526
PB1402. Genetic pleiotropy effects on kidney function and soluble receptor for advanced glycation end-products: The Long Life Family Study (LLFS).....	527
PB1403. Genetic prediction of weight and waist circumference regain after weight loss with Intensive Lifestyle Intervention: A secondary analysis of the Look AHEAD trial.	528
PB1404. Genetic Predictors of Memory Performance in Older Adults	529
PB1405. Genetic profiling and improved predictive capability with protein-based risk scores for IBD and disease progression using UK Biobank data.	530
PB1406. Genetic Regulation of Gene Expression in HIV+ T Cells Associated With Control of HIV spVL in African Populations.....	531
PB1407. Genetic regulation of gene expression in the frontal cortex across the human lifespan	532
PB1408. Genetic susceptibility for autoimmune diseases and white blood cell count variability.....	533

ASHG 2022 Annual Meeting Poster Abstracts

PB1409. Genetic Variability of <i>CYP2B6</i> Polymorphisms (785A>G, 64>CT and 516G>T) in Southwest Nigerian Population: Implications for Malaria Treatment.....	534
PB1410. Genetic variants in Olfactory Response Genes associated with Osteoarthritis progression to Hip or Knee Replacement among African and Hispanic American Veterans.	535
PB1411. Genetic variation in fatty acid amide hydrolase (<i>FAAH</i>) influences early drinking and smoking behaviors.	536
PB1412. Genetically Informed Prediction of Short Term Parkinson’s Disease Progression	537
PB1413. Genetically-associated body-mass index, vitamin D levels, and age at puberty are not associated with Parkinson’s Disease related phenotypes.....	538
PB1414. Genetics determinants of Renin-Angiotensin-Aldosterone System (RAAS) in the Cooperative Health Research In South Tyrol study	539
PB1415. Genetics of Gulf War illness, a genome wide association study.....	540
PB1416. <i>Genetics of ventricular septal defects: A novel genetic interaction between Sox7 and Wnt4 is associated with abnormal endocardial cushion morphogenesis.....</i>	541
PB1417. Genome scans of early childhood caries implicate bitter taste receptors.....	542
PB1418. Genome sequencing in brain samples from the ROSMAP cohort to examine the relationship between telomere length, mitochondrial copy number, and β -amyloid.....	543
PB1419. Genome wide association study of clinically predicted suicide liability.	544
PB1420. Genome-wide analysis in over 1 million individuals reveals over 2,000 independent genetic signals for blood pressure	545
PB1421*. Genome-wide association analyses of sepsis and septic shock in a large practice-based biobank.....	546
PB1422. Genome-wide association analysis of composite sleep scores in 413,904 individuals.....	547
PB1423. Genome-wide association meta-analysis and burden of rare coding variants in African American patients with Acne.....	548
PB1424. Genome-wide association meta-analysis reveals 99 risk loci for pain susceptibility and pleiotropic relationships with psychiatric, metabolic, and immunological traits.....	549
PB1425. Genome-wide association of Copy Number Variation for Cleft Lip with or without Cleft Palate in families from a multi-ethnic study sample.....	550
PB1426*. Genome-wide association of NASH and Mendelian randomization with plasma protein levels identifies putative protein changes resulting from disease.	551
PB1427. Genome-wide association studies identify genetic variants associated with comorbidities in atopic dermatitis.....	552
PB1428. Genome-wide association studies of human and ratbody mass converge on a conserved molecular network.	553
PB1429. Genome-wide association study analysis of glucose response in oral glucose tolerance test (OGTT).....	554
PB1430*. Genome-wide association study for 233 circulating metabolic traits in 136,000 participants reveals extensive pleiotropy and novel associations.....	555

ASHG 2022 Annual Meeting Poster Abstracts

PB1431. Genome-Wide Association Study in a Rat Model of Temperament Identifies Multiple Loci for Exploratory Locomotion and Anxiety-Like Traits.	556
PB1432. Genome-wide association study in mesial temporal lobe epilepsy with hippocampal sclerosis	557
PB1433. Genome-wide association study of diabetic retinopathy in the Million Veteran Program identifies four significant loci	558
PB1434. Genome-wide association study of longitudinal changes in motor symptomology in Parkinson's Disease	559
PB1435. Genome-wide association study of nephrotic syndrome replicates <i>APOLI</i> and MHC loci, while predicted gene expression analysis identifies <i>C4A</i>	560
PB1437. Genome-wide association study of plasma amyloid β levels in older adults	561
PB1438. Genome-wide association study of venous thromboembolism in women using combined oral contraceptives	562
PB1439. Genome-wide association study on skin tags.....	563
PB1440. Genome-wide association study to identify genetic variants for obesity in Korean population	564
PB1441. Genome-wide Interaction Study with Smoking Identifies <i>FHIT</i> and <i>SLC22A23</i> Associated with Alzheimer's Disease.	565
PB1442. Genome-wide meta-analysis identifies a potential therapeutic targets for new loci and functional pathways influencing Alzheimer's disease risk.....	566
PB1443. Genome-wide meta-analysis identifies novel susceptibility loci for acne vulgaris	567
PB1444. Genome-wide metabolite quantitative trait loci analysis in red blood cells from routine blood donors.....	568
PB1445. Genome-wide Polygenic Risk Scores in Diabetes Risk Prediction: Results from the Asian Indian Diabetic Heart Study/Sikh Diabetes Study	569
PB1446. Genome-Wide SNP Interaction Tests with Polygenic Risk Score for Ischemic Stroke Identifies Associations with Alzheimer Disease at <i>VCPKMT</i> and <i>KNDC1</i> by Age-Stratified Analysis	570
PB1447*. Genomic architecture of Autism Spectrum Disorder from comprehensive whole-genome sequence annotation.	571
PB1448. Genomic discovery with functional annotation of <i>CNTNAP5</i> in the phenotypic extremes of primary angle closure glaucoma	572
PB1449. Genomic structural equation modeling to unravel kidney-specific variants from multiple biomarkers.....	573
PB1450. GenoPipeR: An Automated Pipeline for Predictive Modelling of Complex Phenotypes using Genome-wide Data	574
PB1451. Genotypic and Phenotypic Heterogeneity in Prader-Willi Syndrome: A Bangladeshi cohort study.....	575
PB1452. Global analysis of RNA editing in Alzheimer's disease across multiple brain regions.	576

ASHG 2022 Annual Meeting Poster Abstracts

PB1453. Global Long COVID Host Genetics Initiative identifies <i>FOXP4</i> locus as the first genetic risk factor associated to Long COVID.....	577
PB1454. GWAS and Polygenic prediction of anthropometric traits in sub-Saharan African populations	578
PB1455. GWAS of the age-of-onset of type 1 diabetes reveals HTATIP2 as a novel T cell regulator.	579
PB1456. Heterozygous loss-of-function mutation in <i>SORL1</i> causes neuronal dysregulation of endosomal trafficking and processing of APP.	580
PB1457. Heterozygous <i>RYR2</i> , <i>ABCG8</i> , <i>PIK3C2G</i> , and <i>RASSF9</i> variants as the cause of premature coronary artery disease in five Iranian families.	581
PB1458. HLA Haplotype Analysis and Genome Wide Association Study of Nephrotic Syndrome in the Million Veteran Program	582
PB1459. HLA in autoimmune encephalitis.	583
PB1460. HLA types are associated with hundreds of complex traits and diseases in an ancestry-dependent manner	584
PB1461. HLA typing and COVID-19 disease outcome associations: GENCOV Study Canada.	585
PB1462. <i>HLA</i> associations with autoantibody-defined subgroups in idiopathic inflammatory myopathies	587
PB1463. <i>HLA-DRB1*15:01</i> modifies sphingolipid levels in multiple sclerosis	588
PB1464. Host Genetics in Resistance to COVID-19: Hints from recovered Brazilian superelderly ...	589
PB1465. How are <i>APOE4</i> , changes in weight/BMI, and longevity related? Insights from a Causal Mediation Analysis.	590
PB1466*. Hypermethylation of <i>PM20D1</i> promoter is associated with atherosclerosis in Dominican Families.....	591
PB1467. Hypertension Prediction From Metabolomics For Two Cardiovascular Disease Cohorts and Transfer Learning of Resultant Models	592
PB1468. Identification of a locus associated with major depressive disorder in a Colombian population	593
PB1469. Identification of genetic loci associated with peanut specific IgG4 in high risk children from the LEAP study	594
PB1470. Identification of genetic loci with divergent effects between Crohn's disease and ulcerative colitis.....	595
PB1471. Identification of novel genetic variants for association with rheumatoid arthritis susceptibility in Pakistani population.....	596
PB1472. Identification of platelet aggregation genetic determinants: Results from a cross-ancestry GWAS meta-analysis.....	597
PB1473. Identification of shared genes and pathways between psychiatric GWAS and monogenic neurodevelopmental disorders	598

ASHG 2022 Annual Meeting Poster Abstracts

PB1474. Identifying compounds to treat opiate use disorder by leveraging multi-omic data integration and multiple drug repurposing databases.....	599
PB1475. Identifying Developmental Stuttering comorbidities using Electronic Health Records.	600
PB1476. Identifying genetic variations that connect iron homeostasis to metabolic disease.	601
PB1477. Identifying Metabolomic Signatures of Genetic Liability to Type 2 Diabetes in East Asian Adolescents.....	602
PB1478. Identifying novel gene dysregulation associated with opioid overdose death: a meta-analysis of differential gene expression in human prefrontal cortex.	603
PB1479*. Identifying phenotypic signatures of circulatory microRNAs.....	604
PB1480. Identifying Rare Variants Associated with Type 2 Diabetes Mellitus in GENNID: a Multiplex, Multiethnic, Family Study.	605
PB1481. Identifying shared genetic architecture across eye diseases using electronic health records.	606
PB1482. Identifying the genetically dysregulated risk and preventative pathways in Alzheimer's disease.....	607
PB1483. Impact of Interleukin-4, Interleukin-13 Gene Polymorphisms and HLA-DQ Alleles on Genetic Susceptibility of Type-1 Diabetes Mellitus (T1DM) in Kuwaiti Children.....	608
PB1484. Improving Common Disease Risk Prediction with Clinical Risk Factors using Polygenic Scores in a Primary Care Physician Network of 35,000 Patients Followed for 16 years.	609
PB1485. Improving model predictivity and explainability by combining genotypes and immunophenotypes in SARS-CoV-2 infected people.	610
PB1486. Improving the selection of variants included in polygenic risk scoring models.....	611
PB1487. In silico analysis to examine COPD genetic associations using scRNA-seq-based CRISPRi approach in iPSC derived alveolar type 2 (iPSC-AT2) cells.....	612
PB1488. Increased burden of pathogenic variants in known aortopathy genes in individuals with early onset thoracic aortic dissection.....	613
PB1490. Increased frequency of predicted loss of function variants in <i>PKP2</i> in African ancestry individuals.....	614
PB1491. Individuals with GCK-MODY are not at increased risk for common complications associated with T2D as measured by the Diabetes Complications Severity Index (DCSI).....	615
PB1492. Insights from HostSeq :Canada's National Whole Genome Sequencing Cohort of 10,000 Canadians Infected with SARS-CoV-2.....	616
PB1493. Insights into pediatric inflammatory bowel disease subtypes using single-cell RNA-sequencing.....	617
PB1494. Integrating common and rare variants into a genetic risk score for Alzheimer's disease risk prediction.	618
PB1495. Integrating Genomic Risk into Absolute Risk Estimates for Coronary Heart Disease.	619
PB1496. Integration Analysis of Multi-Omics Data for Osteoporosis Biomarker Discovery Identification.....	620

ASHG 2022 Annual Meeting Poster Abstracts

PB1497. Integration network-based analyses and Mendelian Randomization to screen drug repurposing candidates for osteoporosis	621
PB1498. Integration of proteomics quantitative trait loci into genetic association analysis of stroke in the African American population.....	622
PB1499*. Integrative analyses to identify plausible target genes on Alzheimer’s Disease genes with genetic evidence.....	623
PB1500*. Integrative analysis of large-scale multi-ancestry genome-wide association study and single-cell omics data provides high-resolution insight of cell-types in the pathogenesis of type 2 diabetes.	624
PB1501. Integrative genomic and transcriptomic analyses revealed new driver variants associated with progression of the non-alcoholic fatty liver disease.....	626
PB1502. Interactions between Valley Fever and genetic ancestry: does genetic ancestry affect risk of developing disseminated coccidioidomycosis?.....	627
PB1504. Interplay of age and Alzheimer’s disease diagnosis in estimating heritability	628
PB1505. Interrogation of shared genetic architecture of infectious disease traits and Alzheimer’s disease	629
PB1506. Intragenic loci within <i>TOMM40</i> modulate <i>APOE</i> expression in human microglia	630
PB1507. Investigating the cellular and molecular response to hyperglycemia in iPSC-derived cardiomyocytes	631
PB1508*. Investigating the influence of genetic ancestry on gene-environment interactions on polygenic risk score and acculturation: Results from the Hispanic Community Health Study/Study of Latinos.	632
PB1509. Irina St. Louis.....	633
PB1510. Is Whole-Blood Cell Mitochondrial Copy Number Lower in Chronic Obstructive Pulmonary Disease in the NHLBI trans-omics for precision medicine (TOPMED) Program?.....	634
PB1511. Isoform usage differences in schizophrenia.	635
PB1512*. Key discoveries of the genetic basis of stuttering.	636
PB1513. Known Alzheimer disease candidate variants influence the severity of neuropathologic lesions underlying Alzheimer disease and related dementias.	637
PB1514*. Large scale genome-wide association study highlights the causality of vasospastic angina	638
PB1515. Large-scale chromatin accessibility fine-mapping of blood lipids GWAS loci in human liver cell types.	639
PB1516. Large-scale exome sequencing identifies common and rare coding variants contributing to tinnitus	640
PB1517. Leveraging aggregation of juvenile idiopathic arthritis in large pedigrees to investigate susceptibility genes.	641
PB1518. Leveraging genetic variation to evaluate risk factors and therapeutic opportunities for aortic valve stenosis: a mendelian randomization analysis.....	642
PB1519. Leveraging spatial patterns in the distribution of dental caries to resolve genetic heterogeneity.	643

ASHG 2022 Annual Meeting Poster Abstracts

PB1520. Leveraging the nutritional geometry framework to dissect the impact of macronutrient composition on metabolic function and gene regulation in fat tissue	644
PB1521. Limited effect of Y chromosome variation on coronary artery disease and mortality in UK Biobank	645
PB1522. Longitudinal transcriptomic profiling of IBD patients reveals therapeutic decision-making determinants of drug response.	646
PB1523. Loss of chromosome Y in leukocytes as a risk factor for critical COVID-19 in men.	647
PB1524*. Loss-of-function variants association study highlights potential new causal genes implicated in blood pressure regulation.....	648
PB1525. Lung-specific alternative splicing is a shared etiological risk for COVID-19 severity and chronic respiratory diseases.	649
PB1526. Machine learning in COVID-19 phenotype prediction.....	650
PB1527*. Machine learning-based classification of Age-related Macular Degeneration using gene expression profiles	651
PB1528. Mapping Autism Spectrum Disorder Endophenotypes to Genomic Regions in Thousands of Families.....	652
PB1529. Metabolomic signatures of body mass index in the Hispanic Community Health Study/ Study of Latinos (HCHS/SOL)	653
PB1530. Metabolomics profiling upranks the role of natural steroids and phenylalanine metabolism over oxidative stress in insulin resistance amongst lean subjects.	654
PB1531. Metagenomic analysis of the gut resistome in dyslipidemia and healthy individuals.	655
PB1532. Mitochondrial and Nuclear genetic variants demonstrate mitochondrial function determines severity but not risk of amyotrophic lateral sclerosis.....	656
PB1533. Mitochondrial <i>TXNRD2</i> and <i>ME3</i> genetic risk scores are associated with specific primary open-angle glaucoma phenotypes	657
PB1534. Modeling cardiac cell developmental trajectories at high temporal resolution.....	658
PB1535. Modeling gene by environment interactions in post-traumatic stress disorder using hiPSC-derived neurons.....	659
PB1536. Modification of coronary artery disease clinical risk factors by coronary artery disease polygenic risk score	660
PB1537. Molecular and clinical characterisation of Polish Temple syndrome patients with 14q32 alterations.....	661
PB1539. Molecular epidemiologic and family history evidence suggests that myalgic encephalomyelitis (ME)/chronic fatigue syndrome (CFS) may be an autoimmune disorder.	662
PB1540. Molecular states during acute COVID-19 reveal distinct etiologies of long-term sequelae. .	663
PB1542*. Multi-ancestry HLA allele calling using whole-exome sequencing in the UK Biobank reveals 353 novel genome-wide significant HLA alleles associated with 11 auto-immune phenotypes	664

ASHG 2022 Annual Meeting Poster Abstracts

PB1543. Multi-ancestry meta-analysis of asthma improves polygenic risk prediction across populations	665
PB1544. Multi-ancestry meta-analysis of X chromosome-wide associations for height.....	666
PB1545. Multiple HLA haplotypes and a variant altering immunogenicity of minor histocompatibility antigen epitopes encoded by <i>CTSH</i> are associated with age at type 1 diabetes diagnosis.....	667
PB1546. Multi-tissue splicing transcriptome-wide association study identifies 25 new candidate susceptibility genes and a new risk region for Alzheimer’s disease	668
PB1547*. Multi-tissue transcriptome-wide association study (TWAS) uncovers 10 novel genes associated with multiple ageing outcomes.....	669
PB1548. Multitrait analysis genome-wide association study of atherosclerosis phenotypes	670
PB1549. Multi-trait rare variant analysis of cardiometabolic traits	671
PB1550. Multivariate GWAS of 3D cranial vault shape in multi-ethnic and admixed children	672
PB1551. Natural language processing and modelling of clinical brain disease trajectories.....	673
PB1552. Neurobiological pathways underlying opioid addiction identified by systems biology multi-omic gene integration.....	674
PB1553. Neuroimmunogenic architecture in brains of Alzheimer's disease at single-cell resolution..	675
PB1554. Neuronal Assay for Investigating Circadian Rhythm Defects in Prader-Willi and Schaaf-Yang Syndrome	676
PB1555. NGS testing and risk stratification of malignant cardiac arrhythmias	677
PB1556. Non-desmosomal genes in Arrhythmogenic Cardiomyopathy: genetic variants rating.....	678
PB1557. Non-Mendelian inheritance patterns and extreme deviation rates of CGG repeats in autism	679
PB1558. Not one and done: A rare finding of two deleterious variants contributing to the onset of rhabdomyolysis.	680
PB1559. Novel Insights into Pediatric Scoliosis Revealed by Genome-wide Association Study and Whole Exome Sequencing	681
PB1560. Obesity genomic loci are heterogeneously associated with lipid profiles in ancestrally diverse <i>Population Architecture using Genomics and Epidemiology (PAGE)</i> study.	682
PB1561. Opportunities in national SARS-CoV-2 genomic surveillance programs for accelerating study on human genetic determinants of differential COVID-19 severity	683
PB1562. Osteoarthritis has high genetic heritability in 488,421 multi-ancestry participants from MVP and the UK Biobank.....	684
PB1563. Overexpression screen of chromosome 21 genes reveals modulators of Sonic hedgehog signaling relevant to Down syndrome.	685
PB1564*. Parsing Transcriptomic and Variant Signatures in Tetrapartite Brain Regions Uncover a Collection of Novel Genes in the Neuropeptide-Neurotransmitter Axis Conferring Addiction Risk ..	686
PB1565. Partitioned polygenetic risk scores of adipocyte marker genes and sex-specific GWAS variants explain sex-specific differences in abdominal obesity	687
PB1566. Patterns of human germline hypermutability identified with whole-genome sequencing	688

ASHG 2022 Annual Meeting Poster Abstracts

PB1567. Pediatric Manifestations of Genetic Risk Factors for Polycystic Ovary Syndrome	689
PB1568. Perceived benefits and barriers to implementing polygenic risk scores in primary care: results of a national physician survey.....	690
PB1569. Performance of externally developed polygenic risk scores in the All of Us Research Program Database.....	691
PB1570. Phenome-wide analyses with nonsynonymous variants in SOS2 demonstrate remarkable pleiotropic associations in the UK Biobank.....	692
PB1571*. Phenome-wide association studies of deleterious variants in the Han Taiwanese people ...	693
PB1572. Phenome-Wide Association Study (PheWAS) investigates the role of long-read discovered structural variants in cardiovascular disease risk in the UK Biobank using an imputation framework	694
PB1573*. Phenome-wide perspective into the role of Finnish Y chromosome variation in complex diseases	695
PB1574. Phenome-wide PGS portability in the Colorado Center for Personalized Medicine biobank suggests overlooked challenges in diverse populations	696
PB1575. PLA2G6 associated late onset Parkinson's disease in a Sudanese family: A case report.....	697
PB1576. Polycystic Ovarian Syndrome Physiologic Pathways Implicated Through Clustering of Genetic Loci.....	698
PB1577*. Polygenic burden of major depressive disorder and suicidality in children of diverse ancestries: a US-population-based study	699
PB1578. Polygenic risk and complex trait prediction for East Asians using the Taiwan Precision Medicine Initiative and Taiwan Biobank datasets (N = 500k).	700
PB1579. Polygenic risk burden is associated with early psychosis and endophenotypes of psychotic disorders.....	701
PB1580. Polygenic risk for body mass index is associated with atypical antipsychotic induced weight gain.....	702
PB1581. Polygenic Risk score based phenome wide association study for Tourette Syndrome.....	703
PB1582. Polygenic risk scores affected by different strategies of selection of relevant genomic positions and equations used for calculations	704
PB1583. Polygenic risk scores for quantitative estimation of complex vitamin intake needs to lower homocysteine in the adult population: from population based to personalized dietary intake recommendations.	705
PB1584. Polygenic risk scores identify heterogeneity in disease diagnosis and treatment in asthma and COPD.....	706
PB1585. Polygenic score for height influences adolescent growth spurt in Southwestern Indigenous Americans.	707
PB1586. Polygenic score prediction of atrial fibrillation following cardiac surgery: a retrospective hospital biobank study.	708
PB1587. Population level genetic screening for coronary artery disease with polygenic risk scores and rare variants in a diverse electronic health record linked biobank.....	709

ASHG 2022 Annual Meeting Poster Abstracts

PB1588*. Population-scale proteo-genomic analysis reveals novel loci associated with Parkinson's disease	710
PB1589. Portability of a multiethnic polygenic risk score for low-density lipoprotein cholesterol in a Samoan population.....	711
PB1590. Potential effects of SORBS1 variants on serum immunoglobulin E levels specific to milk intake in relation to prevalence of atopic dermatitis in Korean children	712
PB1591. Predicted loss-of-function variants for blood lipids in over a million individuals.....	713
PB1592*. Predicted polygenic transcriptional risk score supports the inference of canalization of polygenic risk of common diseases and traits in the UKBiobank.....	714
PB1593. Prediction of Alzheimer's disease conversion using polygenic risk scores grouped by transcriptome profiles from blood and brain.....	715
PB1594. Pre-infection antiviral innate immunity attenuate SARS-CoV-2 infection and viral load in iPSC derived alveolar epithelial cells type-2.	716
PB1595*. Prioritization of causal genes at Parkinson's disease associated loci using machine learning.	717
PB1596*. Prioritizing genes and gene programs for disease by integrating genetic and perturbation data	718
PB1598. Profiling the neurobiology underlying brain structure in living human subjects.....	719
PB1599. PRS-PheWAS of psychiatric/ behavioral traits to disentangle clinical heterogeneity in Substance Use Disorders.....	720
PB1600*. Quantification of race, ethnicity, and genetic ancestry disparities in anti-hypertensive drug efficacy in the All of Us Research Program	721
PB1601. Quantifying differential multiple sclerosis risk via polygenic risk scores in African American and Hispanic cohorts.....	722
PB1602. Quantifying mitochondrial dysfunction at biobank scale using insights from rare disease ...	723
PB1604. Rare copy number variants in <i>CACNA1H</i> implicated in essential tremor.	725
PB1605. Rare copy number variation in the GR@ACE/DEGESCO dementia dataset of spanish population.	726
PB1606. Rare <i>Plasmodium falciparum</i> coronin gene mutations following ACT treatment of malaria in South Western Nigeria	728
PB1607. Rare protein-truncating DNA variants in APOB or PCSK9, low-density lipoprotein cholesterol, and risk of coronary artery disease.....	729
PB1608*. Rare protein-truncating variants in <i>ZNF518A</i> are associated with shorter female reproductive lifespan.....	730
PB1609*. Rare variant genetic architecture of mitochondrial DNA copy number from 415,422 exomes.	731
PB1610. Rare variants in ADAMTS13 lead to COVID-19 autosomal dominant disorder, conditioned by SARS CoV-2 infection, sex, and age.....	732
PB1611. Rare variation in ADHD and related diagnoses in the Mass General Brigham Biobank.....	733

ASHG 2022 Annual Meeting Poster Abstracts

PB1612. Rare-Inherited variations in Sensory genes increased burden of Autism Spectrum Disorders: A Whole Exome Sequencing study of Indian families	734
PB1613. Relationship between the gut microbiome and depression in individuals with post traumatic stress disorder.....	735
PB1614. Results from largest GWAS in Gastroparesis point to macrophage polarization etiology	736
PB1615. <i>RGS3</i> and <i>IL1RAPL1</i> missense variants implicate defective neurotransmission in early-onset inherited schizophrenias.....	737
PB1616. Risk factors involved in Congenital Diaphragmatic Hernia, a case-control study in Bogotá and Cali, Colombia.	738
PB1617*. Risk factors that affect performance of polygenic risk scores across diverse cohorts	739
PB1618. RNA editing regulates interferon induction and host immune response in SARS-CoV-2 infection	740
PB1619. RNASeq in nasal epithelium from African ancestry subjects in the CAAPA consortium reveal <i>IL4</i> and <i>TGFB1</i> as upstream regulators of differentially expressed networks for asthma	741
PB1620. Role of amino acids in COVID-19 virulence and pathogenicity : an observational study	743
PB1621. SARS-CoV-2 sequencing: A comparison of high-throughput methods.	744
PB1622. Screening of variant rs11190870 nearby <i>LBX1</i> gene for association with Adolescent Idiopathic Scoliosis in the population of North India	745
PB1623*. Secretoglobin family 1D member 2 (SCGB1D2) protein inhibits growth of <i>Borrelia burgdorferi</i> and affects susceptibility to Lyme disease.....	746
PB1624. Self-reported and device-based physical activity and risk of Alzheimer's disease and related dementias: a Mendelian randomization study.....	747
PB1625. Sequencing-based genome-wide association study of triglycerides in East Africans.....	748
PB1626. Serum osteopontin levels in Mexican patients with systemic lupus erythematosus with lupus nephritis and molecular interaction review.....	749
PB1627. Serum proteomic signatures predicting unintentional weight loss in patients with chronic obstructive pulmonary disease	750
PB1628. Sex dependent transcriptional changes in response to stress in patients with myalgic encephalomyelitis/chronic fatigue syndrome.....	751
PB1629*. Sex-specific analysis of rare variant associations with quantitative traits in the UK Biobank.	752
PB1630. Sex-specific and multiomic integration to enhance accuracy of peripheral blood biomarkers of major depressive disorder	753
PB1631. Sex-Specific associations of gene expression in brain with Alzheimer's disease pathology and cognitive performance.	754
PB1632. Sex-specific genetic loci linked to early and late onset type 2 diabetes.....	755
PB1633. Sex-specific genetic predictors of brain amyloidosis in Alzheimer's disease.	756
PB1634*. Sex-specific genetic predictors of memory performance in older adults.....	758
PB1636. Sex-specific risk scores for coronary artery disease using machine learning models.....	760

ASHG 2022 Annual Meeting Poster Abstracts

PB1637. Sex-stratified meta-analysis of age-related cognitive decline across neurocognitive domains	761
PB1638. <i>SHANK3</i> domain specific functional characterization of a spectrum of neurodevelopmental disorders phenotypes.....	762
PB1639. Shared genetic basis informs the roles of polyunsaturated fatty acids in brain disorders.....	763
PB1641. Shriners Children’s international program on the genetics of rare pediatric disorders	764
PB1642. Significant heritability enrichment for Schizophrenia located in single neuron electrophysiology genes.....	765
PB1643. Single cell patch-seq in understanding the regenerative ability of cortical spinal motor neurons	766
PB1644. Single nucleotide variants genotyping of IL-6, IL-1 β , TNF- α genes promoter region and its protein levels associated to inflammatory response in obstructive sleep apnea and periodontal disease	767
PB1645. Single-Cell RNA-Seq Reveals the connections of amyloid-beta and ferroptosis tendency... 768	
PB1646. SnRNA-seq probing of the differential vulnerability of motor neurons in ALS.....	769
PB1647. SNX8 as a novel predisposing factor by an integrative study for the risk of childhood atopic dermatitis in COCOA	770
PB1648. Socioeconomic Status and Risk of Stroke: A Mendelian Randomization Study.....	771
PB1649*. Somatic mutations in chronic lung disease are associated with reduced lung function.....	772
PB1650. Spectrum of causal genetic variants in inherited cardiomyopathies and arrhythmias in a primarily French-Canadian Population: A 14-year retrospective study from a specialized Cardiovascular Genetic Center.	773
PB1651. Stratification of a PMS population based on their response to Human Growth Hormone and Insulin-like Growth Factor 1	774
PB1652. Stratified LD score regression with 3D genomic features in metabolically relevant cell types implicates a role for pancreatic alpha cells in the pathogenesis of childhood obesity.....	775
PB1653. Study of gene-environment interactions suggests a personalised medicine approach for ALS	776
PB1655. Telomere length associated rare variant candidate gene study in idiopathic pulmonary fibrosis	777
PB1656. Telomere length, and Parkinson's disease risk, age at onset, and presentation: Mendelian randomization analyses suggest no causal relationships.....	778
PB1657. THE ASSOCIATION OF <i>FSHR</i> POLYMORPHISMS WITH PCOS IN PUNJABI POPULATION	779
PB1658. The bidirectional causal effects of brain morphology across the life course and risk of Alzheimer’s disease: A cross-cohort comparison and Mendelian randomization meta-analysis.	780
PB1659. The causal effect of blood lipid metabolism related traits with depression phenotypes: the evidence from mendelian randomization study	781

ASHG 2022 Annual Meeting Poster Abstracts

PB1660. The effect of population structure and ancestry-related assortative mating on the genetic architecture of autism spectrum disorder. 782

PB1661*. The first GWAS on intrahepatic cholestasis of pregnancy unveils novel bile acid metabolism related associations. 783

PB1662. The genetic architecture of the corpus callosum and its subregions 784

PB1663. The genetic drivers of young- and early-onset Parkinson’s Disease in India 785

PB1664. The genetic risk score affects the lifetime blood pressure and obesity modify the effect of genetic risk 787

PB1665. The heritable component of human longevity is highly polygenic and pleiotropic. 788

PB1666. *The impact of copy number variants in the genetic testing yield of LQTS patients.* 789

PB1667. The impact of exercise on gene regulation in association with complex trait genetics 790

PB1668. The impact of HIV-1 Vpr on host transcription in macrophages..... 791

PB1669. The interacting effects of environmental factors and polygenic risk scores on BMI: application to the multiethnic GENNID family study 792

PB1670. The kynurenine pathway in Alzheimer’s disease: A systematic review and meta-analysis. . 793

PB1671. The Largest Asian GWAS for Systemic Sclerosis Identified a Novel High Risk Candidate Causal SNP in the Fcγ-Receptor Gene Region. 794

PB1672*. The Pan-UK Biobank project improves locus discovery, facilitates genetic architecture comparisons, and increases the resolution and generalizability across diverse populations..... 795

PB1673. The *PPARG1* gene regulates the aging retina by the modulation of mitochondrial functions, autophagy and senescence in a mouse model of age-related macular degeneration..... 796

PB1674. The prevalence, penetrance, and expressivity of mitochondrial disorders in a population-based cohort. 797

PB1675. The relationship of Alzheimer disease, genetic risk, and education in individuals of African Ancestry. 798

PB1676. The role of host genetics on weight gain following ART initiation in people living with HIV 799

PB1677. The role of *POLG* in the genetic etiology of Amyotrophic Lateral Sclerosis..... 800

PB1678. The variant in *FLNA* identified in a patient with progressive supranuclear palsy 801

PB1679. The whole blood transcriptional regulation landscape in 465 COVID-19 infected samples from Japan COVID-19 Task Force 802

PB1680. Tissue-associated examination of risk factors uncovers genetic risk profiles that vary across patients of inflammatory bowel disease. 803

PB1681. To annotate the neuroprotective role of Camel α Lactalbumin and Oleic acid (CLOA) complex against Parkinson’s induced model. 804

PB1682. Trans-ancestry genome-wide analysis of kidney disease in individuals with type 2 diabetes. 805

PB1683. Trans-ancestry PRSs of lipid traits: characterization and evaluation of clinical utility in the diverse PAGE study 806

ASHG 2022 Annual Meeting Poster Abstracts

PB1684*. Transcriptome analysis of familial dysautonomia reveals tissue-specific gene expression disruption in the peripheral nervous system	807
PB1685. Transcriptome-wide analysis in Cameron County Hispanic Cohort (CCHC) provides novel insights into gene expression signatures of blood pressure	808
PB1686*. Transcriptome-wide and Proteome-wide association study of Tourette’s Syndrome.....	809
PB1687. Transcriptome-wide association analyses in CD4 T-cells identify genes dysregulated in frequent cocaine users and associated with HIV latent reservoir.....	810
PB1688. Transcriptome-wide association study identifies novel candidate susceptibility genes for migraine.	811
PB1689. Transcriptomic analysis of severe obesity identifies novel genes in Hispanic/Latino populations with a high burden of disease.	812
PB1690. Transcriptomic analysis of whole blood in ancestrally diverse Alzheimer Disease cohorts implicates convergent immune and lipid processing molecular pathways	813
PB1691. Transcriptomic and Epigenomic Consequences of Heterozygous Loss-of-function Mutations in <i>AKAP11</i> , a Bipolar Disorder Risk Gene, and their Effects on Responsiveness to Lithium.....	814
PB1692*. Trans-ethnic meta-analysis in a multi-ethnic population refines multiple sclerosis susceptibility loci and identifies novel locus	815
PB1693. <i>TTN</i> truncating variants in cardiac-expressed exons show high penetrance for cardiomyopathy in carriers with atrial fibrillation	816
PB1694. Understanding the effect of non-coding de novo mutations within craniofacial enhancers in trios with orofacial clefts	817
PB1695. Understanding the etiology of Alzheimer's disease using genetically predicted gene expression and brain feature models.....	818
PB1696. Understanding the molecular basis of SARS-CoV-2 pathogenesis at a gene expression level using Applied Biosystems™ TaqMan™ Flexible Array Panels.....	819
PB1697. Undiagnosed complex cases with neurodevelopmental disorders (NDD) after exome sequencing-what’s next?.....	820
PB1698. Universal clinical Parkinson’s disease axes identify a major influence of neuroinflammation. Universal clinical Parkinson’s disease axes identify a major influence of neuroinflammation.	822
PB1699. Unraveling the genetic basis of autoimmune hypothyroidism.....	823
PB1700*. Unsupervised learning revealed metabolic syndrome sub-groups with differing phenotypic and genotypic traits	824
PB1701. Untargeted metabolomics on nutrient density and their genomic interaction among Mexican Americans	825
PB1702*. Using genetics to uncouple highly correlated metabolic phenotypes and test their separate role in disease.....	826
PB1703. Using Independent Component Analysis to disentangle obesity heterogeneity discovers genetic drivers of clinical disease signatures and their variation with BMI status.	828
PB1704*. Using Stimulation of B cells to uncover novel disease-associated QTLs.....	829

ASHG 2022 Annual Meeting Poster Abstracts

PB1705. Using whole genome sequencing to understand the genetic architecture of cerebral palsy in a Canadian cohort.	830
PB1706. USP53 variant associated with psychosis in a consanguineous pedigree.	831
PB1707. Validation of a rapid TTR genotyping assay as a point-of-care tool for cardiac amyloidosis diagnosis in low-income settings	832
PB1708. Variant analysis of disease causing genes and risk factor genes in patients with Parkinson's Disease.	833
PB1709. Variants in <i>GBA</i> and <i>MAPT</i> influence parkinson disease risk and age at onset in a colombian patient.....	834
PB1710. Variant-to-gene mapping in human microglial cell models with clonal CRISPR validation implicates <i>RTFDC1</i> and <i>CASS4</i> at the Alzheimer's disease ' <i>CASS4</i> ' locus.....	835
PB1711. Variation and impact of polygenic hematological traits in monogenic sickle cell disease	836
PB1712. Whole exome sequencing identifies two heterozygous novel variants in <i>CCL21</i> and <i>ITGA2</i> as a cause of Hypertrophic cardiomyopathy	837
PB1713. Whole Exome Sequencing revealed spectrum of mutations associated with different myopathy in clinically suspected DMD patients of Bangladesh	838
PB1714. Whole exome sequencing suggests a role for rare genetic variation in the <i>NLRX1</i> gene in HIV disease progression in pediatric African populations of Botswana and Uganda.	840
PB1715. Whole genome sequence analysis of long non-coding RNAs for plasma lipid traits	841
PB1716. Whole Genome Sequencing Analyses of 87,652 Individuals Reveal Rare Variants in Promoter of <i>HMGAI</i> Associated with Height	842
PB1717. Whole genome sequencing enables identification of a genetic susceptibility locus for idiopathic pulmonary fibrosis in the 16p subtelomere.	843
PB1718. Whole genome sequencing of a diverse Hispanic IBD population in the United States reveals differences in previously identified risk alleles.	844
PB1719. Whole genome sequencing study in multiple sclerosis identifies novel <i>CTBP2</i> association with change in brain lesion burden	845
PB1720. Whole-exome sequencing in five families with specific language impairment (SLI) suggests novel SLI genes and confirms previous findings.	846
PB1721. Whole-genome sequencing analysis of fructosamine and glycated albumin in Black and White participants in the Atherosclerosis Risk in Communities (ARIC) Study.....	847
PB1722. X Chromosome Wide Association Study on Latin American cohort reveals new potential loci associated with Parkinson's Disease	848
PB1723. 4-years of Face2Gene in a General Genetics Clinic: Insights from Retrospective Analysis of Diagnosed Cases	850
PB1724. A recurrent de novo mutation in <i>ZMYND11</i>	851
PB1725. A case of atypical inheritance in late-onset Pompe disease	852
PB1726. A case of Fanconi anemia diagnosed due to <i>BRCA2</i> gene mutation triggered by cancer genome profiling test.	853

ASHG 2022 Annual Meeting Poster Abstracts

PB1727. A case of fatal cardiac arrest in a neonate diagnosed with VLCADD postmortem.....	854
PB1728. A case report of an Egyptian patient with a severe neurodevelopmental disorder and a novel biallelic loss-of-function variant in <i>GOLGA2</i>	855
PB1729. A case with mosaicism pigmentary and Prader-Willi Syndrome: importance of molecular diagnosis in cases of mosaicism.....	856
PB1730. A Difficult Dual Diagnosis: Dilemmas in the Era of Next Generation Sequencing.....	857
PB1731*. A dysmorphology physical examination entry system facilitates structured genetic phenotype capture and natural language processing.....	858
PB1732. A Highly Polymorphic VNTR in the <i>DRD4-DEAF1</i> Intergenic Region.....	859
PB1733. A homozygous variant in a novel gene " <i>AP2A2</i> " causes an early-onset hereditary spastic paraplegia in a Malian family.....	860
PB1734. A multi-branch family with mutation in <i>SLC26A4</i> gene from a village in the southeastern region of Iran.....	861
PB1735. A new case of YARS1 associated multisystemic disorder with compound heterozygous in a patient with Klinefelter syndrome.....	862
PB1736. A novel de novo nonsense variant in <i>TBX2</i> cause osteochondrodysplasia.....	863
PB1737. A novel homozygous missense mutation in <i>ARSK</i> causes a new subtype of MPS.....	864
PB1738. A novel homozygous variant in <i>HECW2</i> gene is associated with intellectual disability and epilepsy in a Malian family.....	865
PB1739. A Novel Homozygous Variant in Homologous Recombination Repair Gene <i>ZSWIM7</i> Causes Azoospermia in Males and Primary Ovarian Insufficiency in Females.....	866
PB1740. A novel homozygous variant in <i>IBA57</i> causing multiple mitochondrial dysfunction syndrome type 3.....	867
PB1741. A novel <i>LZTR1</i> variant in a Brazilian familial case with Noonan syndrome: comparative modeling and structural analysis.....	868
PB1742. A novel <i>PUM1</i> heterozygous missense variant causing severe developmental delay and epilepsy from infancy with dysmorphic features.....	869
PB1743. A novel splice-site variant causes <i>MYH2</i> -associated myopathy in a large family.....	870
PB1744. A novel XL gene for autism with psychiatric illness in adults.....	871
PB1745*. A <i>Plod2</i> mutant mouse model of Bruck syndrome.....	872
PB1746. A prospective, longitudinal observational natural history study of patients with NGLY1 Deficiency.....	873
PB1747. A recessive variant in <i>TFAM</i> causes mtDNA depletion associated with primary ovarian insufficiency, seizures, intellectual disability and hearing loss.....	874
PB1748*. A recurrent missense variant in <i>ITPR3</i> causes demyelinating Charcot-Marie-Tooth neuropathy.....	875
PB1749. A single-center genetic analysis of oculopharyngeal myopathy.....	876
PB1750. A substantial proportion of high myopia is caused by mutations in <i>CACNA1F</i> (XL CSNB).....	877

ASHG 2022 Annual Meeting Poster Abstracts

PB1751. A Zebrafish Model of Congenital Muscular Dystrophy Caused by POMT1 Loss of Function	878
PB1752. Allele-specific inactivation of epidermolysis bullosa simplex mutations using CRISPR-Cas9 and spraying of the corrected cell suspension onto the wounds	879
PB1753. Alteration of mitochondrial proteostasis in Costello syndrome	880
PB1754. An initiative for the diagnosis and study of rare and undiagnosed diseases in Mexico	881
PB1755. Angelman syndrome with mosaic paternal uniparental disomy caused by mitotic nondisjunction	882
PB1756. <i>ARG1</i> variants in Arginase 1 Deficiency: genetic characterization of participants in the pegzilarginase clinical trials	883
PB1757. <i>ARHGAP32</i> , encoding the Rho GTPase Activating Protein 32, is a novel candidate gene involving in autosomal dominant neurodevelopmental disorder spectrum	884
PB1758. ATP2C1 as a candidate for Interstitial Cystitis/Bladder Pain Syndrome	885
PB1759. Atypical 260 kb deletion on distal 22q11.22 involving <i>TOP3B</i> shows the significance of this gene in autism spectrum disorder	886
PB1760. Atypical molecular findings in patients with capillary malformations.	887
PB1761. Bi-allelic ACBD6 variants lead to a neurodevelopmental syndrome with progressive complex movement disorders	888
PB1762*. Biallelic <i>DMD</i> variants in a mildly-affected female without a Duchenne or Becker phenotype	889
PB1763. Biallelic founder mutation in <i>PDE2A</i> causes paroxysmal dyskinesia with Intellectual disability in Pakistani families	890
PB1764*. Bi-allelic <i>LETM1</i> variants perturb mitochondrial ion homeostasis leading to a clinical spectrum with predominant nervous system involvement	891
PB1765. Biallelic Loss of Function Mutations in <i>PYGM</i> Cause Hereditary Macular Dystrophy	892
PB1766. Bi-allelic loss of function variant in the <i>NRCAM</i> gene is associated with hereditary polyneuropathy phenotype	893
PB1767. Biallelic loss-of-function variants in mitochondrial phospholipase <i>PNPLA8</i> decrease in the number of basal radial glial cells and lead to microcephaly with simplified gyral pattern	894
PB1768. Bi-allelic loss-of-function variants in <i>PPFIBP1</i> cause a severe neurodevelopmental disorder with microcephaly, epilepsy and periventricular calcifications	895
PB1769. Biallelic nonsynonymous variants in <i>LTV1</i> are associated with LIPHAK syndrome, a novel poikiloderma-like disorder	897
PB1770. Biallelic pathogenic variants in the endosomal transport regulator <i>SNAPIN</i> cause a rare prenatal neuroanatomical syndrome	898
PB1771*. Biallelic <i>RAD51C</i> loss-of-function variants drive perizygotic SNV/indel hypermutator phenotype in a subject with Fanconi anemia complementation group O.	899
PB1772. Biallelic variants in <i>CSPG4</i> cause a novel neurodevelopmental disorder with intellectual disability, global developmental delay, and facial anomalies	900

ASHG 2022 Annual Meeting Poster Abstracts

PB1773. Biallelic variants in <i>FICD</i> leading to inactivation of BiP cause motor neuron disease.	901
PB1774. Bibliometric analysis of Alport Syndrome - the impact of next generation sequencing on diagnosis.	902
PB1775. Biallelic <i>ATOH1</i> Variant in Siblings With Pontocerebellar Hypoplasia, Developmental Delay, and Hearing Loss.....	903
PB1776. Biotin-thiamine responsive basal ganglia disease: A retrospective review of the clinical, radiological and molecular findings of cases in Kuwait with novel variants.....	904
PB1777. Brachyolmia with amylogenesis imperfecta: Identification of a Novel LTBP3 gene Mutation in several Druze Arab patients from the north of Israel.....	905
PB1780. Characterization of associated non-classical phenotypes in patients with deletion in WAGR region identified by chromosomal microarray: new insights and literature review.....	906
PB1781. Characterization of <i>CFTR</i> Variants Across Ancestries.....	907
PB1782. Characterization of transthyretin <i>TTR</i> missense variants for associations with hereditary transthyretin-mediated (hATTR) amyloidosis and hallmark symptoms in 470,000 UK Biobank whole exome sequences.....	908
PB1783. Characterizing human disease-causing mutations: A zebrafish-based approach.....	909
PB1784. <i>Chd8</i> mutation in mouse impacts cerebellar function across anatomical, genomic, electrophysiological and behavioral axes	910
PB1785. <i>CHP2</i> is a genetic modifier of risk of chronic <i>Pseudomonas aeruginosa</i> airway infection in cystic fibrosis.	911
PB1787. Clinical and molecular characteristics of idiopathic midaortic syndrome in pediatric patients - preliminary results.	912
PB1788. Clinical diversity and molecular mechanism of VPS35L-associated Ritscher-Schinzel syndrome.....	913
PB1789. Clinical usefulness of a custom Next Generation Sequencing gene panel in providing molecular diagnosis to patients with a broad spectrum of Neurodevelopmental Disorders	914
PB1790. Combined brain abnormalities in patients with beta-tubulin mutations.....	915
PB1791. Comparing phenotypes across five developmental and epileptic encephalopathies (DEEs) through evaluation of 2490 patient data years.	916
PB1792. Compound heterozygosity for variants in <i>DCAF1</i> in fraternal twins with neurodevelopmental disorder	917
PB1793. <i>Compound heterozygous IFT81 variants causing brachydactyly with cone-shaped epiphyses and generalized metaphyseal dysplasia</i>	918
PB1794. Compound heterozygous variants in <i>UFMI</i> gene in a case of hypomyelinating leukodystrophy	919
PB1795. Confirmation of association of TGFBI p.Ser591Phe mutation with variant lattice corneal dystrophy.....	920
PB1796. Congenital Anomalies Cluster: Linking clinicians and model system experts.	921
PB1797. Co-occurring anomalies in cases with achondroplasia	922

ASHG 2022 Annual Meeting Poster Abstracts

PB1798*. C-X3-C motif chemokine receptor 1 (<i>CX3CRI</i>) gene variants associated with lesion burden in Cerebral Cavernous Malformation	923
PB1799. De novo (deep-)intronic splicing mutations in patients with neurodevelopmental disorders	925
PB1800. <i>De novo</i> missense variants in <i>RPS5</i> cause a newly characterized ribosomopathy.....	926
PB1801. <i>De novo</i> missense variants in <i>SEPHSI</i> Cause a Novel Neurodevelopmental Disorder with Developmental Delay, Hypotonia, Muscle Weakness, Speech Delay, and Growth delay	927
PB1802. De novo non-synonymous <i>CTR9</i> variants are associated with motor delay and macrocephaly	928
PB1803. <i>De novo</i> truncating <i>ADNP</i> variants as a recurrent cause of neurodevelopmental disorders ..	929
PB1804. Deciphering a Diagnostic Odyssey of Atypical Free Sialic Acid Storage Disorder Associated with Tissue Specific Mosaic Variant in <i>SLC17A5</i>	930
PB1805. Deciphering the phenotypic variability of the most common deafness-causative <i>GJB2</i> p.V371 variant in Asia	931
PB1806. Deep resequencing of the 1q22 locus in non-lobar intracerebral hemorrhage.....	932
PB1807. Defining the clinical heterogeneity of <i>GARS1</i> -related Charcot-Marie-Tooth disease	933
PB1808*. Defining the nuclear genetic architecture of a maternally-inherited mitochondrial disorder.	934
PB1809. Delayed developmental milestones in <i>MBD5</i> -associated neurodevelopmental disorder (MAND) associated with 2q23.1 deletions highlight need for early diagnosis	936
PB1810. Deleterious <i>SNAPC4</i> Variants are Associated with a Neurodevelopmental Disorder.	937
PB1811. Deletions of 14q32.2 result in severe neurodevelopmental outcomes and multiple congenital anomalies: three affected males and review of the literature.	938
PB1812*. Delineation of gain-of-function <i>MYCN</i> -induced novel megalencephaly syndrome and possible out-of-brain complications implicated by a mouse model	939
PB1813. Developmental Delay at the Single Cell Level in Prader-Willi Syndrome.....	940
PB1814. Diagnosis, treatment, and follow-up care for patients with Barth syndrome is presented through the results from 1:1 interviews of pediatric cardiologists.....	941
PB1815. Diagnostic rate and clinical utility of whole genome sequencing in adults with intellectual disability.....	942
PB1816. Diagnostic utility of comprehensive genomic and transcriptomic profiling of anorectal malformations.	943
PB1817. Dilated cardiomyopathy in a 3-month-old female: The answer in genetics.....	944
PB1818. <i>Drosophila</i> model of <i>de novo</i> <i>MRTF-B</i> variants highlights the critical role of actin regulation.	945
PB1819. Effect of <i>FOXO3</i> and Air Pollution on Cognitive Function: A Longitudinal Cohort Study of Older Adults in China	946
PB1820*. Effective Genetic Diagnosis for Unilateral and Asymmetric Pediatric Hearing Loss via Exome Sequencing.....	947
PB1821. Ellis-Van Creveld Syndrome: Clinical and Molecular Analysis of 50 Individuals	948

ASHG 2022 Annual Meeting Poster Abstracts

PB1822. Enrichment of <i>RAI1</i> genetic aberrations associated with sleep disturbances in Smith-Magenis syndrome and autism spectrum disorder.....	949
PB1823. Enzymatic testing for MPSI in Kuwait: A Pilot Study toward newborn screening.	950
PB1824. ERF-Related craniosynostosis: further delineation of the new craniosynostosis syndrome..	951
PB1826. Exome Sequencing in Understanding the Etiologies of Ataxia in Children	952
PB1827. Exome sequencing leads to the identification of a rare case of an autosomal dominant non-syndromic hearing disorder in a German family.....	953
PB1828. Exome sequencing uncovers novel variants in two Malian families with Epileptic encephalopathies	954
PB1829. Expanding the Phenotypic spectrum of <i>SMARCA5</i> -related neurodevelopmental disorder....	955
PB1830. Expanding the phenotypic spectrum of <i>TAB2</i> related disorders.	956
PB1831. Expert perspectives on sequencing newborns for treatable genetic conditions	957
PB1832*. Exploration of Ubiquitin-Proteasome System involvement in neurodevelopmental diseases using cellular models.	958
PB1833. Exploring cancer cachexia on the spatial plane	959
PB1834. Exploring the distinct phenotypes of nine Romanian patients harboring a homozygous Arg355* variant in <i>BBS12</i>	960
PB1835. Family with case of HMG-CoA lyase deficiency with novel pathogenic variant in <i>HMGCL</i>	961
PB1836. First patient reported with a <i>TPP1</i> missense mutation pathogenic variant predicted insilico causing <i>CLN2</i>	962
PB1837. First report of severe hemolysis associated with <i>SCARB1</i>	963
PB1838. Fracture prevalence in children diagnosed with Ehlers-Danlos Syndrome and Generalized Joint Hypermobility	964
PB1839. Fragile X Syndrome in a female patient: A case of strong family history and developmental delay.....	965
PB1840. Frameshift <i>PPP1R12A</i> pathogenic variant in a mexican patient with differences of sex development (DSD), middle line defects and hemangiomas.....	966
PB1841. Frequency of Y chromosome microdeletion and its correlation with spermatogenesis defect in north Indian infertile males.	967
PB1842. Functional analysis of <i>EZH1-A678G</i> in <i>Drosophila melanogaster</i>	968
PB1843. Functional characterization of recurrent truncating variant in <i>UBAP1</i> associated with hereditary spastic paraplegia.	969
PB1844. Gaps in the phenotype descriptions of ultra-rare genetic conditions: Systematic review and recommendations	970
PB1845. <i>GATA3</i> -related pedigrees showing characteristic audiologic profile and inheritance pattern.	971
PB1846. Genealogy as a predictor of disease progression in patients with myotonic dystrophy type 1: A demonstration of the power of intersectorial research.	972

ASHG 2022 Annual Meeting Poster Abstracts

PB1848. Genetic analysis reveals that GNE Myopathy remains an underdiagnosed neuromuscular disorder	973
PB1849. Genetic and clinical landscape of childhood cerebellar hypoplasia and atrophy.	974
PB1850. Genetic and phenotypic heterogeneity of childhood-onset hearing loss and implications for the success of cochlear implants.	975
PB1851. Genetic insights from consanguineous cardiomyopathy families.	976
PB1852. Genetic Mapping Identifies A Founder Mutation of <i>GHRHR</i> Gene in Pakistani Family with Isolated Growth Hormone Deficiency Type 4.	977
PB1853. Genetic modifiers in Niemann-Pick type C1 disease.	978
PB1854*. Genetic modifiers of nutritional status in CF are both unique to CF and shared with the general population.	979
PB1855. Genetic underpinnings of moderate to severe hearing loss in singleton individuals born to consanguineous couples in Pakistan	980
PB1856. Genetics of intellectual disability in pakistani consanguineous families.	981
PB1857. Genome sequencing identifies coding and non-coding variants in hereditary deafness missed by exome sequencing	982
PB1858. Genomic and transcriptomic characterization of a novel <i>RYR1</i> variant in the French-Canadian population	983
PB1859. Genomic ascertainment and reverse phenotyping diabetogenic variants in a phenotypically unselected cohort.	984
PB1860. Genomic region identified on chromosome 22 contributing to lymphedema in Phelan-McDermid syndrome	985
PB1861. GGC repeat expansion within <i>NOTCH2NLC</i> causes behavioral deficits and neurodegeneration in a mouse model of Neuronal Intranuclear Inclusion.	986
PB1862. GWAS of cleft palate trios reveals novel associations and subtype-specific effects.	987
PB1863*. GWAS of Down Syndrome Associated Atrioventricular Septal Defect identifies three novel loci.	988
PB1864*. Haploinsufficiency and loss-of-function of <i>LEF-1</i> cause a novel Mendelian disorder by dysregulating WNT signaling	989
PB1865. Haplotype in interleukin 6 is associated with a reduction of blood transfusion events in sickle cell anemia patients.	990
PB1866. Hemizygous CNV in chromosome region Xq23 is associated with a complex psychiatric and bone mineral density disorder.	991
PB1867*. Here, There and Everywhere: Characterizing Variability in Approaches to Ascertaining Individuals with Rare Mendelian Disorders from EHRs	992
PB1868. Heterogeneity of Inherited Cone Dysfunction Disorders with Normal Fundus Appearances	993
PB1869. High Molecular Diagnostic Yields and New Phenotypic Expansions Involving Anorectal Malformations	994

ASHG 2022 Annual Meeting Poster Abstracts

PB1870. High prevalence of frontotemporal dementia in females of five Hispanic families with R159H VCP multisystem proteinopathy.	995
PB1871*. Highly recurrent histone H4 mutations cause a neurodevelopmental disorder.....	996
PB1872. Homozygous <i>EMCI</i> variant in four families from the same Kuwaiti tribe with cerebellar atrophy, visual impairment, and psychomotor retardation.....	997
PB1873*. How many diseases can one gene cause: Why mechanism matters for gene curation, variant classification, and clinical management.....	998
PB1874*. Human brains with Tay Sachs disease exhibit altered transcriptomes during fetal development.....	999
PB1875. Human mutations in <i>Slitrk3</i> implicated in GABAergic synapse development in mice	1000
PB1876. Hypophosphatemia gene panel sponsored program: a comparison of findings from 2019 through 2022 describing <i>PHEX</i> variant landscape and molecular diagnosis of XLH.	1001
PB1877. Identification and Functional Evaluation of Autosomal Recessive Non-Syndromic Hearing Impairment Genes in Rwanda.....	1002
PB1878. Identification of a novel compound heterozygous mutation in <i>RyR1</i> gene in an Indian family affected with congenital myopathy	1003
PB1879. Identification of a novel pathogenic variant in <i>HIST1H1E</i> encoding the H1 histone linker in a patient with a complex phenotype including progeroid feature, intellectual disability and lipodystrophy. Contribution of additional variants to the complexi.	1004
PB1880*. Identification of a Novel Therapeutic Target Underlying Atypical Manifestation of Gaucher Disease	1005
PB1881. Identification of a pathogenic gene mutation in <i>Med12</i>	1006
PB1882. Identification of <i>INHB</i> -related neurodevelopmental disorder.....	1007
PB1883. Identification of molecular causes of recessively inherited ataxic and neuropathic disorders in consanguineous Pakistani families.....	1008
PB1884. Identifying modifier genes in a PIGA-CDG pedigree with reduced penetrance.....	1009
PB1885. Identifying phenotypic expansions for congenital diaphragmatic hernia plus (CDH+) using DECIPHER data.	1010
PB1886. <i>In Silico</i> genomic interrogation reveals 16 autosomal dominant neurodevelopmental candidate genes at 1p13.3.....	1011
PB1887. Incontinentia pigmenti female with the <i>IKBKG/NEMO</i> del4-10 deletion: A mosaic form. .	1012
PB1888. Integrating Genomic and Phenotypic Analyses of Autonomic Nervous System Dysfunction in a Rare Neurological Disease Cohort.....	1013
PB1889. Intraflagellar transport protein RABL5/IFT22 is required for normal visual function and retinal photoreceptor survival	1014
PB1890. Investigating SLC6A1 targeted therapeutics in <i>Drosophila</i>	1015
PB1891. Investigating the role of seryl-tRNA synthetase in mitochondrial biology and human recessive disease	1016

ASHG 2022 Annual Meeting Poster Abstracts

PB1892*. Investigating tissue specific defects in mitochondrial bioenergetics and quality control: new implications for cellular pathogenesis and therapeutic targeting in Barth syndrome	1017
PB1893. Investigating variants of uncertain significance in orofacial cleft trios.	1018
PB1894. Isolated acanthosis nigricans in a Mexican girl with a nonsense variant in <i>FGFR3</i> gene... ..	1019
PB1895. <i>ITGB8</i> is a candidate disease gene for autosomal dominant and recessive trait forms of muscular dystrophy and neurological disease.....	1020
PB1896. Joint analysis of de novo mutations in multiple traits and gene-expression data improves statistical power for the prioritization of genes associated with disease.....	1021
PB1897. <i>KIF1A</i> novel variant p.(Ser887Profs*64) exhibits clinical heterogeneity in a Pakistani family individuals with HSANIC	1022
PB1898. Large-scale genomic analyses of 150 consanguineous kindreds from the Middle East and North Africa identify novel neurodevelopmental disease mechanisms	1023
PB1899. Large-scale zebrafish-based functional analysis of genes associated with neurodevelopmental disorders.....	1024
PB1900. Leveraging orthogonal sequencing and optical mapping technologies for the precision diagnosis of neurodevelopmental disorders in a Middle Eastern family based cohort.	1025
PB1901*. LIM Homeobox 1 gene variants contribute to Mayer-Rokitansky-Küster-Hauser syndrome	1026
PB1902. Lisch epithelial corneal dystrophy is caused by heterozygous loss-of-function variants in <i>MCOLN1</i>	1027
PB1903. Lissencephaly spectrum disorders: Clinical, radiological molecular results of 70 Egyptian families.....	1028
PB1904. Local versus systemic control of bone and skeletal muscle mass by targeting components of the transforming growth factor- β signaling pathway using pharmacological and transcriptional approaches.....	1029
PB1905*. Long Read Sequencing and Expression Studies of <i>AHDC1</i> Deletions in Xia-Gibbs Syndrome Reveal a Novel Genetic Regulatory Mechanism.....	1030
PB1906. Long term prognosis in two adult patients with malonyl CoA decarboxylase deficiency ...	1031
PB1907. Long-term clinical course of Heyn-Sproul-Jackson syndrome.....	1032
PB1908*. Loss of adipocyte phospholipase gene <i>PLAAT3</i> causes lipodystrophy with neurological features due to inactivated arachidonic acid-mediated PPAR γ ;;; signaling	1033
PB1909*. Loss of C-terminal Mediator Complex subunit-11 impairs fetal brain development and cause severely progressive neurodegeneration	1034
PB1910*. Loss of FOCAD, operating in the SKI mRNA surveillance pathway, causes a pediatric syndrome with liver cirrhosis.....	1035
PB1911. <i>LYRM7</i> mutations causing mitochondrial complex 3 deficiency nuclear type 8 (MC3DN8)in a Five-year-old boy without cavitating leukoencephalopathy	1036
PB1912. <i>MED13L</i> knockout in cerebral organoids leads to a shifted developmental program through abnormal cis-regulatory element activation.....	1037

ASHG 2022 Annual Meeting Poster Abstracts

PB1913. Metatranscriptomics detects emerging multidrug resistant <i>Candida auris</i> in a family with a mild TP63 associated ectodermal dysplasia.....	1038
PB1914. Metatranscriptomics reveals association of α -, β -, and γ -HPVs with typical epidermodysplasiaverruciformis in a large cohort of patients with CIB1, TMC6, or TMC8 mutations	1039
PB1915. Modeling dystroglycanopathy in zebrafish through CRISPR-Cas9-mediated knockout of <i>pomgnt2</i>	1040
PB1916. Molecular and behavioral characterization of a novel mouse model of Snyder-Robinson Syndrome: A path towards therapeutic development	1041
PB1917. Molecular and clinical review of 95 Polish pediatric patients with the clinical suspicion of Alport syndrome.	1042
PB1918. Molecular Genetic Defects of Hereditary Angioedema in Saudi Arabia.	1043
PB1919*. Molecular insights into the pathogenesis of Chediak-Higashi syndrome and the biology of Lysosomal Trafficking Regulator.	1044
PB1920. Monoallelic variation in the DExH-box helicase DHX9, a product of the <i>DHX9</i> gene paralog, perturbs neurodevelopment & causes peripheral nerve axon degeneration.....	1045
PB1921. Monogenic Familial Hypercholesterolemia in the eMERGE Network: Penetrance and Associated Coronary Heart Disease Risk	1046
PB1922. Multi-omic approach identifies a novel non-coding deletion at Xq28 in a patient with X-linked primary immunodeficiency	1047
PB1923. Mutational analysis of a Romanian Bardet-Biedl syndrome cohort revealed an overabundance of causal <i>BBS12</i> variants.....	1048
PB1924. Mutation-specific pathophysiological mechanisms of <i>AFF3</i>	1050
PB1925. Myelomeningocele: targeted search for novel variants in Mexican families.....	1051
PB1926. New <i>trans</i> -modifier genes in the most common monogenic eye disorder, ABCA4/Stargardt disease.	1052
PB1927. Novel clinical entity: Autosomal recessive missense GPC6-related skeletal dysplasia (ARM-GSD)	1053
PB1928. Novel homozygous nonsense mutation of <i>MLIP</i> and compensatory alternative splicing in a late-onset distal myopathy.	1054
PB1929. Novel homozygous variant in <i>TFAP2B</i> in a colombian family with Char syndrome and distinctive neurobehavioral features.	1055
PB1931. Novel locus identification for primary congenital glaucoma in a Pakistani family	1056
PB1932. Novel variants expand understanding of <i>PPP1R13L</i> - related cardio-cutaneous syndrome	1057
PB1933. Novel variants in the <i>PKD1</i> and <i>PKD2</i> genes in German patients with autosomal dominant polycystic kidney disease.....	1058
PB1934. One family, two disorders, three affected: a diagnostic odyssey highlighting the importance of WES as first-tier testing in patients with a specific phenotype but a consanguineous background....	1059

ASHG 2022 Annual Meeting Poster Abstracts

PB1935*. Osteogenically differentiated human fibroblasts - an alternative model to study bone diseases	1060
PB1936. Patients with biallelic GGC repeat expansions in <i>NOTCH2NLC</i> exhibiting a typical neuronal intranuclear inclusion disease phenotype.....	1061
PB1937. Patients with primary familial brain calcification due to variants in <i>SLC20A2</i> are not commonly diagnosed with Parkinson's disease.....	1062
PB1938. PHACE syndrome: New strategies to identify the disease mechanism for this unsolved recognizable condition	1063
PB1939. Phenotypic Characterization of 127 Individuals with <i>MECP2</i> Duplication Syndrome.	1064
PB1940. Phenotypic variation of outcomes for tail-domain <i>KIF5A</i> related variants.....	1065
PB1941. PhenylEc: Novel genotypes and nutritional management in phenylketonuria patients from Ecuador.	1066
PB1942. Polygenic risk scores and protein language models predict phenotype severity in individuals with rare clinical pathogenic variants	1067
PB1943. <i>PPP2R5D</i> -related Neurodevelopmental Disorder can be inherited from a mildly affected parent.	1068
PB1944*. Precision diagnostics for GLUT1 disorder using deep mutational scanning	1069
PB1945. Pre-clinical studies in induced Pluripotent Stem Cell (iPSC) lines with <i>SORD</i> mutations linked to a recessive neuropathy	1070
PB1946. Predictive modeling to define the locus heterogeneity of tRNA synthetase-related peripheral neuropathy.....	1071
PB1947*. Probing variants associated with neurodevelopmental disorders by studying social behaviors in <i>C. elegans</i>	1072
PB1948. Proposal of intervention of the IL-10 rs1800872 (-592 C>A) as a modulator in the cardiovascular variant of Fabry disease	1073
PB1949. Pulmonary function and structure in young adults with types III, IV and VI osteogenesis imperfecta.	1074
PB1950. Rapid Turnover Skeletal Disease from In-frame Tandem Duplication of Exons 4-9 in <i>TNFRSF11A</i> Encoding RANK	1075
PB1951. Rare <i>MECP2</i> variant associated with the development of Rett syndrome: a case report. ...	1076
PB1952. Rare variants in syndromic ciliopathy genes as novel causes of isolated renal disease in adults	1077
PB1953. Rate of Deleterious Copy Number Variants Similar in Early Onset Psychosis and Autism Spectrum Disorders: Implications for Clinical Practice.....	1078
PB1954*. RBM12 Variants Lead to a Novel Neurodevelopmental Syndrome With Atypical Genotype-Phenotype Correlations	1079
PB1955. Reanalysis of whole exome sequencing data of 50 Iranian families with hereditary hearing loss	1080

ASHG 2022 Annual Meeting Poster Abstracts

PB1956. Recessive variants in <i>SLC9A1</i> cause a syndrome of cerebellar ataxia, amelogenesis imperfecta and variable sensorineural hearing loss.	1081
PB1957*. Recessive variants in the C-terminal domain of <i>UFSP2</i> are associated with severe spondyloepimetaphyseal dysplasia.	1082
PB1958. Reconsidering <i>SIX5</i> as a monogenic cause for Congenital Anomalies of the Kidney and Urinary Tract (CAKUT) and Branchio-oto-renal syndrome (BOR).....	1083
PB1959. Report of a novel <i>LAMB3</i> variant and an overview of <i>LAMB3</i> variants associated with junctional epidermolysis bullosa.....	1084
PB1960. Report of <i>PLP1</i> Gene Mutation in Pelizaeus-Merzbacher Disease	1085
PB1961. Rewiring our understanding of the enteric nervous system: Identification of diverse developmental lineages of enteric neurons and cell-state specific roles in maintenance across aging.	1086
PB1962*. <i>RFC1</i> expansions in Italian population: Molecular analysis, structure/size determination and clinical features of patients with CANVAS/spectrum phenotypes.	1087
PB1963. RNF220, RING finger E3 ubiquitin ligase, is required for the generation of p2-derived V2 interneurons and its loss-of-function causes an interneuronopathy with a broad spectrum of neurodevelopmental phenotypes.....	1088
PB1964*. Rothmund Thomson syndrome with congenital cataracts and severe growth restriction is associated with a not previously described gene in an autosomal recessive inheritance pattern in seven families.....	1089
PB1965. Screening for Genetic Modifiers of <i>MED12/kto</i> Using Naturally Occurring Variation in <i>Drosophila melanogaster</i>	1090
PB1966. Seizures in a patient with Autosomal recessive Intellectual disability type 5.....	1091
PB1967*. Sequencing analysis of nonsyndromic cleft lip / palate in Brazilian patients reveals CTNNA1, ESRP1 and PRICKLE1 as novel candidate genes.	1092
PB1968. Severe ABCA4-retinal dystrophy caused by homozygosity for a novel double mutated allele.	1093
PB1969. Shared and divergent transcriptional dysregulation is present across brain regions and cell types in adult <i>Chd8</i> haploinsufficient mice.....	1094
PB1970. Siblings with hypophosphatemic rickets and homozygous <i>CYP27B1</i> variants	1095
PB1971. Six patients with genetic disorders with atypical symptoms: how to perceive them	1096
PB1972. Slc9a6 mutation causes Purkinje cell loss and ataxia in the <i>shaker</i> rat.....	1097
PB1973. Splicing noise is variable across human introns, tissues and age and modelling its characteristics can improve our understanding of age-related diseases of the human brain.....	1098
PB1974*. <i>SRSF1</i> haploinsufficiency is responsible for a syndromic developmental disorder with intellectual disability and variable marfanoid habitus	1099
PB1975. Subtle genomic duplication on Chromosome 6: a hypothesis for a causative mechanism..	1101
PB1976. Sunken symptoms of <i>SRCAP</i> mutation: A case report of atypical phenotype of Floating Harbor Syndrome.....	1102

ASHG 2022 Annual Meeting Poster Abstracts

PB1977. Systematic bioinformatic analysis promotes a missense variant in <i>XK</i> as potential cause of X-linked intellectual disability.	1103
PB1978*. Systematic identification of disease-causing near-coding variants in 29,759 individuals with rare disease.	1104
PB1979. Taking the <i>CPLANE INTU</i> the light: further defining the orofaciodigital syndrome phenotype association with biallelic <i>INTU</i> alterations.	1105
PB1980. Targeted deep sequencing analyses of long QT syndrome in a Japanese population.	1106
PB1981. <i>TCEALI</i> loss-of-function results in an X-linked dominant neurological syndrome and drives the neurological disease trait in Xq22.2 deletion.	1107
PB1982. Temtamy Preaxial Brachydactyly Syndrome: penetrance in heterozygotes for a <i>CHSY1</i> missense variant.	1109
PB1983. TGF β signaling is important for spermatogenesis.	1110
PB1984. The benefits and yields of whole exome sequencing data reanalysis in Iranian undiagnosed patients with intellectual disability.	1111
PB1985. The ClinGen Kidney Disease Clinical Domain Working Group: Defining solutions to the challenge of nephropathy ontology and nomenclature.	1112
PB1986. The genetic and phenotypic landscape of <i>STUB1/TBP</i> -associated disorders: Two genes and three patterns of inheritance sharing a continuous neurodegenerative phenotypic spectrum.	1113
PB1987. The genetic background in a group of 155 Polish pediatric patients with unexplained epilepsy.	1114
PB1988. The genetic landscape of <i>ATP7B</i> provides insights into Wilson Disease variant penetrance and prevalence.	1115
PB1989. The genetic landscape of neurodevelopmental disorders in a large cohort of multiplex consanguineous families from Turkey.	1116
PB1990*. The impact of NGS on rare and complex epilepsies: identification of novel causative variants for an efficient diagnostic pathway and personalized treatments.	1117
PB1991. The mechanism and current concepts and future research direction of rare hereditary disease White Sponge Nevus.	1118
PB1992. The Spectrum of PLEC Sequence Variants and Related Plectinopathies Including Novel Association with Epidermolysis Bullosa Pruriginosa.	1119
PB1993. Three cases of (likely) pathogenic <i>FGF10</i> variants: Examples of phenotypic variability in Lacrimoauriculodentodigital syndrome.	1120
PB1994. Three new founder Bukharian Jewish mutations in <i>COL4A4</i> are causative of autosomal dominant and recessive Alport syndrome.	1121
PB1995. TMEM161B mediates radial glial scaffolding in neocortical development.	1122
PB1996*. Toward precision medicine in Neurodevelopmental disability: identifying phenotypic differences in individuals with global developmental delay.	1123
PB1997. Transcriptional response of <i>Fundulus heteroclitus</i> embryos exposed to carbaryl.	1124
PB1998. Translated mutant <i>DSPP</i> mRNA expression level impacts the severity of dentin defects. .	1125

ASHG 2022 Annual Meeting Poster Abstracts

PB1999. Truncating variants in <i>RFC1</i> cause cerebellar ataxia, neuropathy vestibular areflexia syndrome (CANVAS).....	1126
PB2000. Tumor burden among members of the DiscovEHR cohort harboring deleterious constitutive variants in <i>NFI</i>	1127
PB2001. Two novel compound heterozygous variants identified in a new case of classic Donohue syndrome and literature review.....	1128
PB2002. Understanding the role of <i>TTC5</i> in the developing mouse brain: A novel null <i>TTC5</i> mouse model.	1129
PB2003. Using artificial intelligence driven NGS re-analyses followed by full genome profiling (HiFi-GBS) for unsolved pediatric white matter disorders	1130
PB2004. Using genetic relatedness within biobanks to understand genetic architecture of Retinitis Pigmentosa.....	1131
PB2005. Using remote methods to assess communication in Phelan-McDermid Syndrome.....	1132
PB2006. Variable expressivity in a four-generation ACDMPV family with a noncoding hypermorphic SNV in <i>trans</i> to the frameshifting <i>FOXF1</i> variant.....	1133
PB2007. Variants of <i>LRP2</i> , encoding a multifunctional cell surface endocytic receptor, are associated with hearing loss and retinal dystrophy	1134
PB2008. Vici syndrome in Israel: Clinical and molecular insights	1135
PB2009. Whole exome sequencing identifies novel variants in familial myoclonic epilepsy in Mali.	1137
PB2010. Whole exome sequencing in intellectual disability patients identifies de novo mutations in <i>KCNB1</i> , <i>PPP1R3F</i> , <i>SHANK2</i> , and <i>SYNGAP1</i> genes	1138
PB2011*. Whole exome sequencing reveals known and candidate genes in non-syndromic hearing impairment in Mali	1139
PB2012. Whole exome sequencing will miss Congenital Myotonic Dystrophy in the child on an asymptomatic mother.....	1140
PB2013. Whole genome sequencing identifies biallelic <i>SCLT1</i> splice-region variants in a child with a suspected ciliopathy: Delineating the <i>SCLT1</i> related ciliopathy spectrum.	1141
PB2014. Whole-Exome Sequencing of a French Canadian Cohort Reveals New Candidate Founder Mutations	1142
PB2015. Whole-exome sequencing of Pakistani consanguineous families identified variants in novel and known genes of severe intellectual disability.....	1143
PB2016. Whole-genome sequencing for comprehensive genetic modifier identification in <i>CMT1A</i>	1144
PB2017. A conditional knock-in preclinical mouse model for the <i>TRPV4</i> skeletal disorders.....	1145
PB2018. A cross-species approach using an <i>in vivo</i> evaluation platform in mice demonstrates that sequence variation in the human <i>RABEP2</i> gene modulates ischemic stroke outcomes.....	1146
PB2019. A drug repurposing screen to identify therapies for the rare disease DPAGT1-CDG	1147

ASHG 2022 Annual Meeting Poster Abstracts

PB2020. A Phase 1, Open-Label, Dose-Escalation Study to Evaluate the Safety and Efficacy of HMI-103, a One-Time Phenylalanine Hydroxylase Gene-Editing Vector in Adult Participants with Classical PKU Due to PAH Deficiency	1148
PB2021. A Structural Screen Approach and Molecular Simulation Identifies Potential Ligands Against the K700E Hot Spot Variant and Functional Pockets of SF3B1 to modulate splicing in Myelodysplastic Syndrome.....	1149
PB2022. AAV delivery of <i>ELPI</i> exon-specific U1 snRNA rescues retinal degeneration in a mouse model of familial dysautonomia.	1150
PB2023. Adjunct treatment with glycogen synthase (GYS1) antisense oligonucleotides and Enzyme replacement therapy (ERT) reduced glycogen in the Pompe disease mouse model.....	1151
PB2024. An allele- discriminative approach for efficient and specific CRISPR/Cas9 based gene therapy of Late Onset of Alzheimer'sdisease.....	1152
PB2025. An allele specific strategy for targeting mutant senataxin in patients with amyotrophic lateral sclerosis type 4.....	1153
PB2026*. An alternate translational start site allows 5' nonsense variants to escape NMD so that readthrough compound ELX-02 and modulators can restore CFTR function.	1154
PB2027*. An <i>in vivo</i> screen identifies small molecule modulators of retinitis pigmentosa and the ER stress response.....	1155
PB2028. Antisense oligonucleotide-based exon skipping therapeutic strategy for Cohen syndrome.	1156
PB2029. Characterization of novel mouse model with 35 bp deletion in <i>Champ1</i> gene and exploration of AAV gene therapy to correct syndromes arising from mutations in the <i>CHAMP1</i>	1157
PB2030. Comparison of AAVmyo and MyoAAV as promising vectors to deliver gene therapy in the VCP ^{R155H} KI mouse model.....	1158
PB2031. Deep phenotyping APOC3 Knockouts in a population with high consanguinity	1159
PB2032. Development of an AAV gene therapy for GNE myopathy: AAV9-CK8e-SV40- <i>hGNE1</i> -V5 shows robust GNE expression in mouse muscle tissue.	1160
PB2033. Dkk1 inhibition normalizes limb phenotypes in a mouse model of <i>Fzd2</i> associated omodysplasia Robinow syndromes.....	1161
PB2034. Down-regulation of <i>SCN8A</i> as treatment for developmental and epileptic encephalopathy	1162
PB2035. Drug repositioning network in rare and intractable diseases based on drug target gene analyses.....	1163
PB2036. Dynamin-2 interacts with SPEG and its reduction rescues the skeletal myopathy of SPEG-deficient mouse model	1164
PB2037. ECLIPSE, an automated CRISPR platform for the large-scale generation of cell models for the iPSC Neurodegenerative Disease Initiative (iNDI)	1165
PB2038. Effect of respiratory resistance exercise done remotely on respiratory function in individuals with familial myopathy.....	1166
PB2039. Efficacy and safety of elamipretide in subjects with primary mitochondrial disease resulting from pathogenic nuclear DNA mutations (nPMD) is investigated through this phase 3 study design.	1167

ASHG 2022 Annual Meeting Poster Abstracts

PB2040. Evaluation of Ataluren efficacy in fibroblasts from Neurofibromatosis Type 1 patients with nonsense mutation.....	1168
PB2041. Generation of eight human induced pluripotent stem cell lines from immortalized lymphoblastoid cells	1169
PB2042. Genetic subgroup learnings is presented from the MMPOWER-3 trial: elamipretide improves six-minute walk test in individuals with mtDNA replisome disorders.	1170
PB2043. Genome-wide CRISPR/Cas9 Screening Reveals ATR as a Therapeutic Target That Overcomes Osteosarcoma Chemoresistance.....	1171
PB2044. Genotyping analysis in patients with retinitis pigmentosa due to mutations in the <i>RHO</i> , <i>PDE6A</i> , or <i>PDE6B</i> gene: the PHENOROD2 study.....	1172
PB2045. Harmonization of phenotype ontologies and identification of novel gene/drug-disease associations with Natural Language Processing	1173
PB2046. Harnessing the power of natural variations of the human genome to support the discovery of new therapeutics: A systematic review	1174
PB2047. Implementing a Genetics-First Strategy in Clinical Development for Rare Brain Diseases	1175
PB2048. Interrogation of the role of androgens and their antagonists in the pathogenesis and treatment of Vascular Ehlers-Danlos Syndrome.....	1176
PB2049. Learnings from a randomized clinical trial with ARID1B patients	1177
PB2050. Leveraging Machine Learning to Improve Guide RNA Design for Efficient and Selective ADAR-Mediated RNA Editing.....	1178
PB2051. Localization of KK13 peptide and their potential against ALS.....	1179
PB2052*. Long-term efficacy and safety of elamipretide in patients with Barth syndrome is presented through the 192-week open-label extension results of TAZPOWER.....	1180
PB2053. New class of anaplerotic compounds ameliorates substantial loss of lysine succinylation in propionyl-CoA carboxylase deficient cells.....	1181
PB2054*. Novel ASO based treatment shows promise in treating CMT2S, preclinical data	1182
PB2055. Novel siRNA-based therapeutic approach for Megacystis Microcolon Intestinal Hypoperistalsis Syndrome (MMIHS).....	1183
PB2056*. Personalized RNA interference approach as a nanotherapy for Crouzon syndrome: design of allele-specific siRNAs targeting FGFR2 mutant allele delivered by recombinant human nanoferritin.	1184
PB2057. Pre-clinical antisense oligonucleotide treatment of CMT2E in a human induced pluripotent stem cell (iPSC)-derived motor neuron model.....	1185
PB2058*. RNA editing of founder nonsense mutations causing inherited retinal diseases using site-directed endogenous adenosine deaminase acting on RNA.....	1186
PB2059. Sustained efficacy and safety up to 3.5 years in adults with glycogen storage disease type Ia (GSDIa): results from a phase 1/2 clinical trial of DTX401, an AAV8-mediated, liver-directed gene therapy.	1187
PB2060*. Testing gene therapy for achromatopsia-like visual deficits in nonhuman primates affected by naturally occurring missense mutations in <i>PDE6C</i>	1188

ASHG 2022 Annual Meeting Poster Abstracts

PB2061. The Jackson Laboratory Center for Precision Genetics: Preclinical research for rare diseases	1189
PB2062. Two years of venglustat combined with imiglucerase shows continued positive effects on neurological features and brain connectivity in adults with Gaucher disease type 3.....	1190
PB2063. Venglustat, a novel, investigational, brain-penetrant glucosylceramide synthase inhibitor, for Gaucher disease type 3: phase 3 LEAP2MONO trial design	1191
PB2064. A novel human mitochondrial donating cell type promotes maturation of GV oocytes.....	1192
PB2065. Assessing the accuracy of variant detection and coverage bias in mitochondrial DNA (mtDNA) after whole genome amplification of single fibroblast cells with known mtDNA heteroplasmic mutations to effectively optimize preimplantation genetic testing of mtDNA related disorders.....	1193
PB2066. Assessing the magnitude of clinical utility of preconception expanded carrier screening for prevention of neurodevelopmental disorders	1194
PB2067. Assisted reproductive technology and DNA methylation at imprinted genes	1195
PB2068. Capacity Building: Exploring the Use of Newborn Screening Data to Understand Neurodevelopmental Outcomes	1196
PB2069. Cell free fetal (cff) DNA assessment using Y- chromosome specific sequences and amplicon NGS of targeted SNPs for possible implementation in non-invasive prenatal testing of common chromosomal aneuploidies.....	1197
PB2070. Cell-informed atlas of gene, isoform, and splicing regulation in the developing human brain identifies candidate causal variants for neuropsychiatric GWAS.....	1198
PB2071. Characterization of the prenatal presentation of novel truncating variants in <i>ANKRD11</i>	1199
PB2073*. Comprehensive preimplantation genetic testing for monogenetic disorders (PGT-M) and aneuploidy (PGT-A) using whole genome sequencing and haplotyping.....	1200
PB2074. Contribution of BBS9 isoforms to the etiopathogenesis of non-syndromic craniosynostosis	1201
PB2075. <i>De novo</i> mutations disturb early brain development more frequently than common variants in schizophrenia	1202
PB2076. Developmental programming of vitamin b12 deficiency model in zebrafish.....	1203
PB2077. Developmental trajectory analysis of differentiating mouse sensory interneurons to further improve stem-cell differentiation protocols.....	1204
PB2078. Discordances among NIPT, QF-PCR and cytogenetic results in prenatal cases.....	1205
PB2079. DNA damage and antioxidant capacity as prenatal markers of placental epigenetic aging clock.....	1206
PB2080. Down Syndrome therapeutics: The integrated stress response and restoration of protein synthesis.....	1207
PB2081. Explore SOX9 haploinsufficiency in neural stem cells.	1208
PB2082. Exploring roles of distal enhancers in <i>SOX9</i> haploinsufficiency in sex determination	1209
PB2083. Fetal genetic autopsy: A five-year retrospective audit monocentric audit.....	1210

ASHG 2022 Annual Meeting Poster Abstracts

PB2084. Gaucher disease: systematic review.....	1211
PB2085. Genetic Characterization of VACTERL Association-Like Congenital Malformations.....	1212
PB2086*. Genetic deconvolution of fetal and maternal cell-free DNA in maternal plasma enables next generation non-invasive prenatal screening.....	1213
PB2087. Genetic determinants of gestational length in Indian women: A two-stage genome-wide association study.....	1214
PB2088*. Genome-wide analysis of preeclampsia identifies novel loci and loci previously associated with blood pressure traits.....	1215
PB2089. Heteroplasmy dynamics and segregation in pre-implantation embryos: Rare insights from complete mitochondrial genome sequencing of clinical trophectoderm biopsy.....	1216
PB2090. High-risk <i>APOL1</i> genotypes and infant growth restriction associations in Black women affected by preeclampsia.....	1217
PB2091. Host genomics and human milk exposure are associated with sex-specific changes in the gut microbiota of infants with asthma in the CHILD Cohort Study.....	1218
PB2092. Human engineered heart tissues uncover <i>MYH7</i> alleles associated with a frail sarcomere and ventricular noncompaction.....	1219
PB2093. Identifying genes associated with preeclampsia using whole exome sequencing data from participants in the Penn Medicine BioBank.....	1220
PB2094. Improving rare conditions diagnostic rates by standardizing practice and offering preclinical testing.....	1221
PB2095. Investigating uptake and impact of genetic and genomic evaluation following a perinatal demise.....	1222
PB2096. Knowledge and attitude of pregnant women in the Kingdom of Saudi Arabia toward Noninvasive prenatal testing: A single center study.....	1223
PB2097. Large-scale genome-wide association study meta-analysis of Hyperemesis Gravidarum confirms the nausea and vomiting hormone gene <i>GDF15</i> is the greatest genetic risk factor and identifies additional risk loci.....	1224
PB2098. Lethal genes and Mendelian disorders.....	1225
PB2099. Leveraging duplex DNA sequencing to investigate germline mutation rates in the context of sperm selection and paternal age.....	1226
PB2100*. Metabolomic analysis reveals altered omega-6 fatty acid processing as a marker of <i>FMR1</i> premutation carriers with FXPOI compared to premutation carriers without ovarian insufficiency.....	1227
PB2101. Molecular dysregulation in <i>Sox2</i> -expressing pituitary stem cells lacking <i>Prop1</i>	1228
PB2102. Molecular Epidemiological Status of Group B Streptococcus in Ile Ife South Western Nigeria.....	1229
PB2103. Next-Generation Sequencing based pre-implantation genetic Testing in Nepal.....	1230
PB2104. <i>Nlrp2</i> is a Maternal Effect Gene Required for both Placental and Embryonic Development.....	1231

ASHG 2022 Annual Meeting Poster Abstracts

PB2105. Non-invasive prenatal testing (NIPT): A reliable accurate prenatal non-invasive diagnosis setting in Nepal	1232
PB2106. Ovarian subfertility genetic variants in females with unexplained infertility.	1233
PB2108*. Phenome and genomewide recessive scan of coding variants reveals novel associations for female infertility.....	1234
PB2109. Phenomic and genomic analyses of endometriosis in the All of Us Research Program.	1235
PB2110. Placenta fraction of maternal blood cell-free RNA associates with pre-eclampsia during pregnancy.....	1236
PB2111. Preeclampsia risk is influenced by apolipoprotein L1 genotype and micronutrient deficiencies.	1237
PB2112. Prenatal and postnatal changes in chromatin accessibility and gene expression in mouse gut development for understanding Hirschsprung disease (HSCR).....	1238
PB2113. Prenatal cytogenomic microarray analysis (CMA) reference model.	1239
PB2114. Prenatal diagnosis of recurrent severe form of infantile galactosialidosis exemplifying a pitfall in parental carrier screening.....	1240
PB2115. Prenatal diagnostic testing in cases of cystic hygroma or increased nuchal translucency: A single center experience.	1241
PB2116. Prenatal onset of Spondyloepimetaphyseal Dysplasia SIK3 Type (SEMDK).....	1242
PB2117. Rapid testing of congenital diaphragmatic hernia novel variants using CRISPR/Cas9 mouse embryo editing.	1243
PB2118*. Reconsidering the “advanced” maternal age for invasive prenatal testing	1244
PB2119. Recurrent constitutional chromosome 5 inversion.....	1245
PB2120. Rescue of limb hemorrhage in a cohesinopathy mouse model with severe limb reduction.	1246
PB2121. Sex-specific variation in the commonly shared mothers’ milk and infant gut microbiota are associated with sex disparity in childhood asthma risk.	1247
PB2122. Single-nucleus chromatin accessibility and RNA sequencing reveal impaired brain development in prenatally e-cigarette exposed neonate’s brain.....	1248
PB2123*. Sperm mosaicism predicts transmission of <i>de novo</i> mutations to human blastocysts.	1249
PB2124. Successful implementation of a short turn-around time prenatal exome sequencing process in a perinatal centre.	1250
PB2125. Tensor decomposition on electronic health records identifies latent subtypes of preterm birth that associate with polygenic risk scores	1251
PB2126. The abnormal expression of has-miR-196b-5p, FAS, and FAS-AS1 genes in Iranian patients with recurrent pregnancy loss	1252
PB2127. Understanding caudal developmental abnormalities using single-nucleus multi-omics data from wild type and Danforth's short tail mouse E9.5 tailbuds.....	1253
PB2128. Understanding the burden of clonal and non-clonal sperm mosaicism by bulk and single-cell genome sequencing.....	1254
PB2129. Utility of preconception carrier screening in couples with unexplained infertility.....	1255

ASHG 2022 Annual Meeting Poster Abstracts

PB2130. Variant reclassification in expanded carrier screening brings challenges for clinical management: experience of a Hong Kong prenatal diagnosis centre.	1256
PB2131. Whole genome sequencing and analysis of 4,053 individuals in trios and mother-infant duos from the Born in Guangzhou Cohort Study	1257
PB2132. Whole-genome sequencing analysis in families with pregnancy loss	1258
PB2133. "Grown up" minors: Recontact after withdrawal or attrition	1259
PB2134. "Frankly, I thought this was one and done": Patient and provider perspectives on return of reclassified variants in clinical oncology	1260
PB2135. "I just wanted more": Hereditary cancer syndromes patients' perspectives on the utility of circulating tumour DNA testing for cancer screening	1261
PB2136*. A New System for Triaging Patients in a Medical Genetics Clinic Saves Time and Improves Quality.	1262
PB2137. A Sociotechnical Analysis of Returning Genomic Informed Risk Assessments in Primary Care Pediatrics	1263
PB2138. A systematic review of health policy guiding the identification, analysis, and management of genomic secondary findings.....	1264
PB2139. Access to Clinical Genetics Service and Testing for Patients with Retinal Dystrophies: a Quality Improvement Study.....	1265
PB2140. Advancing Health Equity in Genomics: Reflections and Recommendations for Future Research Directions from an NHGRI Workshop.....	1266
PB2141. Anxiety is a prominent neurobehavioral feature in Kabuki syndrome that significantly affects quality of life.....	1268
PB2142. Assessing the performance of the Clinician-reported Genetic testing Utility InDEx (C-GUIDE): Further evidence of inter-rater reliability	1269
PB2143. Assessment of cancer patients using a distress thermometer to evaluate perceived and calculated stress in order to identify individuals requiring additional psychological support.	1270
PB2144. Assessment of reading comprehension levels of interpretation sections of genetic testing reports	1271
PB2145. Awareness about rare diseases through Indian cinema	1272
PB2146. Barriers to genetic testing when indicated at diagnosis for autism spectrum disorder.	1273
PB2147*. Benefits, harms and costs of newborn genetic screening for long QT syndrome: estimates from the PreEMPT Model.....	1274
PB2148. Cancer patients' experience of receiving variant of uncertain significance results: An Asian perspective.	1275
PB2149*. Challenges in incorporating polygenic risk scores in the electronic health record.	1276
PB2150. Clinical cancer and direct-to-consumer genetic test result sharing behavior among US adults: Findings from HINTS 2020	1277
PB2151. Clinical utility of returning all clinically relevant secondary findings from genomic sequencing: A systematic review	1278

ASHG 2022 Annual Meeting Poster Abstracts

PB2152. Communicating genetic and socioeconomic factors underlying racial health disparities in medical education	1279
PB2153. Community data-driven approach for generating cross-ethnic population carrier screening panel.....	1280
PB2154. Community Engagement in genetic and genomic research in a low literate setting: Challenges and perspectives	1281
PB2155. Considerations for assessing and optimizing remote participant recruitment for a pragmatic trial of polygenic risk score testing.....	1282
PB2156. Considerations for embedding genomic discovery research in clinical settings: Perspectives from medical geneticists	1283
PB2157. Cultivating a data-driven computational culture within biomedical institutions by empowering graduate students with code-based data science skills.....	1284
PB2159*. Delivery of an unfavorable genetic diagnosis: Developing a methodology framework....	1285
PB2160. Designing a Comprehensive Education Program for Participants, Providers, and Research Teams: The eMERGE Consortium Experience	1286
PB2161. Developing Recommendations for Direct-to-Consumer Genetic Testing	1288
PB2162*. Development and utility of a clinical research informatics application for participant recruitment and management for a return of results pilot trial in the Million Veteran Program	1289
PB2163. Do Automated Family Histories Significantly Improve Risk Prediction in an EHR?	1290
PB2164. Do digital tools support a diversity of patient values? A qualitative study using the Genetics Adviser.....	1291
PB2165. Early Mentoring for Medical Students in Research of Medical Genetics.....	1293
PB2166. Electronic health record-based recruitment for genomic research studies.....	1294
PB2167. Engagement and Feedback on Genetic Ancestry and Trait Results from >100,000 All of Us Participants.....	1295
PB2168*. Engaging a local community & participant advisory board in a national precision medicine study: Experiences from the New York City Consortium <i>All of Us</i> Research Program.....	1296
PB2169. Engaging Families and Clinicians to Advance Care Coordination and Improve Health Outcomes in Genetic Disease: The Long-Term Follow-Up Cares and Check Initiative.....	1297
PB2170. Ensuring participant understanding of informed consent through an automated consent chatbot for a national genome sequencing study.	1298
PB2171. Ethical considerations for genomic newborn hearing screening from the SEQaBOO (SEQuencing a Baby for an Optimal Outcome) study.....	1299
PB2172. Evaluation of out-of-province and self-pay for genetic testing in two publicly funded provincial programs in British Columbia, Canada.....	1300
PB2173. Evidence-Based Assertions of Clinical Actionability of Secondary Findings: Update from the ClinGen Actionability Working Group.....	1301
PB2174. Expanding ethics education for trainees: creating and evaluating a course on genetics, ethics, and society.	1302

ASHG 2022 Annual Meeting Poster Abstracts

PB2175. Expanding genomic testing to underserved pediatric populations.....	1303
PB2176. Exploration of a Novel Cancer Genetics Learning Format: A Pilot Self-Directed Course in Genetic Cancer Risk Assessment.....	1304
PB2177. Exploring stigma and discrimination among adults after receipt of neurodevelopmental/psychiatric genetic results in a population-based genomic screening program.	1305
PB2178. Exploring the ethical implications of genomic data sharing in cloud-based ecosystems: Views of developers and users.	1306
PB2179. Factors contributing to preimplantation genetic testing for polygenic disorders (PGT-P) screening and recommendations by reproductive medical professionals.	1307
PB2180. Family communication choices about neurodevelopmental/psychiatric genetic results: A social-network assessment of adult participants in a population-based genomic screening program.	1308
PB2181. <i>FLCN</i> variant of unknown significance complicating a kidney donation process in a family with Birt-Hogg-Dubé.	1309
PB2182. Fragile X syndrome: Experiences with receiving a genetic diagnosis and communicating genetic risk in a Cameroonian family.	1310
PB2183. Fragmented systems of care: An overview of Canadian health system care models for hereditary cancer syndromes.....	1311
PB2184. Gender medicine, Gender-Omics and Sex-Omics: What we are dealing with?.....	1312
PB2185. Genetic counselors and legal recognition: a made-for-Canada approach	1313
PB2186. Genetic Counselors' Attitudes Towards Religiously-Based Policies and Their Impact on Professional and Ethical Obligations in the Prenatal Setting.....	1314
PB2187. Genetics education in neurodevelopmental disorders for non-genetic health professionals: A scoping review.	1315
PB2188*. Genome sequencing and preferences for secondary findings in ostensibly healthy COVID-19 positive individuals: GENCOV study Canada.....	1316
PB2189*. Genomic knowledge, orientation, and empowerment in a demographically diverse population: Preliminary findings from the Texome project.....	1317
PB2190*. Healthcare provider confidence in understanding unsolicited genomic results: Insights from the eMERGE III experience.....	1318
PB2191. Healthcare providers' experiences counseling patients with results from consumer genomic testing.....	1319
PB2192. Homozygous <i>APC</i> p.I1307K Mutations in an Ashkenazi Jewish Female: Implications for Genetic Counseling and Clinical Management.....	1320
PB2193. Identifying ELSI needs in newborn screening research.....	1321
PB2194. Implementation and evaluation of digital patient-reported oncology questionnaire at the East Genomic Medicine Service Alliance	1322
PB2195. Implementing human genetics training and research at a university hospital in Nairobi, Kenya.	1323
PB2196*. Implementing public health genomics: The familial hypercholesterolemia model	1324

ASHG 2022 Annual Meeting Poster Abstracts

PB2197. Improving Community Based Genomics Research in Black Populations: A Review of Lessons Learned and Ideas for Need Based Solutions.....	1325
PB2198. Incidental findings in study participants: What is the researcher's obligation?.....	1326
PB2199. Individual attitudes towards identity, privacy, and health insights from personal genomics: Evidence from a full archival search on Twitter.....	1327
PB2200. Integrating Predictive Genetic Testing into clinical care in the Preventive Genomics Clinic	1328
PB2201. International Best Practices for Genomic Data Sharing in the Health Technology Industries	1329
PB2202*. Is it time for a paradigm shift? Design and formative evaluation of a randomized clinical trial of the sequence of genetic counseling and testing to optimize efficiency, patient empowerment and engagement, and medical adherence for cardiovascular genetic testing indications (<i>RESEQUENCE-GC</i>).....	1330
PB2203. Lessons learned from population-based biobank participants' responses to individual genomic results	1331
PB2204*. Low interest and high drop off in population genetic screening of adults: Results from a "real world" pilot study.....	1332
PB2205. Masking for whole-exome sequencing data: An analysis of the impact of masking choices in rare coding variants association tests.	1333
PB2206. Medical genetics and genomics services <i>across the lifespan</i> : Patient and referral trends over 17 years at the Maritime Medical Genetics Service.....	1334
PB2207. Newborn Screening: a multidisciplinary approach based on landscape analysis, genetic screening, whole genome sequencing, and artificial intelligence to achieve an early diagnosis of rare diseases	1335
PB2208. Non-inferiority of letter versus telephone genetic counseling for negative exome-based cancer genetic test results disclosure in a historically underserved population.....	1336
PB2209. Parents' attitudes towards return of results from genome sequencing research for their healthy children: A qualitative study	1337
PB2210*. Partnering with patients to explore the psychosocial and socioeconomic impacts of hereditary cancer syndromes.....	1338
PB2211. Patient reported utility of cancer results from genomic sequencing.	1339
PB2212. Patterns and barriers to clinical trial participation for patients with primary mitochondrial disease: A survey of patients/caregivers, physicians, and pharmaceutical industry professionals.	1340
PB2213. Personal utility in genomic medicine: what is it and can we measure it?	1341
PB2214. Population DNA screening for medically actionable conditions in young adults: The DNA Screen national pilot study.....	1342
PB2215. Predictors and Trends in Cancer Genetics Clinic Attendance Rate After the Adaptation of Telemedicine During the COVID-19 Pandemic	1343
PB2216. Preferences of parents from diverse backgrounds on genomic screening in newborns.	1344

ASHG 2022 Annual Meeting Poster Abstracts

PB2217*. Priorities to promote participant engagement in cancer genomics research: A report from the Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network.....	1345
PB2218*. Prioritizing Research in Newborn Screening: Tools from the Newborn Screening Translational Research Network.....	1346
PB2219. Psychosocial impacts in patients/parents with secondary findings reporting from exome sequencing: 1st French multicenter qualitative and quantitative study (FIND Study).	1347
PB2220. Q-A sequences in prenatal genetic counseling in Japan; conversation analytic study of NIPT and NT consultations	1349
PB2221. Raising Genetic Disease Awareness: NBSTRN Efforts to Leverage Social Media to Realize the Promise of Genomics	1350
PB2222. Real-world implementation of an efficient germline Point-of-Care Testing model in a multidisciplinary cancer center.	1351
PB2223. Rethinking general consent for stem cell-based embryo model research	1352
PB2224. Return of secondary genomic findings: Experiences of sickle cell disease research participants.....	1353
PB2225. Secondary findings in a large Pakistani cohort tested with whole genome sequencing	1354
PB2226. Shared in Rare: Engaging stakeholders to develop a shared ELSI research agenda across rare diseases.	1355
PB2227. The future of diversity, equity, and inclusion: Achieving the vision of genetic counseling for the All of Us Research Program	1356
PB2228. The Million Veteran Program - Return Of Actionable Results (MVP-ROAR) Study: Participant Baseline Characteristics.....	1357
PB2229. The Quebec Participatory cohort, a citizen science projet for direct-to-consumer genetic testing participants	1358
PB2230*. The role of oncology nurses in the genetic counseling workflow: A literature scan	1359
PB2231. The SeDeN-p2 study : Perceptions of French health professionnals on the extension of newborn screening with or without genetics as a first-line test.	1360
PB2232. The Texome Project: Initial enrollment in a rare and undiagnosed disease program for underserved communities in Texas.....	1361
PB2233. Transforming genetics service delivery: Genetics providers' perspective on integrating digital tools into clinical practice.	1362
PB2234*. Tribal community perspectives on genomics research and data sharing: a mixed-methods study.....	1363
PB2235. Understanding views towards gene editing in Switzerland.....	1364
PB2236. Users' evaluation of artificial intelligence-based diagnostic program in pediatric rare diseases	1365
PB2237*. Using variant databases to estimate disease prevalence for rare recessive disease.....	1366
PB2238. Utilizing chatbots for family communication: Uptake and engagement among familial hypercholesterolemia probands from a prospective, pragmatic trial.	1367

ASHG 2022 Annual Meeting Poster Abstracts

PB2239*. Validation of automated electronic health record (EHR) data capture of hereditary breast and ovarian cancer and Lynch syndrome phenotypes	1368
PB2240. Voices from the online right: A qualitative study of race and genetics discourse.	1369
PB2241. Whole-exome sequencing in Puerto Rican patients: approach to diagnosis in complex traits.	1370
PB2242. Your genes & you: patient motivations regarding secondary findings from genomic studies	1371
PB2243. A calibrated automated patch clamp assay can interpret any VUS in <i>KCNH2</i>	1372
PB2244. A Case of Concurrent 3p26.3-p25.3 Deletion and 9q34.2-q34.3 Duplication in Monozygotic Twin Males: A Case Report.	1373
PB2245. A combination of exome sequencing and optical genome mapping unveils a dual molecular diagnosis in a case with an unknown neurodevelopmental disorder.	1374
PB2246. A framework designed to improve the evaluation and reporting of incidental findings: Implementation and experience of a clinical laboratory in the first 15 months.	1375
PB2247. A heterozygous de novo VUS (p.M130I) in candidate gene <i>PSMC5</i> in a male with autism, mild developmental delay, speech deficiency, dysmorphic features, hypospadias and chordee.	1376
PB2248. A multiomics approach to resolving small supernumerary marker chromosomes	1377
PB2249. A multiple congenital anomalies newborn with coexistence of extra der(22) chromosome marker and balanced t(11;22) maternally inherited	1378
PB2250. A national initiative working toward equitable translational genomic research for Canadian Indigenous families: Early findings indicate the need to reduce uncertainty in variant interpretation.	1379
PB2251*. A new neurodevelopmental disorder with microcephaly and neural tube defect Loss of <i>NARS1</i> leads to microcephaly and neural tube defects	1380
PB2252. A novel <i>FAME1</i> repeat configuration in a European family identified using a combined genomics approach.	1381
PB2253. A rare association of congenital hypohydrotic ectodermal dysplasia, secondary hypogonadism, and severe iron deficiency anemia.	1382
PB2254. A recurrent <i>AFF3</i> pathogenic variant associated to KINSSHIP syndrome in a Mexican patient	1383
PB2255. A robust NGS detection workflow for the <i>IKBKG</i> gene.	1384
PB2256. A variant interpretation framework based on probabilistic graphical modeling that provides continuous high-resolution estimates of variant pathogenicity with improved interpretation accuracy over current discrete five-category frameworks	1385
PB2257. Accurate Total and Allele-Specific Copy Number Quantification of the Cardiovascular Disease-Associated Lipoprotein (a) Kringle-IV Type-2 Tandem Repeat Using Illumina Sequencing	1386
PB2258. Advancing Biobanking in the Genomics Era: Development of Molecular Omics Approaches for Interrogating the Diversity of Population Genetics for Early and Late-Stage Human Disease Prevention	1387

ASHG 2022 Annual Meeting Poster Abstracts

PB2259. Analysis of mRNA and protein expression levels of MECP2 in missense mutant iPSC and lymphoblast lines against isogenic controls.....	1388
PB2260. Anthropometric parameters in males with chromosome 9p microdeletion syndrome in the first years of life: preliminary evidence.....	1389
PB2261*. B-allele frequency based approach to detecting absence of heterozygosity using optical genome mapping.....	1390
PB2262. Beyond NGS: RNA-based functional evaluation improves ACMG classification of VUS with potential Loss of function	1391
PB2263. Challenges of characterisation of medically relevant tandem repeats in whole genome sequencing data.....	1392
PB2264. Child with Rhabdomyosarcoma and unusual features: suspected Rasopathy.....	1393
PB2265. Chromosomal and molecular investigations of Senegalese children bearing disorders of sex development.....	1394
PB2266*. Chromosome conformation capture to detect structural variation in medical diagnostics	1395
PB2267. Clarifying genetic disease with long range sequencing: Tipping the balance for structural variant detection in SMAD3.	1396
PB2268. Clinical description of patients with CNV in the 15q11.2 region: What do we know?	1397
PB2269*. Clinical long-read genome sequencing: Analytical performance of germline small variant detection using HiFi genome sequencing	1398
PB2270. Clinical utility of parental testing for the reclassification of likely pathogenic and uncertain copy number and sequence variants in a pediatric neurodevelopmental disorders cohort.	1399
PB2271. Clinical validation of end-to-end whole genome sequencing and interpretation applied to both panel-based and diagnostic genome testing.....	1400
PB2272. Clinical whole-genome sequencing of oral samples and microbial contamination	1401
PB2273. Comparing the analytical performance of exome sequencing and traditional panel testing in a cancer population.....	1402
PB2274. Comparison of chromosomal inversions in three different datasets	1403
PB2275. Comparison of diagnostic yield between exon-targeted microarray and standard microarray for patients with congenital heart defects (CHD).	1404
PB2276. Comparison of positive findings between CMA/exome combo and Whole Genome Sequencing: Aclinical lab experience	1405
PB2277. Copy Number Variations (CNVs) detected in 14% of patients presenting with peripheral neuropathies in a French cohort.....	1406
PB2278. CREBBP related Rubinstein Taybi syndrome: Report of 3 patients.....	1407
PB2279*. Deciphering the genetic etiology of idiopathic male infertility by low-pass mate-pair genome sequence analysis.	1408
PB2280. Deep exome sequencing to detect mosaic variants in craniofacial microsomia	1409
PB2281. Development and analytical validation of an innovative 397 gene genome-based inherited kidney disease panel	1410

ASHG 2022 Annual Meeting Poster Abstracts

PB2282. Development and clinical implementation of genome-based germline exome testing.	1411
PB2283. Development of a comprehensive whole genome sequencing test for cardiovascular disease patients.	1412
PB2284. Diagnosis of genetic causes of intellectual disability and multiple congenital abnormalities in center of excellence for human genetics in Egypt	1413
PB2285. Diagnostic approach of transthyretin amyloidosis in a Colombian patient.....	1414
PB2286. Diagnostic Utility and Lessons Learned from Deep Sequencing Vascular Malformations.	1415
PB2287. Diagnostic utility of the targeted next generation sequencing panel test for suspicious genetic glomerular diseases	1416
PB2288. Diagnostic yield of trio and individual Whole Exome Sequencing in a group of colombian patients with a suspected monogenic disease.	1417
PB2290. Does a larger gene panel size change the mutation detection rate in hypertrophic or dilated cardiomyopathy ?.....	1418
PB2291. <i>DST</i> variants are responsible for neurogenic arthrogyriosis multiplex congenita confirming the large clinical spectrum of type VI hereditary sensory autonomic neuropathy.....	1419
PB2292. Evaluate using artificial intelligence and machine learning in to predict deletion detection possibility in a cohort of chromosome 22q11.2 deletion syndrome cases	1420
PB2293. Evaluating the potential for using internal data in assigning benign classification to variants of uncertain significance.	1421
PB2294. Evidence review of the use of first-line genome sequencing to diagnose rare germline disorders.....	1422
PB2295. Examining the factors impacting molecular diagnosis during clinical exome sequencing re-analysis.....	1423
PB2296. Gabriella Miller Kids First Data Resource Center (DRC): Collaborative Platforms for Accelerating Pediatric Research	1424
PB2297. Generic genome sequencing: one lab flow for all.....	1425
PB2298. Genetic Testing Interpretation Consult (GTIC): A Novel Electronic Health Record (EHR) Support Tool to Further Characterize Variants of Interpretation for the Nongenetic Provider to Support Their Clinical Practice.	1426
PB2299. Genetic testing of <i>STRC</i> for hearing loss: a clinical lab's approach to pseudogenes.....	1427
PB2300. Genetic testing of the <i>ATP7B</i> gene for patients with suspected Wilson disease.	1428
PB2301. Genome sequencing enables parent of origin analysis of the X chromosome in probands with Klinefelter syndrome	1429
PB2302. Genome Sequencing Enables Precision Clinical Care in Genetic Leukoencephalopathies.	1430
PB2303. Genome-wide investigation of potentially pathogenic copy number variants & mechanisms fomenting their origins.....	1431
PB2304*. Genomic sequencing generates less uncertainty than panel-based testing: Results of over 1.5 million tests.....	1432

ASHG 2022 Annual Meeting Poster Abstracts

PB2305. GMCK-RD a success story: Implementation of genomic medicine in Stockholm healthcare region, update on first 10,000 samples	1434
PB2306. High diagnostic yield from clinical genome sequencing supports genome sequencing as first-tier genetic test: Evidence-based from 2002 index cases.....	1435
PB2307. Identification of a small duplication in the upstream noncoding enhancer of <i>SOX9</i> in a family with 46,XX disorder of sexual development by clinical whole genome sequencing.....	1436
PB2308. Identification of Balanced and Unbalanced Complex Chromosomal Rearrangement Involving Chromosomes 2, 4,5,18 and 21 in a family	1437
PB2309. Identification of somatic Neurofibromatosis 1 mosaicism using targeted next-generation sequencing of café-au-lait-macules.....	1438
PB2311. Importance of the reclassification of the clinical significance of new genetic variants in non-classical mucopolysaccharidosis type IV A.....	1439
PB2312. Inadequate parental sequencing depth burdens African ancestry patients with inaccurate de novo inheritance calls.	1440
PB2313. Increased diagnostic yield from negative whole genome-slice panels using automated reanalysis.....	1441
PB2314. Increased Diagnostic Yield With Long Read Sequencing In Patients With Undiagnosed Neurodevelopmental Disorder	1442
PB2315. Increasing global access to clinical whole genome sequencing for under-resourced undiagnosed individuals.....	1443
PB2316. Infinium Global Screening Array for CNVs detection.	1444
PB2317*. Integrating population-scale data into structural variant visualization.....	1445
PB2318*. Investigating the frequency of parental gonosomal mosaicism in Neurodevelopmental Disorder cohorts to improve the genetic diagnostic rate.....	1446
PB2319*. Is 22q11.2 deletion syndrome truly less common in African American patients?.....	1447
PB2320. <i>KANSL1</i> , mosaic or not?	1448
PB2321. Leveraging genomic advances to ascertain the cause of congenital limb defects in Indian subcontinent.	1449
PB2322. Long-read HiFi genome sequencing reveals a 2.7 kilobase intronic insertion in <i>NR5A1</i> as a cause of 46,XY disorder of sexual development	1450
PB2323. Low-level paternal mosaicism of novel pathogenic <i>DNM1L</i> variant confirms EMPF1 diagnosis in proband, highlighting importance of deep sequencing when parental carrier status is ambiguous.	1451
PB2325. Minimizing false negatives in NGS-based diagnostic testing by optimizing variant filtering strategy	1452
PB2326. Mitochondrial variants and short tandem repeat expansions detected by clinical whole genome sequencing.....	1453
PB2327. Molecular characterization of a rare case of chromosome 20p duplication syndrome in an Egyptian male patient with multiple congenital anomalies	1454

ASHG 2022 Annual Meeting Poster Abstracts

PB2328. Monozygous twins with distal 5p duplication: a case report.	1455
PB2329*. Multiple molecular diagnoses identified by clinical whole genome sequencing in an international cohort of more than 1,200 individuals	1456
PB2330. Mutation Spectrum of Polycystic Kidney Disease-Toronto Genetic Epidemiology Study .	1457
PB2331. Non-invasive detection of <i>TEK</i> mutations in a cohort of children with isolated venous malformations using cell-free DNA.....	1458
PB2332. Novel potential biomarkers for <i>DICER1</i> syndrome	1459
PB2333. Optical genome mapping and whole genome sequencing in a case of multiple chromosomal rearrangements	1460
PB2334. Optical genome mapping identified a likely pathogenic <i>POLR3B</i> variant in an undiagnosed male with ataxia, hypotonia, and cerebellar atrophy.....	1461
PB2335*. Optical genome mapping identifies double parental paracentric inversions as risk factor for atypical monocentric recombinant chromosomes in offspring	1462
PB2336. Optical genome mapping improves clinical interpretation of constitutional copy number gains.	1463
PB2337. Pallister-Killian syndrome: Characterization of the isochromosome 12p by fluorescent in situ hybridization: A Case Report.....	1464
PB2338. Pathogenic variants detected by RNA sequencing in Cornelia de Lange syndrome	1465
PB2339. Positive Findings in Clinical Mitochondrial Genome Testing by NGS.....	1466
PB2340. Potentially noteworthy intronic variants in the <i>COL5A1</i> and <i>COL5A2</i> genes in patients with classical Ehlers-Danlos syndrome: The necessity for RNA study and genetic diagnostic approaches	1467
PB2341. Rapid whole genome sequencing in the neonatal intensive care unit of Brazilian hospitals.	1468
PB2342. Recombinant chromosome with proximal 14q duplication in two sibs with mental retardation and facial dysmorphism due to maternal translocation t(14q21;21p11).	1469
PB2343*. Return of incidental genetic findings to pediatric patients: considerations and opportunities from experiences in genome sequencing	1470
PB2344. Review of Cytogenetic Findings of patients with Turner Syndrome and its variants among Filipinos from 1991 to 2020	1471
PB2345. Somatic Mutational Landscape of Extracranial Arteriovenous Malformations and Phenotypic Correlations.....	1472
PB2346. Statistical method for detection of uniparental disomy using SNP microarray or NGS technologies	1473
PB2347. Strategies to Reduce Inconclusive Genetic Testing Results for Inherited Cardiomyopathies: Perspectives from a Clinical Diagnostic Laboratory	1474
PB2348*. The perplexing 22q11.2 duplication syndrome: an important cause of congenital anomalies, medical conditions, cognitive deficits, and behavioral phenotypes - or nothing at all	1475

ASHG 2022 Annual Meeting Poster Abstracts

PB2349. The power of whole genome sequencing: Identification of an undocumented homozygous partial gene deletion in a patient with a neurological phenotype.....	1476
PB2350. The utility of whole exome sequencing in atypical cases of heterogeneous neurological disorders.....	1477
PB2351. Towards dried blood spot-based RNAseq in routine diagnosis of genetic disease.....	1478
PB2352. Transposable element insertions in 1000 Swedish individuals.....	1479
PB2353. Two new Tunisian cases of achondroplasia/hypochondroplasia : phenotype-genotype correlation.	1480
PB2354. Ultra-low DNA input with long-read sequencing identified complex chromosomal rearrangements involving <i>NIPBL</i> in a Cornelia de Lange syndrome patient.....	1481
PB2355. Unexpected CNV anomalies using exome or genome-wide approach for complex phenotypes: report of a Regional Genetic Center experience	1482
PB2356. Unmasking of a chromothripsis event using the integrated approach of Chromosomal Microarray Analysis (CMA) and Optical Genome Mapping (OGM).....	1483
PB2357. Unusual X-chromosome composition in mother and daughter with different karyotype constitutions	1484
PB2358*. Updated clinical practice guidelines for managing children and adults with 22q11.2 deletion syndrome.....	1485
PB2359. Utilization of the Age-based Semi Quantitative Metric (ASQM) to evaluate gene-condition pairs for inclusion in Early Check's newborn sequencing panel.	1486
PB2360. Value of DNA testing in the diagnosis of Sickle Cell Anemia in Childhood in an environment with a high prevalence of other causes of anemia.....	1487
PB2361. A Bayesian interaction model to estimate cell-type-specific Methylation Quantitative Trait Loci (meQTLs) incorporating priors from flow-sorted sequencing data	1488
PB2363. A long short-term memory autoencoder approach for detecting CRISPR genomic edits. ..	1489
PB2364. A methylation-based COVID-19 classification model to predict severe disease in vaccinated individuals.....	1490
PB2365. A simple workflow for methyl DNA analysis in FFPE-preserved Alzheimer and other neurodegenerative brain samples	1491
PB2366. A single cell chromatin accessibility and transcriptome atlas of the human heart improves the identification of risk variants and genes of Atrial Fibrillation.....	1492
PB2367. A single-cell atlas of 3D genomic structure and DNA methylation in human subcutaneous adipose tissue.	1493
PB2368. Activation of <i>Xist</i> by an evolutionarily conserved function of KDM5C demethylase.....	1494
PB2369. Additive effects of stress and alcohol exposure on accelerated epigenetic aging in Alcohol Use Disorder	1495
PB2370*. Alcohol use-disorder associated DNA methylation differences in human nucleus accumbens and dorsolateral pre-frontal cortex.....	1496
PB2371. Alterations of regulatory regions in T-cell Prolymphocytic Leukemia.	1497

ASHG 2022 Annual Meeting Poster Abstracts

PB2372*. Altered and allele-specific open chromatin landscape reveals epigenetic and genetic regulators of innate immunity in COVID-19.....	1498
PB2373. Analysis of methylation QTLs in breast cancer characterises the influence of germline variation on the abnormal cancer methylome	1499
PB2374. <i>ASXL3</i> links chromatin biology to neurodevelopment disorders.....	1500
PB2375. Atypical Prader-Willi and Angelman syndrome deletion: Importance of parent of origin detection.....	1501
PB2376*. Cardiac development in mammalian models lacking the PR-DUB component <i>ASXL3</i> ...	1502
PB2377. Catalysis of histone H3K27me3 by Polycomb-independent mechanisms.....	1503
PB2378. Characterization of nucleosome positioning and DNA methylation signatures using prostate cancer cells from diverse ethnic groups using NOME-EM-seq.....	1504
PB2379. Characterization Of Pten Expression Cis-Regulatory Elements Involved In Cowden Disease	1505
PB2381. ChromGene: Gene-Based Modeling of Epigenomic Data.....	1506
PB2382. Clinical Epigenomic Testing in Canada: Discovery and Clinical Assessment of Episignatures	1507
PB2383. Conserved GATA4/5/6-mediated gene regulation in cardiac development and disease	1508
PB2384*. Cord blood DNA methylation alterations potentially mediate associations between adverse pregnancy outcomes and childhood blood pressure.....	1509
PB2385. Could the difference in DNA methylation status explain phenotypic variability in patients with 5p- syndrome?.....	1510
PB2386. Critical epigenomics association between methylation and clinical phenotype	1511
PB2387. Cross-species and tissue imputation of species-level DNA methylation samples.....	1512
PB2388*. Cross-tissue patterns of DNA hypomethylation reveal genetically distinct histories of cell development.....	1513
PB2389. Defining the gene regulatory roles of non-coding variants in the pathogenesis of autism. .	1514
PB2390*. Defining the genetic background of immunoglobulin G galactosylation in humans.....	1515
PB2391. Differential newborn DNA methylation among individuals with complex congenital heart defects and pediatric lymphoma.....	1516
PB2392. Differentially methylated probes and regions associated with the Healthy Eating Index....	1517
PB2393. Digitally untangling the 3D genome	1518
PB2394. Disparities in epigenetic modifications in <i>NRF1</i> and <i>FTO</i> genes in children.....	1519
PB2395. Dissecting mechanisms underlying expression divergence of human duplicated genes	1520
PB2396*. Dissecting the molecular machinery and sequence of histone to protamine transition during spermiogenesis.....	1521
PB2397. Distinct signatures in malignant PEComa and leiomyosarcoma identified by integrative RNA-seq and H3K27ac ChIP-seq analysis	1522
PB2398. Divergent age-related methylation patterns in long and short-lived mammals.....	1523

ASHG 2022 Annual Meeting Poster Abstracts

PB2399. DNA methylation changes among Ethiopian women diagnosed for cervical cancer. 1524

PB2400. DNA methylation changes associate with measured glomerular filtration rate in an American Indian cohort with type 2 diabetes. 1525

PB2401. DNA methylation epi-signature and biological age in attention deficit hyperactivity disorder patients 1526

PB2402. DNA Methylation Episignature of Valproate Embryopathy..... 1527

PB2403. DNA methylation in twins discordant for van der Woude syndrome 1528

PB2404. DNA methylation profile in multiple system atrophy and progressive supranuclear palsy. 1529

PB2405. DNA methylation signatures in Arab and Middle Eastern families with neurodevelopmental disorders provide insight to disease etiology. 1530

PB2406. Dysregulation of serum and exosomal miR-7-1-5p and miR-223-3p in Parkinson’s disease 1531

PB2407. Early Embryos Employ a Unique Mechanism of CENP-A Equalization during the First Cell Cycle 1532

PB2408*. Epigenetic Age Acceleration is Suggestively Associated with Stroke Precursor Phenotypes 1533

PB2409. Epigenetic profiling of isolated blood cells reveals highly cell-type specific smoking signatures and links to disease risk 1534

PB2410. Epigenetic regulation of Wnt signaling pathway associated with age-related mobility loss. 1535

PB2411. Epigenetic signatures of asthma in nasal epithelium from African ancestry populations from the CAAPA consortium 1536

PB2412*. Epigenome-wide association meta-analysis of DNA methylation with lifetime cannabis use 1537

PB2413. Epigenome-wide association of DNA methylation markers for dilated cardiomyopathy in left ventricular heart tissues..... 1538

PB2414*. Epigenome-wide association study in multi-ethnic Asian populations identifies novel markers for incident type 2 diabetes 1539

PB2415. Epigenome-wide association study reveals CpG sites associated with thyroid function and regulatory effects on *KLF9* 1540

PB2416. Epstein-Barr Nuclear Antigen 2 (EBNA2) types 1 and 2 have shared and distinct host DNA binding partners. 1541

PB2417*. Expanded studies of a methylation-based COVID-19 classification model to predict severity of disease and its ability to differentiate from other respiratory viruses. 1542

PB2418. Exploring novel regulatory elements in X-chromosome inactivation..... 1543

PB2419. Extracellular vesicles microRNAs in traumatic brain injury patients with and without intracranial abnormality 1544

PB2420. Finding the Right Match: the Role of ZCWPW1 in Mammalian Meiosis..... 1545

PB2421. First trimester human placenta DNA methylation correlates to maternal estradiol levels... 1546

ASHG 2022 Annual Meeting Poster Abstracts

PB2422. Functional Analysis of CHD2 and CHD8 <i>De Novo</i> Mutations and Related Molecular Events Involved in Neurodevelopmental Disorders	1547
PB2423*. Gene regulatory network synchronizes genetic and epigenetic signals, prioritizes GWAS SNPs, and identifies repurposable drug candidates for multiple sclerosis.....	1548
PB2424*. Genetic variation in correlated regulatory region of immunity.....	1549
PB2425. Genome-wide DNA methylation profiling in subcortical regions reveals epigenetic signatures associated with PTSD and MDD	1550
PB2426*. Genome-wide dysregulation of R-loops in Ataxia Telangiectasia neurological pathogenesis.	1551
PB2427*. Genome-wide evaluation of the effect short tandem repeat variation on local DNA methylation	1552
PB2428*. Genome-wide mapping implicates 5-hydroxymethylcytosines in diabetes and Alzheimer's disease	1553
PB2430. Genome-wide neonatal epigenetic changes associated with maternal exposure to the COVID-19 pandemic	1554
PB2431. Genome-wide screening of epigenetic variations in the EpiSign Knowledge Database.....	1555
PB2432. Global identification of cortical circRNAs and circRNA clusters associated with Alzheimer's disease	1556
PB2433*. High dimensional co-expression network analysis unravels transcriptomic drivers of diverse biological systems.....	1557
PB2434. High-dimension to high-dimension screening for detecting genome-wise epigenetic regulators of gene expression.....	1558
PB2435. High-resolution analysis identifies ancestry-specific chromatin 3D interactions	1559
PB2436. Human epi- and transcriptomic trajectories in the postprandial state.	1560
PB2437. Human gene regulatory evolution is driven by divergence in both <i>cis</i> and <i>trans</i>	1561
PB2438. Identification of diagnostic DNA methylation epesignatures associated with large structural chromosomal variations in patients with neurodevelopmental disorders.	1562
PB2439. Identifying functional interactors of conserved meiotic histone reader ZCWPW1	1563
PB2440. Immune cell-type level epigenome-wide association analysis on chronic HIV infection highlights hallmark genes for HIV pathogenesis and cancer.....	1564
PB2441*. Investigating aberrant DNA methylation in pediatric epilepsies	1565
PB2442. Investigation of the effects of alcohol exposure on the chromatin binding and localization of cohesin and CTCF.....	1566
PB2443. Investigation of Tissue Level of Molecular Pathways Effective in Vascular Calcification at Chronic Renal Insufficiency Patients.....	1567
PB2444. KDM6A/B inhibition rescues an osteoblast phenotype and altered gene expression in a novel mouse model of Weaver Syndrome with skeletal overgrowth.	1568
PB2445. Learning a generalized regulatory model from paired chromatin accessibility and transcriptome.....	1569

ASHG 2022 Annual Meeting Poster Abstracts

PB2446. Learning the cis and trans sequence features of gene regulatory divergence between human and rhesus macaque	1570
PB2447. Linking genetic variants associated with gluteofemoral fat storage to determine their role in adipogenesis.....	1571
PB2448*. Loss of DNMT3A or TET2 activity in stimulated macrophages alters transcription factor binding and enhances inflammatory gene expression.....	1572
PB2449*. Mapping chromatin accessibility QTL in 138 liver tissue samples identifies coordinated regulation, links regulatory elements to genes, and predicts mechanisms at GWAS loci.	1573
PB2450. Methylation-based machine learning classification of osteoarthritis in the Multicenter Osteoarthritis Study.	1574
PB2451*. MIMOSA: A method for improved methylome imputation increases power to identify CpG site-phenotype associations.....	1575
PB2452. Multiscale phase separation by explosive percolation with the single chromatin loop resolution.....	1576
PB2453*. NANOME: A Nextflow pipeline for haplotype-aware allele-specific consensus DNA methylation detection by nanopore long-read sequencing.....	1577
PB2454. Nascent RNA reveals new links between genetic variation and disease at multiple levels of regulation	1578
PB2455. Network analysis across cellular reprogramming states.	1579
PB2456. Nucleosome spike-in controls identify best-in-class antibodies and enable reliable next-generation epigenomic mapping approaches	1580
PB2457*. Peripheral immune roles in bipolar disorder: genomic, epigenomic, and phenotypic integration	1581
PB2458. Rare inherited Copy Number Variants as genetic modifiers of Developmental Disorders..	1582
PB2459. Reduced-Representation Methylation Sequencing (RRMS): Oxford Nanopore’s alternative to RRBS enables accurate and easy to use methylation calling in CpG islands, shores, shelves, and promotor regions using a MinION flowcell.....	1583
PB2460. RFX transcription factors regulate genes involved with primary cilium in glioma.....	1584
PB2461*. Role of TET1-mediated epigenetic modulation in Alzheimer’s disease.....	1585
PB2462. Scoring individual cells in snATACseq dataset based on GWAS fine-mapping identifies disease-relevant cell-types in neurodegenerative diseases.....	1586
PB2463*. Serum miR-451a is up-regulated in sarcopenic patients compared to healthy controls.....	1587
PB2464. Simultaneous sequencing of genetics and epigenetics provides opportunities for new biological insights.	1588
PB2465*. Single-cell sequencing links reduced glucose metabolism to cocaine addiction-like behavior in rats.....	1589
PB2466. Single-molecule architecture and heterogeneity of human telomeric DNA and chromatin	1590
PB2467. SnapHiC-G: identifying long-range enhancer-promoter interactions from single cell Hi-C data via a global background model	1591

ASHG 2022 Annual Meeting Poster Abstracts

PB2468. SpliceVI: a visualization tool for predicting protein consequences based on SpliceAI data.	1592
PB2469. Stability of genetic regulation of gene expression over time.	1593
PB2470*. Subsetting systemic lupus erythematosus patients based on clustering of DNA methylation at the time of disease flare	1594
PB2471. TDP-43 depletion results in aberrant gene expression and TE activation via dysregulation of R-loop homeostasis.....	1595
PB2472. The association of cigarette smoking with DNA methylation and gene expression in human tissues.....	1596
PB2473. The interplay of epigenetic, genetic, and traditional risk factors on blood pressure: Findings from the Health and Retirement Study.	1597
PB2474. The role of genetic and epigenetic mechanisms in the definition of the circular RNA landscape in multiple sclerosis.....	1598
PB2475. Tobacco smoke-derived metabolites in plasma are associated with cardiometabolic traits, are genetically regulated, and may modulate adipose tissue DNA-methylation in African Americans...	1599
PB2476. Transcriptome and chromatin accessibility dynamics across 25 brain regions identify novel susceptibility gene sets for neuropsychiatric disorders	1600
PB2477. Transfer learning of a methylation based COVID19 severity model for DNA viral infection by eczema herpeticum.....	1601
PB2478. Transposable elements are associated with the variable response to influenza infection	1602
PB2479. Trinity of chromatin architects - The coordination of CTCF, RNAPOL2 and Cohesin in shaping the genomic landscape.....	1603
PB2480. Uncovering Novel Functions of Histone Demethylase KDM5 Through a Genome-wide Approach.....	1604
PB2481. Using allele-specific expression to assess intra- and inter- individual heterogeneity in X-chromosome inactivation within human tissues	1605
PB2482*. Using iPSC-derived microglia and oligodendrocytes to study ancestry-specific gene expression in the context of Alzheimer’s Disease (AD).....	1606
PB2483. Using machine learning to predict 3D genome organization across thousands of diverse individuals reveals conservation and population differentiation.....	1607
PB2484. Validation and Methylome analysis of the CRISPR-edited DUX4 locus in immortalized myoblast cell lines modeling FSHD using long-read sequencing.	1608
PB2485. Variablecis-regulatory landscapes associate with gene function and expression patterns...	1609
PB2486. Variation in DNA methylation is associated with cognitive function in post-surgery breast cancer patients prior to adjuvant therapy.	1610
PB2487. <i>Variation in miRNAs may contribute to orient future therapeutic strategies in hidradenitis suppurativa; An epigenome-widescreening approach.....</i>	1611
PB2488. Vitamin c contributes to epigenetic regulation of genes related to diabetic retinopathy in retinal endothelial cells.	1612

ASHG 2022 Annual Meeting Poster Abstracts

PB2489. Whole genome bisulfite sequencing identifies methylation markers in cord blood linked to maternal hyperglycaemia and childhood metabolic abnormalities	1613
PB2490. Whole-methylomics reveals differentially methylated genes in blood associated with Late-onset Alzheimer's disease	1614
PB2491*. Widespread age-associated changes in the chromatin architecture of skeletal muscle cell type populations	1615
PB2492. A Catalogue of Transcriptomes and Associated Genetic Effects on 2,000 Qataris Uncovering the Functional Impact of Middle Eastern Genetic Variation and Identifying Novel Pathways Underlying Human Traits and Diseases	1616
PB2493. A gain-of-function recurrent missense variant leads to a GABAergic/glutamatergic imbalance in a forebrain organoid model of PACS1 syndrome	1617
PB2494. A genome-first approach characterizing the frequency and phenotype of potentially pathogenic germline <i>DGCR8</i> variants in two large, unselected cohorts	1618
PB2495*. A Homozygous IER3IP1 Mutation Causes Secretory Protein Trafficking Defects in Neural Progenitor Cells	1619
PB2496. A humanized yeast model to identify disease-associated, dominant-negative effects on protein function	1620
PB2497. A lupus-associated variant in <i>IRF7</i> amplifies <i>IFN-α</i> production.	1621
PB2498. A Novel De novo Likely Pathogenic Variant of <i>WFS-1</i> gene in a Pakistani Child	1622
PB2499. A rare genetic variant in the cleavage site of <i>prepro-orexin</i> is associated with idiopathic hypersomnia	1623
PB2500. A risk variant for Barrett's esophagus and esophageal adenocarcinoma at chr8p23.1 affects enhancer activity and implicates multiple gene targets	1624
PB2501. A single-nucleus transcriptome-wide association study implicates novel genes in depression pathogenesis	1625
PB2502. A splicing variant found in the human myostatin gene encodes an isoform that inhibits myostatin	1626
PB2503. A structural variant of the C-terminal prion-like domain of TDP-43 causes vacuolar muscle degeneration	1627
PB2504*. A uniquely prevalent deleterious variant within <i>POMC</i> gene identified in Estonian Biobank cohort	1628
PB2505. <i>A2m11</i> -Knockout Mouse as a Potential Model of Chronic Otitis Media	1629
PB2506*. Aberrant splicing prediction across human tissues.	1630
PB2507. Abnormal cell fate allocation in embryogenesis caused by <i>Zic3</i> loss-of-function	1631
PB2508. Abundance of immunoprotein CD99 affected by loss of chromosome Y	1632
PB2510. All-but-one conditional analysis of eQTL isolation in peripheral blood sheds light onto causal variant prioritization	1633
PB2511. Allele-specific expression in blood cells during aging	1634

ASHG 2022 Annual Meeting Poster Abstracts

PB2512. Amplicon sequencing-based noninvasive fetal genotyping for RHD-Positive D antigen-negative alleles.....	1635
PB2513. An African-specific Alzheimer disease-associated <i>ABCA7</i> frameshift deletion results in altered microglia functionality.....	1636
PB2514. An Alzheimer's disease risk variant in <i>TTC3</i> modifies the growth and transcriptional profile of iPSC-derived forebrain neurons.....	1637
PB2515. An integrated machine learning and functional analysis approach for resolution of variants of uncertain significance (VUS) in <i>PCDH19</i>	1638
PB2516*. An updated genetic atlas of the plasma proteome: Findings from the UK Biobank Pharma Proteomics Project (UKB-PPP).....	1639
PB2517. Antibody repertoire heavy chain gene usage is explained by common genetic variants in the immunoglobulin heavy chain locus.	1640
PB2518. <i>APOE</i> Genotype Contributes to Distinct Profiles in Human iPSC Astrocytes and Neurons	1641
PB2519. Assessing the contribution of variants with functional effects on post-transcriptional regulation to genetics of psychiatric disease.....	1642
PB2520. Assessing the genetic control of whole blood RNA editing and its role in modulating risk of common diseases.	1643
PB2521. Assessment of genotype-phenotype correlation of human <i>TEX11</i> variants associated with non-obstructive azoospermia in mouse.....	1644
PB2522. Association of GWAS and candidate gene loci of dopaminergic system with major depression, schizophrenia and bipolar disorder in the Pakistani population.....	1645
PB2523. Association of MTHFR C677T polymorphism with obesity in women: A case-control study and meta-analysis.....	1646
PB2524. Bayesian fine-mapping of sex-specific gene regulation during neurodevelopment shows link to neurodevelopmental disorders.....	1647
PB2525. Body mass index as an environmental context in functional genomic annotations.	1648
PB2526. Calibration set for high throughput functional assay for <i>RYRI</i> Variants.....	1649
PB2527. CD4+ T cell gene expression influences COVID-19 severity and susceptibility: results from single cell sequencing and Mendelian randomization.....	1650
PB2528. Cellular consequences of Hirschsprung disease-associated RET variants studied utilizing CRISPR-mediated genome engineering.	1651
PB2529. Characterization of a leaky splice mutation in <i>AIRE</i> , associated with milder phenotype and late debut of autoimmune polyendocrine syndrome type 1.	1652
PB2530. Characterization of haplotype diversity in the human immunoglobulin heavy chain constant region using long-read sequencing.....	1653
PB2531. Characterization of mutants in <i>SOS1</i> in an endothelial cell model of lymphatic anomaly..	1654
PB2532. Characterizing a novel missense variant in the methionine salvage gene <i>MRI1</i> that conveys protection for type 2 diabetes.....	1655

ASHG 2022 Annual Meeting Poster Abstracts

PB2533. Characterizing the effects of a novel <i>RDH12</i> donor splice site variant using <i>in vivo</i> , <i>in vitro</i> , and <i>in silico</i> methods	1656
PB2534*. Chromatin conformation during CD4+ T cell activation implicates putative autoimmune disease candidate genes.....	1657
PB2535. Clonal hematopoiesis of indeterminate potential and kidney function decline in the general population	1658
PB2536. Clustering of missense mutations follows a non-random distribution across the human genome and correlates with specific classes of genes.....	1659
PB2537. COL9A1 and FLNB: Variants corresponding to complex clinical phenotype.	1660
PB2538. Colocalization of blood cell traits GWAS associations and variation in <i>PU.1</i> genomic occupancy prioritizes causal noncoding regulatory variants.....	1661
PB2539. Combining human genetics and functional genomics identifies regulators of the lysosomal lipid BMP.....	1662
PB2540. Comparative titin gene frequencies and distributions across four population cohorts.....	1663
PB2541. Comparison of pedigree-based graph workflow to more traditional workflows for rare candidate variant analysis	1664
PB2542. Comprehensive detection of trans-regulatory signal from scRNA-seq and perturbation datasets.....	1665
PB2543. Comprehensive signal identification highlights the contribution of protein QTLs to complex trait genetics.....	1666
PB2544*. Conditionally distinct adipose eQTL signals in 2,256 individuals identify hundreds of colocalized genes for cardiometabolic traits.....	1667
PB2545. Congenital Myasthenic Syndrome Due to <i>DOCK7</i> Gene Mutation Case Study Report	1668
PB2546. Connecting rare variation to extremes of plasma protein levels	1669
PB2547*. Consequences of loss of function burden effects on the human plasma proteome in 54,306 UK Biobank participants.....	1670
PB2548. CRISPR perturbation screens identify predicted regulatory sequences and genes within GWAS loci associated with erythrocyte density traits.....	1671
PB2549. CRISPR/Cas9-mediated knockout and cDNA rescue using sgRNAs that target exon-intron junctions in <i>Drosophila</i> enable functional analysis of pathogenic variants in a tissue-specific manner.	1672
PB2550. Cryptic AS-NMD elements harbor relevant variants in probands with early onset genetic disease.	1673
PB2551. CYP2C19 Polymorphism in Ischemic Heart Disease Patients Taking Clopidogrel after Percutaneous Coronary Intervention in Egypt.....	1674
PB2552*. Defining the effects of noncoding genetic variation on human regulatory element activity	1675
PB2553*. Deletion mapping of regulatory elements for <i>GATA3</i> reveals a distal T helper 2 cell enhancer involved in allergic diseases.	1676

ASHG 2022 Annual Meeting Poster Abstracts

PB2554. Detection of rare Synaptogyrin 3 (<i>SYNGR3</i>) gene missense mutations in patients with schizophrenia.	1677
PB2555. Determining the rate of unintended on-target and off-target effects induced by CRISPR-Cas gene editing in iPSC clones	1678
PB2556. Development of a multiplexed assay of variant effects for the DNA repair gene <i>MUTYH</i>	1679
PB2557. Development of a yQTL Discovery Pipeline Applicable for Both Unrelated and Related Individuals.....	1680
PB2558. Discovery and finemapping of eGFR loci in metaanalysis GWAS of 80K African ancestry individuals.....	1681
PB2559. Disease-specific analysis improves prioritization of non-coding genetic variants	1682
PB2560. Dissecting the downstream effects of IL-6 signaling on atheroprogession: a proteome-wide Mendelian randomization study.....	1683
PB2561. <i>Drosophila</i> assays for variant interpretation in the polycomb repressive complex 2 related syndromes	1684
PB2562. Dual inhibition of mTORC1 and mTORC2 activation effectively rescues hyperproliferative lymphatic sprouting in a cell-based model system of complex lymphatic anomalies	1685
PB2563. Ehlers-Danlos Syndrome: A Likely Pathogenic Variant Found in <i>COL5A1</i> gene	1686
PB2564. Elucidating genomic disease associations using transcriptomic, proteomic and metabolomic data in 4,732 individuals	1687
PB2565. Elucidating mechanisms of genetic interactions and oligogenic inheritance in Autism Spectrum Disorder.	1688
PB2566. Elucidating the mechanism of a neurological syndrome caused by germline mutations in <i>H3F3A</i> and <i>H3F3B</i>	1689
PB2567. Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease.....	1690
PB2568. Epitranscriptomic m6A alterations in C9orf72 ALS cerebellum.....	1691
PB2569*. Escape from Nonsense-Mediated Decay: Annotation of Transcripts with Protein Truncating Variants	1692
PB2570. Establishing a deep mutational scan for functional classification of <i>MSH6</i> missense variants	1693
PB2571. Establishment of <i>in vitro</i> assay system for evaluating of <i>RERE</i> variants.....	1694
PB2572. Estimating Diagnostic Noise in Panel-Based Genomic Analysis	1695
PB2573. Evaluation of regulatory potential of European ancestry <i>APOEε3 TOMM40-523'</i> repeat haplotypes with differential risk effects.....	1696
PB2574. Evaluation of the ACMG PM5 variant pathogenicity evidence guideline.....	1697
PB2575. Evidence from a massively parallel reporter assay and genome engineering for non-coding variants at 1q32- <i>IRF6</i> that directly influence risk for orofacial cleft.....	1698
PB2576. Evolutionary study and variant analysis in the <i>mcph1</i> gene in patients diagnosed with non-syndromic hearing impairment	1699

ASHG 2022 Annual Meeting Poster Abstracts

PB2577. Examination of predicted loss of function (pLoF) variants in the general population reveals high rates of rescue and artifacts.....	1700
PB2578. Exome analysis for bronchial asthma in the Mennonite population reveals an association with <i>NOTCH4</i> polymorphisms	1701
PB2579. Exome-wide association analyses of 1,463 proteins in 31k participants of the UK Biobank reveal the spectrum of coding variant effects on protein levels.....	1702
PB2580. Expanding the rare <i>DHCR24</i> -related sterol biosynthesis disorder: genotype -phenotype correlation.	1703
PB2581. Expansion of the histidine-rich domain of <i>DYRK1A</i> in a patient with developmental delay, short stature, and microcephaly: A case report.	1704
PB2582. Explicable prioritization of genetic variants by integration of rule-based and deep learning algorithms for diagnosis of rare Mendelian disorders.....	1705
PB2583. Exploiting a human stem cell-based modeling approach to functionally characterize loss-of-function of lncRNA RMST as a cause of Kallmann Syndrome.	1706
PB2584. Fast Forward - is Multiomics a resurgence of old?	1707
PB2585*. Fine-mapping causal tissues and genes at disease-associated loci.....	1708
PB2586. FOXI3 pathogenic variants cause one form of craniofacial microsomia.....	1709
PB2587. Fragile X related genes and Dlg4 gene are involved in process formation in neuronal cells	1710
PB2588. Full penetrance of a hypomorphic TYR gene variation in oculocutaneous albinism.	1711
PB2589. Functional analysis of genetic variations found in clinical exomes of ASD patients.	1712
PB2590. Functional and <i>in silico</i> analyses of variants found in mitochondrial trifunctional protein deficiency patients	1713
PB2591*. Functional characterization of a heat shock regulated enhancer at coronary artery disease associated locus that controls CALCRL expression in endothelial cells	1714
PB2592. Functional characterization of new non-coding pathogenic variants creating upstream Open Reading Frames in the 5UTR of the Endoglin gene and causing Hereditary Hemorrhagic Telangiectasia.	1715
PB2593. Functional Characterization of Novel UBE3A Variants Associated with Angelman Syndrome	1716
PB2594. Functional characterization of polycystic ovary syndrome-associated risk loci identifies genetic regulatory regions.....	1717
PB2595. Functional impact of rare variants and sex across the x chromosome and autosomes.....	1718
PB2596. Gene discovery and pathogenic mechanisms of disease in worms, flies, and fish: The Model Organisms Screening Center for the Undiagnosed Diseases Network	1719
PB2597. Gene expression profiling with direct long-read RNA sequencing uncovers functional variation affecting transcripts production.	1720
PB2598. Gene prioritization for rare diseases integrating genotype, RNA-seq and phenotype - lessons from a CAGI 6 challenger team.....	1721

ASHG 2022 Annual Meeting Poster Abstracts

PB2599. Genetic control of mRNA splicing as a potential driver for incomplete penetrance of coding variants.....	1722
PB2600*. Genetic determinants of blood transcript splicing and impact on molecular phenotypes in 4732 healthy individuals.....	1723
PB2601. Genetic determination of fetal hemoglobin levels in patients with sickle cell disease	1724
PB2602. Genetic effects on chromatin accessibility and nucleosome positioning using snATAC data across 284 skeletal muscle biopsies.....	1725
PB2603. Genetic inactivation of zinc transporter SLC39A5 improves liver function and hyperglycemia in obesogenic settings	1726
PB2604. Genetic regulation of OAS1 nonsense-mediated decay underlies association with COVID-19 hospitalization in patients of European and African ancestries	1727
PB2605. Genetic regulation of splicing in macular and peripheral retina reveals novel molecular mechanisms for age-related macular degeneration.....	1728
PB2606*. Genetically Decreased CPS1 Activity Attenuates Atherosclerosis in Humans and Mice Through Sexually Dimorphic Patterns.....	1729
PB2607. Genome-First Approach to Explore Prevalence and Cancer Risk of Adults with Pathogenic and Likely Pathogenic Variants in RASopathy Genes	1730
PB2608. Genome-wide association study of serum metabolites and lipids in Multiple Sclerosis	1731
PB2609. Genome-wide characterization of selective constraint on variation within CTCF Binding Sites.....	1732
PB2610. Genome-wide microRNA expression quantitative trait loci from 596 adult brains.....	1733
PB2611*. Genome-wide pleiotropy study identifies association of <i>PDGFB</i> with Age-Related Macular Degeneration and COVID-19 infection outcomes.....	1734
PB2612. Genome-wide transcriptome correlation analysis identifies shared pathways between coronary heart disease and hematologic traits.....	1735
PB2613. Genomic characterization of the immunoglobulin light chain lambda locus from individuals of European, Asian and African origin.....	1736
PB2614*. Glucocorticoids unmask silent non-coding genetic risk variants for common diseases. ...	1737
PB2615*. Hematopoietic Loss of Y Chromosome Leads to Cardiac Fibrosis and Heart Failure Mortality	1738
PB2616. Human genetic variation reveals regulators of <i>Yersinia pestis</i> cellular Infection.....	1739
PB2617. Human pancreatic islet microRNAs implicated in diabetes and related traits by large-scale genetic analysis.....	1740
PB2618. Hypotonia due to homozygous mutation of <i>NALCN</i>	1741
PB2619. Identification of bidirectional regulatory regions for <i>FADS1</i> and <i>FADS2</i>	1742
PB2620. Identification of polyadenylation signals relevant to Mendelian disease variant interpretation.	1743
PB2622. Identification of susceptibility loci for frailty through a genome-wide scans and genetic association study of aging-related variants in Korean population	1744

ASHG 2022 Annual Meeting Poster Abstracts

PB2623. Identifying context-specific genetically regulated expression changes in monocytes associated with resilience to Alzheimer’s Disease	1745
PB2624*. Identifying functional variants related to heart development in African populations.	1746
PB2625. Impact of rare heterozygous mutations of <i>PCSK1</i> on obesity: implication for treatment with MC4R agonists.....	1747
PB2626. Improving generalizability of gene expression prediction models from African-ancestry populations.....	1748
PB2627. Inhibiting aurora kinase A in the HMC3 immortalized human microglia cell line increases phagocytosis of amyloid beta <i>in-vitro</i>	1749
PB2628. Integrated analysis of retinal eQTLs, meQTLs and eQTM identifies target genes and epigenetic mechanisms contributing to age-related macular degeneration.....	1750
PB2629. Integrative 3D structural analysis of <i>de novo</i> missense variants and their associated protein structures inferred by AlphaFold2 identifies the importance of β -strand/sheet features for congenital heart disease	1751
PB2630*. Investigating the impact of non-coding structural variants flanking epilepsy-associated gene, <i>FOXG1</i> on chromatin architecture and gene regulation.	1752
PB2631. Investigating the molecular and cellular effects of pathogenic variants of <i>GNAI1</i> in developmental and epileptic encephalopathy	1753
PB2632. Investigating the potential roles of Importin-7 and Importin- β in LINE-1 retrotransposition.	1754
PB2633. KDM5A, a chromatin remodeler and newly identified autism gene, has a cell type-specific function in the hippocampus.	1755
PB2634. KMT5B is required for early motor development.	1756
PB2635. Large-scale integrative analysis of 1.5 million single cells across diverse human tissues and immune cell types reveals dynamic regulatory effects on <i>HLA</i> gene expression.	1757
PB2636. Leveraging Mosaic Epileptogenic Human Brain Tissue to Study Cell-type Specific Transcriptional Changes Associated with a Pathogenic <i>PIK3CA</i> Variant.....	1758
PB2637. Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery	1759
PB2638. Long-chain fatty acid oxidation disorder gene variants identified through a gene panel sponsored program: a diverse gene variant landscape.	1760
PB2639. Mapping the effects of genomic deletions and duplications on cognitive ability across cortex	1761
PB2640*. Massively parallel reporter assays with multi-layer annotations identified cell-type-specific functional variants and genes associated with melanoma.....	1762
PB2641. Massively parallel saturation genome editing of an essential mitochondrial targeting sequence.	1763
PB2642. Maternally Inherited novel <i>SUMO2</i> variant associated with a complex phenotype	1764
PB2643. Mechanisms of adaptation in high-altitude pregnancy: association of genotype with oxygen delivery and placental metabolism.....	1765

ASHG 2022 Annual Meeting Poster Abstracts

PB2644. Mendelian randomization reveals the causal interaction of <i>MICA</i> transcription levels with the risk of Graves' disease	1766
PB2645. Metabolomics to clarify the mechanism of <i>ALDH2</i> polymorphism-induced alcoholic liver injury.....	1767
PB2646. Mimicking NDD phenotypes and functional characterization of <i>PPP1R9A</i> ^{+/-} mutation using iPSCs- derived neurons and single-cell transcriptomics	1768
PB2647. Misexpression of inactive genes is associated with nearby rare structural variants.....	1769
PB2648. Mitochondrial Complex III (CIII) deficiency case study: Analysis of a likely pathogenic variant in <i>UQCRC2</i> gene as the etiology of intellectual disabilities and obesity.	1770
PB2649*. Mitotic recombination is a common mechanism of cellular mosaicism in Fanconi anemia.	1771
PB2650. Modeling single-cell activation states enhances power to identify ex vivo stimulation response eQTLs.	1772
PB2651. Modeling the effects of <i>KCNQ1</i> SNPs on type 2 diabetes risk in iPSC-derived pancreatic islets identifies an early role of methylation that alters islet composition and beta-cell mass.....	1773
PB2652. Mosaicism of a <i>de novo TP53</i> mutation in a patient with breast cancer: value of two germ layers study	1774
PB2653. Multiomics identifies dysregulation of Wnt-signaling through epigenetic gene regulation across tissues in Bohring-Opitz Syndrome	1775
PB2654. Mutant <i>Wac</i> mice exhibit phenotypes relevant to DeSanto-Shinawi Syndrome and provide initial insights into molecular and developmental mechanisms.....	1776
PB2655*. Mutational scanning to determine pathogenicity of Variants of Uncertain Significance in genes in the Sonic Hedgehog Pathway	1777
PB2656. NASP contributes to autism via epigenetic dysregulation of neural and immune signaling pathways	1778
PB2657*. Natural variants in <i>SEL1L</i> modify lethality, ERAD, and proteasome function in a model of NGLY1 deficiency.....	1779
PB2658. Novel actionable targets discovered by proteomic genome-wide association of 1715 unstudied proteins.....	1780
PB2660. Novel non-coding variant in <i>NMNAT1</i> reduces the transcription of <i>NMNAT1</i> , which leads to Leber Congenital amaurosis.....	1781
PB2661. Novel plasma and brain proteins that are implicated in multiple sclerosis risk	1782
PB2662. Novel splice variant in <i>PRUNE1</i> leads to aberrant neurodevelopment and neurodegeneration.	1783
PB2663. Nuclear abnormalities in novel <i>LMNA</i> variant segregating with <i>LMNA</i> -associated cardiocutaneous progeria syndrome.....	1784
PB2664*. Organ-specific prioritization of non-coding regulatory variants with stacking generalization.	1785
PB2665. Parallel functional screening of human duplicated genes in neurodevelopment at single-cell resolution using zebrafish	1786

PB2666. Pathogenic mechanisms associated with a recurrent *CUX2* missense variant in epilepsy... 1787

PB2667*. Pathogenic variants in *PLEKHO2* predispose to heritable thoracic aortic disease 1788

PB2668*. Patient-derived iPSC-CMs from TANGO2-deficient disorder revealed multi-channel defects, functional association between ERK1/2 and TANGO2, and folate as a potential therapeutic agent to mitigate arrhythmias in patients 1789

PB2669*. Phenotypic and functional genomic impact of complex structural variation at the 17q21.31 locus. 1790

PB2670. PIEZO2 as a molecular candidate of respiratory phenotype induced by Congenital Central Hypoventilation Syndrome mutations in *PHOX2B*. 1791

PB2671. Pooled RNA-IP approach to investigate variant effects on RBP binding and splicing..... 1792

PB2672. Population genetic study for 16 autosomal STR loci < and > for > in unrelated 15,546 Korean individuals 1793

PB2673. Population-specific non-coding and coding putative causal variants shape quantitative traits 1794

PB2674. Population-specific RNA-seq data empowers the functional follow-up of disease-associations from FinnGen..... 1795

PB2675. Possible complex Di-Genetic inheritance in Primary Congenital Glaucoma 1796

PB2676. Predicted nasal epithelial transcriptome-wide association study in African-ancestry populations..... 1797

PB2677. Predicting functional and fitness effect of missense variants..... 1798

PB2678*. Profiling Local Ancestry Effects on Gene Regulation in An Admixed Population..... 1799

PB2679*. Quantifying negative selection in human 3'UTRs uncovers deleterious non-coding genetic variation. 1800

PB2680. Rare protein-coding variants in the insulin receptor gene (*INSR*) identified in women with polycystic ovary syndrome (PCOS)..... 1801

PB2681. Regulation of stress granules by VCP in health and genetic degenerative diseases 1802

PB2682. Regulatory rare variant trait associations in 127,724 UK Biobank genomes..... 1803

PB2683. Role of SNPs of Leptin gene in Etiology of Venous Thromboembolism among Females contraceptive users in Pakistan Population..... 1804

PB2684. Scalable mQTL analysis of biochemical pathways with the UK Biobank using REVEAL:Biobank 1805

PB2685. Seamless integrative pipeline for QTL datasets enhance the discovery of putative causal variants for Alzheimer's Disease. 1806

PB2686. SEPT-GD: a decision tree to prioritise potential RNA splice variants in cardiomyopathy genes for functional splicing assays in diagnostics..... 1807

PB2687. Sex differences in gene expression affect host response to influenza A infection..... 1808

PB2688. Short chain fatty acid abnormalities:Genotype-Phenotype correlation in thirteen Puerto Rican patients. 1809

ASHG 2022 Annual Meeting Poster Abstracts

PB2689*. Single-cell caQTL reveals cell-state-dependent regulation in immune cells in rheumatoid arthritis.	1810
PB2690. Single-nucleus transcriptome analysis of 424 aging brains identifies cell type and subtype specific <i>cis</i> -eQTL and their relation to neurodegenerative disease susceptibility.	1811
PB2691*. SMN modulates adverse effects associated with GEMIN5-mediated neurodevelopmental syndrome.....	1812
PB2692. Spectrum of pathogenic ATM variants in Indian ataxia telangiectasia patients.	1813
PB2693. Structural and functional characterization of missense variants in <i>NPR2</i> augments genetic associations, phenotypic prediction, and mechanistic understanding of skeletal development.	1814
PB2694. Systematic analysis of the impact of short tandem repeats on gene expression	1815
PB2695. Targeted-NGS across the β -globin gene cluster identifies multiple linkage disequilibrium patterns in north Indian non-transfusion dependent β -thalassemia patients	1816
PB2696. Testing the association of Grantham score with functional impact for clinical variant interpretation.....	1817
PB2697. The contribution of mitochondrial DNA variation to disease risk in a diverse cohort	1818
PB2698. The contribution of short tandem repeats to splicing variation in humans	1819
PB2699. The effect of rare protein-coding sequence variants on the human plasma proteome	1820
PB2700. The genetic architectures of gene expression in individuals of African and European ancestry: results and consequences of eQTL studies	1821
PB2701. The human myocardial transcriptomic response to the interaction effect of genetic variants and ischemic stimulus identified by a novel approach.....	1822
PB2702. The importance of the use of genomics in early diagnosis of tuberous sclerosis: case report.	1823
PB2703. The Multiple <i>de novo</i> Copy Number Variant (<i>MdnCNV</i>) phenomenon: peri-zygotic DNA mutational signatures and multilocus pathogenic variations	1824
PB2704*. The role of introgressed archaic DNA in the human genome.....	1825
PB2705. The Unmasking Of ‘Mitochondrial Adam’ And Structural Variants Larger Than Point Mutations As Stronger Candidates For Traits, Disease Phenotype And Sex Determination.....	1826
PB2706. <i>TP53</i> codon 72 polymorphism modulates macrophage polarization through altered PI3K/Akt signaling pathways.....	1827
PB2707. Two types of variants pointing out a crucial role for RNF216 in healthy and diseased brain.	1828
PB2708. TxEWAS identifies genetic modifiers of drug response and side effects through retrospective gene-environment association studies.....	1829
PB2709. Types of cis- and trans-gene regulation of expression quantitative trait loci across human tissues.....	1830
PB2710*. UBA5-related Epilepsy: From Cellular Models to Novel Therapies	1831
PB2711. UK Biobank whole genome sequencing and large-scale proteomics identifies novel protein quantitative trait loci (pQTLs) contributing to human disease.	1832

ASHG 2022 Annual Meeting Poster Abstracts

PB2712. Unbiased linkage of SNV to expression in single cells for regulatory variant discovery influencing cancer drug resistance.....	1833
PB2713. Uncovering sources of human gene expression variation in a globally diverse cohort.	1834
PB2714. Unsupervised Density Estimation for Noncoding Variant Effect Prediction.....	1835
PB2715. Use of reference population data for resolving variants of uncertain significance in hematopoietic genes.....	1836
PB2716. Using massively parallel reporter assays to dissect context-specific regulatory grammars in type 2 diabetes.....	1837
PB2718. Variant-to-gene mapping at the insomnia <i>WDR90</i> locus and subsequent luciferase assay analyses implicates <i>PIGQ</i> as an effector gene in sleep dysregulation.....	1838
PB2719. Variation of 5'untranslated regions by dosage sensitivity reveals differences in post-transcriptional regulation important for interpreting variant effect.....	1839
PB2720. x-QTL: A mixture-model for identification and characterization of trans-eQTLs.....	1840
PB2721. A Pharmacogenetic Integration Solution for Clinical Decision Support Software.....	1841
PB2722. A systematic review and comparison of statistical methods for testing genotype-treatment interaction effects in trans-ethnic pharmacogenomics GWAS.	1842
PB2723. Accurate CYP2D6 star (*) allele diplotyping for long-read PacBio HiFi sequencing.....	1843
PB2724. Allelic diversity of the pharmacogene <i>CYP2D6</i> in New Zealand Māori and Pacific peoples.	1844
PB2725. An NGS approach to detecting novel LOF variants in <i>DPYD</i> across >69,000 individuals .	1845
PB2726. Clinical genetic testing and reporting of homoplasmy or low-level heteroplasmy for <i>MT-RNR1</i> m.1555A>G conferring an increased risk for aminoglycoside induced ototoxicity and irreversible hearing loss.	1846
PB2727. Convergence of bipolar disorder treatments and gene knockdown on the transcriptome....	1847
PB2728. CRISPR enrichment of complex <i>CYP2D6-D7-D8</i> loci allows for accurate detection of <i>CYP2D6</i> structural variation and phased haplotyping.....	1848
PB2729. DEEPCT(Deep-learning-based Clinical Trial) simulates a drug clinical trial by exploiting germline variants as random assignment of drug.....	1849
PB2730*. Dynamics in blood transcriptomic networks responding to TNF inhibitors in patients with rheumatoid arthritis.....	1850
PB2731. Elucidating the multi-omic drivers of hepatic drug metabolism in African Americans.....	1851
PB2732*. Estimating UK Biobank population-specific PGx allele and phenotype frequencies using PharmCAT.....	1852
PB2733. Evaluating the frequency and the impact of pharmacogenetic variants in an ancestrally diverse Biobank population.....	1853
PB2734. Exploring polygenic risk scores as a biomarker of treatment response in systemic lupus erythematosus.....	1854
PB2735. Genetic associations of iron absorption among individuals of East Asian and European descent.	1855

ASHG 2022 Annual Meeting Poster Abstracts

PB2736. Genomic analysis of hydroxychloroquine-associated retinal toxicity in a large cohort	1856
PB2737. Impact of global and local ancestry on pharmacogenetics in African American persons with multiple sclerosis.	1857
PB2738. Implementation of clinical pharmacogenomics: How NCBI's Medical Genetics Summaries and MedGen can help educate and support clinicians.	1858
PB2739. Investigating pharmacogenetics of dupilumab in the treatment of atopic dermatitis in an Asian clinical cohort.....	1859
PB2740. LOF <i>CYP2C19</i> allele evaluation identifies significant number of Indian CAD patients are low responders to Clopidogrel.....	1860
PB2741. Meta-analyses of genome wide association studies of toxicity to oxaliplatin and fluoropyrimidine chemotherapy in 1800 patients with advanced colorectal cancer.....	1861
PB2742*. Novel genetic signals identified for angiotensin-converting enzyme inhibitor-induced cough	1862
PB2743. Pharmacogenetics of tenofovir clearance among Southern Africans living with HIV.	1863
PB2744. Pharmacogenomic analysis of drug metabolizing enzymes associated with plasma clozapine/N-desmethylclozapine ratio.	1864
PB2745. Pharmacogenomics for All of Us: Approach to testing and return of results in 1M participants	1865
PB2746. Predicted expression of enzymes in the thiopurine metabolic pathway and risk of side effects in patients with inflammatory conditions treated with azathioprine.....	1866
PB2747. <i>PTPN2</i> and Discontinuation of Azathioprine Attributed to Myelotoxicity.....	1867
PB2748. Testing for common pharmacogenomic predictors of vigabatrin-associated visual field loss, under univariate & polygenic models.	1868
PB2749. The Alabama Genomic Health InitiativeIntegrating Genomics into Primary Care	1870
PB2750. Vaccinations against pneumonia and the flu may protect carriers of particular genotypes against Alzheimer's disease.....	1871
PB2751. Variant Calling and Orthogonal Verification of 20 Pharmacogenomics (PGx) Clinical Pharmacogenetics Implementation Consortium (CPIC) Level A Genes including 4 Challenging Genes Using Illumina DRAGEN Bio-IT Platform.	1872
PB2752. Variation in <i>CYP2A6</i> , a nicotine metabolism gene, associates with lung diseases: A Phenome-wide Association and Mendelian Randomization study.	1873
PB2753*. Whole Exome Sequencing adds power to prioritize drug targets when combined with other genetic data types.....	1874
PB2754. A Bayesian framework to capture the probability of polymorphism with sequence context across the human genome.	1875
PB2755. A likelihood-based framework for demographic inference from genealogical trees.....	1876
PB2756. A robust kinship inference approach based on machine learning method for missing person identification	1877
PB2757. Addressing population stratification in GWAS with variational autoencoders.	1878

ASHG 2022 Annual Meeting Poster Abstracts

PB2758. Analysis of evolutionary pressure on gene expression identifies patterns of selection in tissues and disease	1879
PB2759. Analysis of human disease variants from ancestrally diverse Asian genomes	1880
PB2760. Analysis of Latvian genome diversity combining the whole-genome sequencing with genome-wide genotyping data.	1881
PB2761. Ancestral genetic tracing defies socio-cultural structure in the north western human populations of India	1882
PB2762. Ancestral haplotype prediction using neural networks	1883
PB2763. Ancient trans-species polymorphism at the Major Histocompatibility Complex in primates	1884
PB2764. Ancient virome analyses of ancient individuals who lived in the Japanese archipelago	1885
PB2765. Between-population genetic differences for human complex traits.....	1886
PB2766. Biological functions and spatial distribution of rapidly evolving brain-expressed genes in the human lineage	1887
PB2767. CAFE: Ancestry and genotype calling uncertainty-adjusted ancestry-specific allele frequency estimation from admixed samples.....	1888
PB2768. Cardiovascular associations and signals of selection at the <i>Adrenergic Receptor Alpha-1A</i> gene (<i>ADRA1A</i>) in high-altitude Andeans.....	1889
PB2769. Challenging continental groupings of humans using a data-driven approach.....	1890
PB2770. Characterizing the origin of African ancestral tracts in admixed individuals of the UK Biobank	1891
PB2771. <i>Cis</i> -regulatory genetic variants influence distinct expression profiles of adipose and muscle tissue transcripts between African and European ancestry individuals.	1892
PB2772. Coding variants associated with healthy aging: Findings from the Long Life Family Study (LLFS)	1893
PB2773. Comparing human centromeres with TandemAligner	1894
PB2774. ConsensuSV - HPC-ready, ML-enhanced automated pipeline for Illumina-based variant detection.....	1895
PB2775. Conservation and function of the <i>NXF2</i> palindrome in male fertility.....	1896
PB2776. Continuous measures of genetic diversity in biomedical research.....	1897
PB2777*. Contrasting patterns of positive selection at <i>Plasmodium falciparum</i> and <i>Plasmodium vivax</i> associated loci in Oman.....	1898
PB2778. <i>De novo</i> assembled and phased human genomes from Persian Arab trios show divergent and novel sequence versus CHM13 and GRCh38, providing valuable population specific reference genomes for middle eastern region.	1899
PB2779. Detection of Genomic Regions undergoing Selection in post-Neolithic Transition.....	1900
PB2780. Did Neanderthals have different 3D genome folding from modern humans?	1901
PB2781. Discovering determinants of short tandem repeat stability in local sequence context using gradient-based attribution scores.	1902

ASHG 2022 Annual Meeting Poster Abstracts

PB2782. Disease-associated STR loci in diverse populations 1903

PB2783. Evaluation of imputation performance of different reference panels in a Pakistani population.
..... 1904

PB2784. Evidence of potential natural selection in African American individuals post admixture... 1905

PB2785. Examining polygenic adaptation in time-stratified genome samples..... 1906

PB2786. Exploring population health disparities through identity by descent genetic communities in
The Biobank at The Colorado Center for Personalized Medicine 1907

PB2787. Extensive sex-biased gene flow in Native Americans across the Andes 1908

PB2788. Fine-scale ancestry mapping revealed population-specific associations with complex disease
..... 1909

PB2789. Generation of synthetic genotypes with neural networks. 1910

PB2790*. Genetic and phenotypic high-altitude adaptation at*EPAS1/HIF2A* in Andean highlanders
..... 1911

PB2791. Genetic landscape and demographic history of Britain and Ireland 1913

PB2792. Genetic variation related to the regulation and expression of hemoglobin isoforms among
global populations..... 1914

PB2793. Genome-wide patterns of tandem repeat abundance across species are primarily driven by
retrotransposons 1915

PB2794*. Genomic deserts: A global survey of underrepresented populations in genomics research
..... 1916

PB2795. Genomic variation and chromatin structure across the human-specific NOTCH2NL
duplications..... 1917

PB2796*. Germline genetic determinants of T-cell fraction in ~110,000 diverse individuals..... 1918

PB2797. Guaranteeing unbiasedness in selection tests based on polygenic scores. 1919

PB2798. Haplotype-based analyses of phylogeny and regional genome diversity among laboratory rats
using the latest rat reference genome..... 1920

PB2799*. High coverage whole genome sequencing of the expanded 1000 Genomes Project cohort
including 602 trios 1921

PB2800. Human populations and demographics in Qatar from the Neolithic to the Late Iron Age. . 1922

PB2801. Improved detection of evolutionary selection highlights potential bias from different
sequencing strategies in complex genomic-regions..... 1923

PB2802. Imputation around the world: Assessing imputation quality across diverse global populations.
..... 1924

PB2803. Imputation of ancient genomes. 1925

PB2804*. Inference of natural selection on epigenetic marks: implications for the evolution of gene
regulation and germline mutation rates..... 1926

PB2805. Inference of the distribution of fitness effects using local genealogical trees..... 1927

ASHG 2022 Annual Meeting Poster Abstracts

PB2806. Inferences about archaic introgression and positive selection from Lithuanian whole-genome sequences	1928
PB2807. Inferring between-population genetic variation using admixed populations: challenges and limitations.	1929
PB2808*. Inferring germline CpG methylation signature accumulated along the human history from genetic variation catalogs.....	1930
PB2809. Insights into non-equilibrium population genetics under strong natural selection from a new theoretical formalism	1931
PB2810. Integrative approaches linking divergent selection to disease association studies.....	1932
PB2811. Introgressed sequences and positive selection profiles of the Japanese population revealed by the analysis of 3,256 Japanese whole-genome sequencing (WGS) data.....	1933
PB2812. Investigating human migrations and genetic diversity in Northeast India over the last ~4,000 years using ancient genomics.....	1934
PB2813. Investigating selection and the genetics of height in Central African Rainforest Hunter-gatherers.....	1935
PB2814. Kidd Lab 55 AISNP panel shows a better performance than the SNPforID 34-plex for genetic ancestry estimates in a Brazilian population sample.....	1936
PB2815. Large-scale analysis of 6,000 whole genomes from Qatar uncovers genetic structure of Arab and Middle Eastern populations and establishes a valuable resource for understanding personalized disease risk and causality	1937
PB2816. Late Night Eating or Circadian disruption could be more dangerous than our think Late Night Eating or Circadian disruption could be more dangerous than our think.....	1938
PB2817. Leveraging Base Pair Mammalian Constraint to Understand Genetic Variation and Human Disease	1939
PB2818. Long-range regulatory effects of Neandertal DNA in modern humans.....	1940
PB2819. Lower allele frequency among males is a marker for pathogenicity among non-pseudoautosomal X-chromosome variants	1941
PB2820*. <i>MaLAdapt</i> reveals novel targets of adaptive introgression from Neanderthals and Denisovans in worldwide human populations.....	1942
PB2821. Modeling recurrent mutations predicts allele frequencies and enables precise inference in large samples.....	1943
PB2822. Mosaic chromosomal alterations and longevity.....	1944
PB2823*. Neanderthal introgression modifies anti-inflammatory response in modern humans.....	1945
PB2824. Newfoundland and Labrador: A mosaic founder population of an Irish and British diaspora from 300 years ago.....	1946
PB2825. No reduction in diagnostic yield of exome sequencing in prenatal and pediatric patients with non-European ancestries.	1947
PB2826. On the Genes, Genealogies, and Geographies of Quebec.....	1948
PB2827. Patterns of selection on human gene regulatory variation	1949

ASHG 2022 Annual Meeting Poster Abstracts

PB2828. Phased, long read assemblies from Central African participants represent a more comprehensive, inclusive future for human pangenomics	1950
PB2829. Population-specific Recombination Maps Reduce False Positive Findings in Genomic Scans of Adaptations	1951
PB2830. Presenilin-1 G206A mutation in Puerto Ricans with Alzheimer Disease reveals it is as a founder event on the African haplotype.....	1952
PB2831. Primate comparative genomics and infectious diseases; COVID-19, Monkey Pox and what comes next	1953
PB2833. Recovering signatures of ghost admixture using ancestral recombination graphs.....	1954
PB2834. Relatedness inference up to 3 rd degree from low coverage ancient genomes in presence of contamination, long runs of homozygosity and ascertainment bias.....	1955
PB2835. Relating the evolutionary fitness costs of loss-of-function mutations to their pathogenic consequences in humans	1956
PB2836. Repeats R/T ratio is associated with Human Y chromosome Haplogroups.....	1957
PB2837. Resolution of structural variation in diverse mouse genomes reveals dynamic chromatin remodeling due to transposable elements	1958
PB2838. Resurrecting the alternative splicing landscape of archaic hominins using machine learning.	1959
PB2839. Risk of venous thromboembolism in individuals with supernumerary sex chromosome aneuploidies in two large population-based cohorts	1960
PB2840. SALAI-Net: Species-agnostic local ancestry inference network	1961
PB2841. Sequence variants affect the genome-wide rate of germline microsatellite mutations.	1962
PB2842. Sequencing of 230 primate species identifies conserved enhancers underlying complex human disease	1963
PB2843. Signatures of Mutational Processes in Human DNA Evolution	1964
PB2844. Simple scaling laws control the genetic architectures of complex traits.....	1965
PB2845. Simulating effects from genetic and experimental perturbations to gene regulatory networks	1966
PB2846. Simulating Genome-phenome Dataset of 1,000,000 Individuals for the European 1+ Million Genomes Initiative	1967
PB2847. South Asian and Greater Middle Eastern Enrichment of <i>NOTCH3</i> p.Arg1231Cys Variant and Associated Stroke Risk	1968
PB2848. Structure-function analysis of a tRNA-derived SINE.....	1969
PB2849. The differential identification of cross-species lung scRNA expression via integrative analysis	1970
PB2850. The genetic diversity of Japanese populations inferred from the whole genome sequencing data.....	1971
PB2851. The genetic making of Bangladesh	1972

ASHG 2022 Annual Meeting Poster Abstracts

PB2852. The genomic history of human populations in the North American Central and Southern Plains.....	1973
PB2853. The impact of background selection on deleterious alleles and its consequences on trait evolution and DFE inference	1974
PB2854. The influence of demographic history and genetic architecture on complex phenotypes via runs of homozygosity.....	1975
PB2855. The landscape of tolerated genetic variation in humans and primates.....	1976
PB2856*. The lingering effects of Neanderthal introgression on human complex traits	1977
PB2857. The NHGRI Sample Repository for Human Genetic Research: biospecimens and a new genomic data search tool.....	1978
PB2858. The spatial distribution of rare deleterious alleles: implications for study design in human genetics	1979
PB2859. Tracing the genetic link of Mizo people of India: Genetic & archaeological evidence.....	1980
PB2860. Understanding ancestry-specific disease allelic effect sizes by leveraging multi-ancestry-matched single-cell RNA-seq and GWAS datasets	1981
PB2861*. Understanding natural selection in the European Holocene using Ancient DNA.....	1982
PB2862. Understanding the structure of regional populations with founder effects.	1983
PB2863. Unsupervised Learning of Ancestry Informative Markers and Genetic Admixture Coefficients From Biobank Data.....	1984
PB2864. Using a population-specific reference panel improves genotype imputation accuracy in individuals of African ancestry	1985
PB2865. Variable representation of duplicated sequence in recent canine genome assemblies.....	1986
PB2866. A bioinformatics pipeline to aid in the design of allele-specific antisense oligonucleotides for patients with rare genetic conditions.....	1987
PB2867. A Cloud-scalable Software Suite for Large-Scale Proteogenomics Data Analysis and Visualization	1988
PB2868. A ComparativeBioinformatics Analysis to Propose a COVID_19 Candidate Vaccine Epitope at the ACE-Bionformatics in Mali (West Africa).....	1989
PB2869*. A comparison of the sensitivity and diagnostic rate of long-read vs. short-read whole genome sequencing in critically ill infants in the SeqFirst Project.....	1990
PB2870*. A computational approach to design a COVID-19 vaccine against a predicted SARS-CoV-2 variant: high immunogenicity, efficacy and safety of DELLERA vaccine.	1991
PB2871. A deep learning framework for structural variant discovery and genotyping.....	1992
PB2872. A Deep-learning based RNA-seq Germline Variant Caller	1993
PB2873. A demonstrated workflow to unleash the multiomic potential of a single blood draw.....	1994
PB2874. A framework for evaluation of new or modified sequencing technologies for use in human genomics	1995
PB2875. A High Throughput Dual Nucleic Acid Extraction Method Enables Whole Genome Sequencing from PAXGene Blood RNA Tubes.....	1996

ASHG 2022 Annual Meeting Poster Abstracts

PB2876. A Hybrid Machine Learning and Regression Method for Cell Type Deconvolution of Spatial Barcoding-based Transcriptomic Data.....	1997
PB2877*. A Machine Learning-based approach to extract the gene-disease association discovery information from OMIM.....	1998
PB2878*. A massive proteogenomics screen identifies thousands of novel human coding sequences	1999
PB2879. A method for ABO genotyping by Sanger DNA sequencing	2000
PB2880. A multi-omics approach to study monozygotic twins discordant for amyotrophic lateral sclerosis.....	2001
PB2881. A new approach to assess the allele frequency of small insertions and deletions.....	2002
PB2882*. A new framework for efficient Perturb-seq enables cheap large-scale dissection of the innate immune response and provides insight into regulatory eQTL relationships.....	2003
PB2883*. A new method for detecting genes differentiated in expression variance using single-cell RNA sequencing data	2004
PB2884. A novel auto-normalizing pooled library construction method that streamlines sample preparation prior to targeted hybrid capture.....	2005
PB2885. A novel Bayesian factor analysis method improves detection of genes and biological processes affected by perturbations in single-cell CRISPR screening.....	2006
PB2886. A novel enzymatic fragmentation library preparation workflow that prevents sequencing artifacts for production-scale DNA-seq	2007
PB2887. A platform for high-resolution morphology analysis reveals tumor heterogeneity and enables label-free enrichment of target cell subpopulations	2008
PB2888. A rigorous benchmarking of methods for SARS-CoV-2 lineage detection in wastewater..	2009
PB2890. A software platform for real-time and automated simultaneous analysis and detection of genetic diseases.....	2010
PB2891*. A spatial map of neurodevelopmental disorder risk in the developing human cortex	2011
PB2892. A submodular fairness metric for sequence-to-function models in precision genomics.....	2012
PB2893. A systems biology framework to evaluate the contribution of cellular crosstalk in Alzheimer's disease genetic risk and other psychiatric traits	2013
PB2894. A two-channel deep learning framework for accurate prediction of the change in protein folding free energy upon mutations	2014
PB2895. A unified computing environment for genomics data storage, management, and analysis: NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space (AnVIL).....	2015
PB2896. A validated clinical whole genome sequencing system for the detection of germline variants.	2016
PB2897. Age-dependent accumulation of somatic mutation and cell fusion in human cardiomyocytes revealed by single-cell whole-genome sequencing.....	2017
PB2898. Algorithmic and Assay-based Simplification of Multiallelic Variants for Genotyping.....	2018

ASHG 2022 Annual Meeting Poster Abstracts

PB2899*. An Algorithm for Sequence Location Approximation using Nuclear Families (ASLAN) Validates Regions of the Telomere-to-Telomere Assembly and Identifies New Hotspots for Genetic Diversity.....	2019
PB2900*. An algorithmic framework for isoform-specific functional analysis	2020
PB2901. An empirical comparison of calling pipelines for whole genome sequencing demonstrates the importance of mapping and alignment.....	2021
PB2902*. An evaluation of the Ultima Genomics sequencing platform: Scalable, high-throughput sequencing for low-cost whole genome sequencing.....	2022
PB2903. Analysis of alternative polyadenylation from long-read or short-read RNA-seq with LAPA	2023
PB2904. Analysis of pathogenic tandem repeat variation in congenital disorders.....	2024
PB2905. Application of a novel scRNA-seq method based on Pre-templated Instant Partitions (PIPseq) to evaluate therapeutic induced intestinal epithelium regeneration	2025
PB2906. Application of DRAGEN Graph read alignment to challenging medically relevant genes and other difficult regions in GRCh38 and T2T-CHM13 genomes.	2026
PB2907. ASCoT - Application for Sanger COncfirmatory Testing - A graphical software interface for management of confirmatory sequencing in genomic medicine.....	2027
PB2908. Assessing the impact of bioinformatics tools on genomic reproducibility: opportunities and pitfalls.	2028
PB2909. Assessing the utility of genomic deep learning models for disease-relevant variant effect prediction	2029
PB2910. Assessment of Whole Genome Sequencing Quality Metrics from a Large Cohort of Saliva-derived DNA Samples.	2030
PB2911. Associations between genetically predicted levels of blood metabolites and pancreatic cancer risk	2031
PB2912. Autoencoder-based Model for Genotype Imputation.....	2032
PB2913. Automated Walk away NGS Sample Preparation Using a Flexible Library Prep System with Enhanced Error Correction	2033
PB2914. Automation of single-cell NGS library preparation for Tapestry platform to accelerate oncology discoveries.....	2034
PB2915*. AUTOSurv: Interpretable autoencoder aided deep learning framework for breast cancer survival analysis incorporating multi-omics data	2035
PB2916*. Bayesian Causal Inference Method applied to CRISPR perturbations Estimates the Causal Gene Regulatory Network of CD4+ T cells.....	2036
PB2917. Benchmarking algorithms for joint integration of unpaired and paired single-cell RNA-seq and ATAC-seq data.....	2037
PB2918. Benchmarking NGS Integration Site Analysis Methods in Support of Long-Term Safety Monitoring of Cell and Gene Therapy Products.....	2038
PB2919*. Benchmarking of computational methods for ancestry prediction from RNA-seq data. ...	2039

ASHG 2022 Annual Meeting Poster Abstracts

PB2920. Benchmarking splice variant predictive algorithms with saturation screens.	2040
PB2921. Beyond the "one gene, one disease" paradigm with DIVAs, an Explainable AI tool for phenotype-driven digenic variants interpretation.....	2041
PB2922. Biallelic CORO1A Mutation in a Patient with Late-onset Cutaneous Tuberculosis, Epidermodysplasia Verruciformis, and StaphylococcusAureus Infection.....	2042
PB2923*. Blended Genome Exome (BGE) Sequencing as an Alternative to SNP Arrays for Cost Effective Imputation of Variants in Globally Diverse Cohorts.....	2043
PB2924. Brain transcriptomic profiling of Parkinson´s disease patients reveals disease stage specific gene expression patterns.	2044
PB2925. Breast Cancer Specific Enhancer-Interactome Analysis Identifies Candidate Target Genes Enriched in Common Variant Risk Regions.....	2045
PB2926. Building a transcriptomic atlas of adult human ovarian tissue to gain insight into folliculogenesis.	2046
PB2927. bulk and single-cell gene co-expression network analysis of AD using PyWGCNA.....	2047
PB2928. Calling structural variants in extended rat pedigree using PacBio HiFi sequencing.	2048
PB2929. Cas9 targeted enrichment and adaptive sampling for characterizing repetitive elements....	2049
PB2930. Causal relationship analysis of DNA openness and RNA expression at the single cell level using scRNA-seq and scATAC-seq data.	2050
PB2931. CCC: An efficient not-only-linear correlation coefficient based on machine learning.....	2051
PB2932. Cell manufacturing genome integrity analysis by optical genome mapping.	2052
PB2933. Characterizing the genetic etiology of patients with neuromuscular disease through the integration of omics data.....	2053
PB2934. Characterizing the immune dysfunction in X-CGD using single-cell transcriptomics.	2054
PB2935. Circulating proteomic signatures of pulmonary function.....	2055
PB2936*. Clinical genome and RNA sequencing with a novel, cheaper sequencing-by-synthesis technology.....	2056
PB2937*. Clinical use of the amniotic fluid cells transcriptome in deciphering Mendelian disease .	2057
PB2938*. Closing the gap: Solving complex medically relevant regions of the human Genome.....	2058
PB2939. Combination of annotation enrichment approaches for characterization of blood-based RNA sequencing biomarkers.....	2059
PB2940. Combination of low-coverage whole genome and deep-coverage exome sequencing is the cost-effective way to drive large-scale genetics studies forward.....	2060
PB2941. Combined multiple structural variant algorithms achieved the highest recall rate in GIAB large deletion callsets.....	2061
PB2942. Community detection analysis in multilayer COVID-19 patient similarity networks.....	2062
PB2943. Comparison of exome performance in an inherited cancer context.....	2063
PB2944. Comparison of genotyping arrays and low-pass sequencing for predictive genomics research.	2064

ASHG 2022 Annual Meeting Poster Abstracts

PB2945. Competing sampling biases balance concordance of single-cell and nucleus RNA-seq	2065
PB2946. Comprehensive characterization of structural variants in human brain genomes using single flow cell nanopore sequencing protocol	2066
PB2947. Comprehensive detection of trait-associated structural variations using short read sequencing data	2067
PB2948. Comprehensive discovery of CRISPR-targeted sequences in the human gut metagenome	2068
PB2949. Comprehensive Structural Variant Detection: From Population to Mosaic level.....	2069
PB2950. CRISPR Streamline: Creating a Pipeline That Evaluates Phenotype Information From Off-Target Sequence Identification	2070
PB2951. Cryptic splice variants contribute to the genetic etiology of congenital heart disease.....	2071
PB2952. Custom morphology markers allow better tissue stratification to study tumor heterogeneity using NanoString® GeoMx® Cancer Transcriptome Atlas	2072
PB2953. D-CoMEx: Differential co-expressed module extraction to identify phenotype-specific miRNA biomarkers related to diseases.	2073
PB2954. <i>De novo</i> variant detection with HiFi reads.....	2074
PB2956. Deep immune repertoire analysis of SARS-CoV-2 antibodies from a vaccinated human patient via mass-spectrometry and synthetic long reads.	2075
PB2957. Deep learning approaches for predicting virus integration sites in the host genomes	2076
PB2958. Deep plasma protein profiling of Alzheimer’s subjects with a novel unbiased and scalable proteogenomics approach	2077
PB2959. DeepCell, A novel cell type specific marker detection machine learning tool for single cell transcriptome.....	2078
PB2960. DeepConsensus v0.3: Gap-aware sequence transformers for sequence correction.	2079
PB2961. <i>DeepLoop</i> robustly maps chromatin interactions from sparse allele-resolved or single-cell Hi-C data at kilobase resolution	2080
PB2962. DeepMod2: A deep learning framework for DNA and RNA modification detection using Oxford Nanopore sequencing.	2081
PB2963. Delivering Comprehensive, Single-molecule Proteomics Using Protein Identification by Short-epitope Mapping (PrISM).....	2082
PB2964. Depth normalization for single-cell genomics count data.....	2083
PB2965. Detecting and assembling non-reference LINE-1 insertions using clustered long reads.	2084
PB2966. Detection of 5-hydroxymethylcytosine using a modified EM-seq protocol.	2085
PB2967. Detection of somatic variation in human neurons using long-read sequencing approaches	2086
PB2968. Developing local compute capacity - a Ugandan experience Integrated Biorepository of H3Africa Uganda, Makerere University College of Health sciences.	2087
PB2969. Development of computational methods for analyzing single-cell spatial transcriptomic data with applications to murine spermatogenesis.	2088

ASHG 2022 Annual Meeting Poster Abstracts

PB2970. Diagnosis of neurodevelopmental disorders by RNA-Seq : From urine derived stem cells to induced neural stem cells, a promising alternative tissue.	2089
PB2971. Diagnostic metabolomic profiling of hypomyelinating leukodystrophy caused by DEGS1 deficiency.....	2090
PB2972*. Direct haplotype-resolved 5-base HiFi genome sequencing allows for linking rare disease variants to non-coding function	2091
PB2973. Magnified Convolutional Enrichment Representation Model	2092
PB2974. Efficiency of mutagenesis promoted by CRISPR-Cas9 in <i>THRAB</i> gene in zebrafish.	2093
PB2975. Efficient DNA sample contamination metric estimation using a novel variant representation and algorithm.	2094
PB2976. Empowering Discovery in Childhood Cancer: Genomic Harmonization at the Kids First Data Resource Center	2095
PB2977. Enabling flexible low throughput sample preparation for multiple sequencing platforms using the Miro Canvas	2096
PB2979. Ensure sample identity in sequencing workflows with the Twist Sample ID Kit.....	2097
PB2980. Establishing a baseline transcriptomic profile of human eccrine sweat glands via single-cell RNA-sequencing with validation by 3D imaging.....	2098
PB2981. Establishing best practice for structural variant discovery with long read sequencing in the Gabriella Miller Kids First Pediatric Research Program using Sentieon’s haplotype-resolved variant evaluation tool.....	2099
PB2982. Estimating and testing cell-type-specific gene co-expression from single-cell data: A robust and efficient method that mitigates technical confounding.	2100
PB2983. Evaluating deep learning for predicting epigenomic profiles	2101
PB2984. Evaluation of cell type annotation for peripheral blood mononuclear cells using single-cell RNA versus a deconvolution approach with bulk RNA sequencing.	2102
PB2985. Evaluation of single nuclei RNA sequencing in brain tissue using Pre-templated Instant Partitions (PIPseq)	2103
PB2986. Evercode™ Whole Transcriptome Single-Cell RNA-Seq with Focal Barcoding Enables Single-Cell Transcriptomes Coupled with Gene or Vector-Enriched Readouts.....	2104
PB2987. Exome sequencing reanalysis complemented with combined multi-omics approach reached to 60% diagnostic yield in previously undiagnosed rare disease cohort.....	2105
PB2988. Exome technology innovations advancing personalized medicine.....	2106
PB2989. Extensive differential cell type-specific gene expression and regulation by sex in human skeletal muscle.	2107
PB2990. Extracellular protein monitoring in the ResolveOME genomic and transcriptomic dual workflow to uncover cancer pathology mechanisms in single cells	2108
PB2991. faiGP: an evolutionary approach to discover governing equations in high-dimensional genomic data	2109
PB2992. Fast and gentle microfluidic cell sorting upstream of single cell transcriptomics	2110

ASHG 2022 Annual Meeting Poster Abstracts

PB2993. FastRNA: an efficient exact solution for PCA of single-cell RNA sequencing data based on a batch-accounting count model	2111
PB2994. FedExosome: mitochondrial DNA & microRNA profiling of neuronal-derived packages in Alzheimer's disease	2112
PB2995*. Fixing falsely duplicated and collapsed regions of the GRCh38 reference genome.....	2113
PB2996. Functional characterisation of the amyotrophic lateral sclerosis risk locus <i>GPX3/TNIP1</i> ..	2114
PB2997. Gene regulatory network inference using single-cell multiome ATAC-seq and RNA-seq data.	2116
PB2998. Generating long-range sequencing information without long-read sequencing.....	2117
PB2999. Generative Adversarial Networks Create High-Quality Synthetic Prostate Cancer Magnetic Resonance Images.....	2118
PB3001*. Genetics guided approach to infer unobserved covariates in functional genomics data....	2119
PB3002. <i>Genetics of Cardiovascular diseases: Mathematical Modeling & Development of a Predicting Simulator Prediction of the heart attack, the treatment to prescribe and the risk of secondary complications</i>	2120
PB3003. Genome-wide prediction of chromatin profiles from gene expression	2121
PB3004*. <i>Genome-wide prediction of pathogenic gain- and loss-of-function variants from ensemble learning of diverse feature set</i>	2122
PB3005. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases: an analysis of the Canadian Longitudinal Study on Aging (CLSA).....	2123
PB3006. Genotyping known disease-associated short tandem repeats in 7,000 Project MinE genomes reveals their relation to ALS and assists in disease-association analysis of unknown short tandem repeats.	2124
PB3007*. GestaltMatcher supports classification of ultra-rare disorders and delineation of novel syndromes by facial phenotype descriptors	2125
PB3008*. GTAGMe-seq reveals the genome-wide long-range coordinated cis-regulatory elements at the single-molecule level	2126
PB3009. HD-Seq: A novel method for high accuracy sequencing.....	2127
PB3010. High accuracy Sequencing By Binding (SBB) improves human germline variant calling .	2128
PB3011. High throughput single-cell epigenomic profiling through split-pool combinatorial barcoding.	2129
PB3012. High throughput workflow for human whole genome sequencing using PacBio HiFi.	2130
PB3013. High-throughput CRISPR Inhibition Screen to Improve Interpretation of Noncoding Variants in Developmental Epileptic Encephalopathies	2131
PB3014. High-throughput metabolomics enabling discovery of tens of thousands of small molecule biomarkers that link genetics and disease	2132
PB3015. High-throughput RNA isoform sequencing using programmable cDNA concatenation.....	2133
PB3016. High-throughput RNA sequencing directly from cell lysates enables reproducible phenotypic profiling for CRISPR treatment phenotyping and cell response screening applications	2134

ASHG 2022 Annual Meeting Poster Abstracts

PB3017. High-Throughput Sample Processing workflow for rapid Vaginal microbiota profiling of women's health samples.	2135
PB3018*. High-throughput targeted enrichment for long read third-generation sequencing: A flexible, scalable, and cost-efficient method achieving similar throughput as for short read NGS.	2136
PB3019. Horizon - an ultrarapid platform for NICU/PICU whole genome clinical pathogenicity prediction	2137
PB3020*. Human cytomegalovirus infection extensively re-organizes the human genome at human disease risk loci	2138
PB3021*. Human microglia transcriptional changes associated with progression and development of Alzheimer's Disease.	2139
PB3022. Identification of FZD3 and SEMA3A as potentially secreted proteins involved in proliferative sickle cell retinopathy.	2140
PB3023. Identification of novel long non-coding RNA with distinct expression patterns in different multiple myeloma subtypes.	2141
PB3024. Identifying high-confidence de novo mutations in somatic and germline cells through duplex sequencing of diverse tissue types.	2142
PB3025*. Illuminating Dark Proteins using Reactome Pathways	2143
PB3026. Improved data quality for renal cell carcinoma samples with HIVE scRNAseq integrated storage.	2144
PB3027. Improved detection of functionally relevant aberrant splicing using the intronic Jaccard index	2145
PB3028. Improved sequence mapping using a complete reference genome and lift-over	2146
PB3029*. Improving gene detection sensitivity from clinically relevant low input and FFPE RNA samples utilizing a rapid whole transcriptome library prep workflow.....	2147
PB3030. Improving sensitivity of combinatorial barcoding-based single cell sequencing.	2148
PB3031. Imputing functionally impactful <i>CYP2A6</i> structural variants from SNP array data.	2149
PB3032. Imputing the whole metagenomic shotgun sequencing data using the 16s amplicon sequencing data.....	2150
PB3033. Increased mitochondrial expression is associated with increased <i>APOEε4</i> expression and affected by ancestry and sex.	2151
PB3034. Increasing throughput and decreasing cost for single-cell profiling using combinatorial indexing	2152
PB3035. In-depth analysis and comparison of different whole genome methylation sequencing library preparations with low DNA input.....	2153
PB3036. Individualized cellular ancestry: efficient reconstruction of cell lineage trees	2154
PB3037. Induced pluripotent stem cells (iPSCs) for functional genetic screen applications.	2155
PB3038. Infection with SARS-CoV-2 leads to variant-specific changes in gene expression in airway cell lines and primary cell cultures.	2156

ASHG 2022 Annual Meeting Poster Abstracts

PB3039. Inference of menopausal status from female reproductive tissue gene expression signatures 2157

PB3040. Insights from HLA transcriptome analysis of a cohort from the Qatar Genome Program .. 2158

PB3041. Insights into Molecular Pathways Associated with Juvenile Myositis through Single-Cell Transcriptome Profiling 2159

PB3042. Integrated genetic analysis of transcriptome sequencing data in congenital diaphragmatic hernia. 2160

PB3043. Integrated Proteotranscriptomics Analysis of Breast Cancer using Variational Autoencoders to Prioritize Pathogenic Genes 2161

PB3045. Integrating Knowledge in Breast Cancer Subtyping Deep Learning Classifiers 2162

PB3046. Integrating multiple human toxicogenomics datasets for drug repositioning 2163

PB3047. Integration of single nucleus and bulk adipose RNA-seq reveals distinct adipocyte subtypes for obesity and insulin resistance. 2164

PB3048*. Integrative computational analysis of long-read transcriptomes identifies alternatively spliced poison exons in iPSC-derived brain organoids 2165

PB3049. Integrative single-nucleus multi-omics across 287 skeletal muscle biopsies reveals context-specific e/caQTL and extensive caQTL-GWAS colocalization 2166

PB3050. Integrative single-nucleus multi-omics analysis identified candidate regulatory elements and variants and their target genes in Alzheimer's disease brains 2167

PB3052. Investigating somatic mutation rates and signatures across the length of the colon with duplex sequencing..... 2168

PB3053. ioSearch: A tool for searching disease-associated interacting omics..... 2169

PB3054. Is single nucleus ATAC-seq accessibility a qualitative or quantitative trait? 2170

PB3055. IsoMiGA: Combining long and short read RNA-seq in human microglia reveals novel isoforms and splicing events and insight into Alzheimer's disease 2171

PB3056. Large-scale single-cell RNA-sequencing in Japanese identified biological and host genetic involvement of innate immune cells in COVID-19 severity..... 2172

PB3057. Learning nonlinear causal relations in complex biological systems using a deep-learning approach and knockoff statistics 2173

PB3058. Leveraging a multi-‘omics Parkinson’s Disease dataset to understand the role of mitochondrial transcriptional control in the aetiology of Parkinson’s Disease. 2174

PB3059. Linking GWAS risk variants to disease genes by epigenomic mapping and prediction of functional enhancer-promoter interactions. 2175

PB3060. Long read transcriptomes identify features not found with very deep short read sequencing 2176

PB3061. Long-amplicon variant-robust genome capture of SARS-CoV-2 using molecular inversion probes..... 2177

PB3062. Long-read capture with Twist target enrichment system 2178

ASHG 2022 Annual Meeting Poster Abstracts

PB3063. Long-read isoform sequencing reveals aberrant splicing of <i>PSEN2</i> , but not <i>PSEN1</i> , in individuals with sporadic Alzheimer’s disease.	2179
PB3064*. Long-read RNASeq in human brains aligned to T2T CHM13 complete human genome reveals new gene bodies and new transcripts, exons, & exon junctions in known genes.	2180
PB3065. LongReadSum: A fast and flexible quality control tool for long-read sequencing data.	2181
PB3066*. Low cost, noninvasive RNA-sequencing to enable massive scaling of transcriptome studies.	2182
PB3067. Low-pass whole genome sequencing as a cost-effective and improved alternative to genotyping arrays.....	2183
PB3068. Machine learning approach embedding an alternative splicing as a strategy for a multi omics integration.....	2184
PB3069. Major cell-types in imbalanced multiomic single-nuclei datasets impact statistical modeling of links between regulatory sequences & genes.....	2185
PB3070*. Making population-scale toxicogenomic analysis performant and cost-effective using REVEAL Biobank.	2186
PB3071. Massively parallel reporter perturbation assays uncover temporal regulatory architecture during neural differentiation.....	2187
PB3072. Max Read™: a novel high-throughput sequencing format for patterned flow cells.....	2188
PB3073. Metabolomic analysis of vitreous humor from the eyes with different uveal melanoma prognostic subtypes defined by tumor gene expression profiling.....	2189
PB3074. Metagenomic analyses coupled with metabolic and deep immune profiling reveal coordinate effects on host-microbe interactions in chronic kidney disease.....	2190
PB3075. Methods for screening candidate causal regulatory variants in primary immune cells by CRISPR ribonucleoprotein.	2191
PB3076. MHConstruct creates personalized full-length MHC locus haplotype sequences by finding optimal paths through a population variation graph.	2192
PB3077. MiRNA regulates potential genetic pathways involved in diabetic retinopathy: Implication for an early disease risk prediction.....	2193
PB3078. Mitigating challenges of large-scale single cell data management, querying, and analysis with REVEAL SingleCell.....	2194
PB3079. Mitochondrial Genomic ratio (mitscore) is an efficient biomarker to predict the implantation efficacy of human embryos.....	2195
PB3080. More bang for your buck! Universal Sequence Detection: A novel full spectrum variant genotyper for NGS data.....	2196
PB3081. Multi-Omics Integration for Osteoporosis Prediction through Mixture-of-Experts Framework.....	2197
PB3082. Multi-omics integration via similarity network fusion to detect subtypes of aging.....	2198
PB3083*. Multiset correlation and factor analysis enables exploration of multi-omic data.....	2199

ASHG 2022 Annual Meeting Poster Abstracts

PB3084. Neuroimmune insights through single-cell transcriptomics of paired brain and blood from living human subjects	2200
PB3085. New generative deep models to discover novel disease-gene associations in large-scale genomic cohorts	2201
PB3086. New genotyping and sequencing methods for high throughput molecular blood group typing	2202
PB3087. Novel Principal Component Analysis reveals rich gene expression contexts from snRNA-seq within and across cell-types.	2203
PB3088*. Novel sequencing platform using an open fluidics architecture and mostly natural chemistry for cost-efficient whole genome sequencing.....	2204
PB3089. NTSM: Fast sample swapping detection and ancestry estimation on unprocessed heterogeneous raw sequencing data	2205
PB3090. ODER: Optimising the Definition of Expressed Regions, a publicly available R package to improve annotation of RNAseq data and investigate novel transcription using short-read RNA-sequencing data.....	2206
PB3091. On statistical modeling of single-cell RNA-seq data accounting for technological and biological dropouts.....	2207
PB3092. Optimizing microbiome extraction methods for human intestinal resections.....	2208
PB3093. OrphaID: a new platform for rare intellectual disabilities in Orphanet in partnership with ERN-ITHACA	2209
PB3094. Outcomes from the Care4Rare Canada unsolved research pipeline for 950 patients with non-diagnostic exome sequencing data.....	2210
PB3095. Overcoming FFPE hurdles to enable high quality hybrid capture libraries and somatic mutation detection in matched tumor-normal patient samples.	2211
PB3096. PacBio HiFi sequencing provides highly accurate CpG methylation calls without bisulfite treatment.	2212
PB3097*. Paired whole genome and whole transcriptome sequencing of a large cohort of undiagnosed pediatric trios with neurological phenotypes	2213
PB3098. Papillary renal cell carcinoma (pRCC) functional heterogeneity unraveled by single nucleus sequencing (Sn-Seq) technologies	2214
PB3099. Peripheral leukocyte transcriptome dynamics following ischemic stroke of large vessel and cardioembolic etiologies	2215
PB3100. Perturb-seq analysis reveals key mediators of TNF α -induced transcriptional response.....	2216
PB3101. Peruvian Hospital<Urban Antibiotic Resistance found in Local Wastewater	2217
PB3102. PhaseDancer: Targeted assembly of the complex syntenic regions in non-human primates of the HSA2 fusion site.	2218
PB3103. Phenome-wide gene prioritisation leveraging Knowledge Graphs, Graph Convolutional Networks and UK Biobank PheWAS.	2219
PB3104. PIPseq, a novel and highly scalable technology based upon Pre-templated Instant Partitions (PIPs), is powering single-cell RNA sequencing into the million cell era.....	2220

ASHG 2022 Annual Meeting Poster Abstracts

PB3105. Population diversity and selection of recent gene duplications detected using a complete human genome sequence	2221
PB3106. Population Scale Genetic Interpretation Software for Reporting Pathogenic and Likely Pathogenic Variants Impacting the ACMG59 Genes.	2222
PB3107. Predicting the transcriptional activity and mechanism of action of small molecules using deep learning	2223
PB3108. Prokaryotic and viral genomes recovered from 787 Japanese gut metagenomes revealed microbial features associated with diets, populations, and diseases	2224
PB3109. Proteoform inference using proteogenomic approach in non-small cell lung cancer and healthy control plasma proteomes reveals disease-associated protein isoforms	2225
PB3110. Proteogenomic approaches improve genetic discovery across diverse human diseases.	2226
PB3112*. Pseudogenes limit the identification of common, functionally important transcripts generated by their parent genes	2227
PB3113. PubCaseFinder: Implementation of the new function designing Mondo-based virtual gene panels to increase diagnostic efficiency for rare genetic diseases.	2228
PB3114*. Pyro-Velocity: Probabilistic and scalable RNA velocity inference from single-cell data .	2229
PB3115. pyTCR: a comprehensive and scalable platform for TCR-Seq data analysis to facilitate reproducibility and rigor of immunogenomics research	2230
PB3116. Quantifying regional DNA methylation improves detection of biologically relevant associations in Alzheimer’s Disease	2231
PB3117*. Rapid whole genome and whole exome variant detection using a novel fluorescently labeled reversible terminated nucleotide sequencing system and GPU-based accelerated analysis	2232
PB3118. RedSpa: a method for Reference-based domain detection in multiple-sample Spatial transcriptomics studies	2233
PB3119. Regulome-wide association study identifies chromatin accessibility associated with pancreatic cancer risk	2234
PB3120. Revealing the sequence of neutralizing antibodies produced during infection using VyCAP B cell screening equipment coupled to an optimized single cell RNA sequencing protocol.	2235
PB3121. Rigorous benchmarking of HLA callers for RNA sequencing data	2236
PB3122. rMATS-turbo: An efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data.....	2237
PB3123. RNA monsters generated by human mitochondrial RNA polymerase	2238
PB3124*. Robust mapping of cell states from multiomics data using implicit feature selection.	2239
PB3125. SARS-CoV-2 genomic surveillance in Rwanda: Introductions and local transmission of the B.1.617.2/Delta variant of concern.	2240
PB3126. scNanoGPS enables high throughput single cell nanopore sequencing of same cell mega-omics in human tumors	2241
PB3127. Searching for Somatic Variation in Rapid-onset Obesity with Hypothalamic dysfunction, Hypoventilation and Autonomic Dysregulation (ROHHAD).....	2242

ASHG 2022 Annual Meeting Poster Abstracts

PB3128. SeeNV: a pipeline ready tool for the integration and visualization of CNVs and their callers	2243
PB3129. Sentieon DNAscope LongRead: a highly accurate, fast, and efficient pipeline for germline variant calling from PacBio HiFi reads.....	2244
PB3130. Sentieon DNAscope: high accuracy small variant calling using machine learning	2245
PB3131. Sequence-to-expression models of compact promoters for cell-type-specific promoter design	2246
PB3132. Sequencing By Binding (SBB) demonstrates superior performance in low-pass whole-human-sequencing applications	2247
PB3133. Sequencing By Binding (SBB) enables a lower limit of detection for Tuberculosis resistance genes <i>gyrA</i> and <i>katG</i>	2248
PB3134. <i>SigAlign</i> (Similarity-guided Align): A new, high-performance generalized pairwise alignment algorithm using simplified and intuitive regularization criteria.....	2249
PB3135. Simplified RNA-Seq Library Prep: Improved RNA sequencing results for FFPE and blood samples.....	2250
PB3136. Simulating pathogenic and likely pathogenic variants for bioinformatics pipeline testing .	2251
PB3137*. Simultaneous dimensionality reduction and cell-type annotation of single-cell RNA-seq data using marker enriched uniform manifold approximation and projection.	2252
PB3138*. Single cell dissection of Alzheimer's disease and related dementias.	2254
PB3139. Single cell RNA-seq analysis of Bone Marrow-derived Osteoblasts.....	2255
PB3140. Single neuron whole genome sequencing for somatic CNVs using three different genome amplification methods.....	2256
PB3141*. Single-cell allele-specific expression analysis reveals dynamic and cell-type-specific regulatory effects	2257
PB3142. Single-cell dissection of ALS and frontotemporal dementia in the human motor and prefrontal cortices.	2258
PB3143*. Single-cell genome-wide association reveals a nonsynonymous variant in <i>ERAPI1</i> confers increased susceptibility to influenza virus.	2259
PB3144. Single-cell multi-omics reveals gene regulation dynamics during mouse secondary palate development.....	2260
PB3145. Single-cell RNA sequencing (scRNA-seq) of fresh, neuronal tissue using Pre-templated Instant Partitions (PIPseq) and comparison to alternative scRNAseq methods.....	2261
PB3146. Single-molecule, modified base sequencing to identify frequency and cause of rAAV vector breakpoints.....	2262
PB3147. Single-nucleus profiling of the cerebellar cortex in essential tremor reveals Bergmann gliosis and oligodendrocyte myelin abnormalities as hallmarks of disease	2263
PB3148. Single-stranded library preparation for cfDNA identifies unique fragment length signature in myelodysplastic syndrome with multi-lineage dysplasia and ring sideroblasts.....	2264
PB3149*. SnapFISH: a computational pipeline to identify chromatin loops from DNA FISH data..	2265

ASHG 2022 Annual Meeting Poster Abstracts

PB3150*. Somatic genomic alterations in single neurons from brains with chronic traumatic encephalopathy (CTE)	2266
PB3151. sparQ mRNA-Seq: Consistent and high-quality mRNA library preparation from both abundant and limiting samples.....	2267
PB3152. STOC: A Simple Tool Of Compression	2268
PB3153. STRT-N: A newly optimised single-cell RNA sequencing method.	2269
PB3154. Structural and copy number variant detection, filtering, annotation, and classification by optical genome mapping in constitutional disorders.....	2270
PB3155. Structuring information via an immune-focused ontology enables the construction of a high-quality knowledge graph for the study of autoimmune diseases	2271
PB3156. SVPred: An integrated framework for Structural Variant Discovery	2272
PB3157. Systematic evaluation of transcriptome sequencing applied to real-world rare disease cohorts: Insights and limitations	2273
PB3158*. Tagmented, Indexed, and Pooled followed by ChIP Sequencing (TIP-ChIP) to generate high-throughput multi-target ChIP-Seq Results.....	2274
PB3159. Targeted sequencing of a 1 Mb carrier screening panel using molecular inversion probes.	2275
PB3160. Targeted sequencing of native telomeres reveals patterns of telomere length and subtelomere methylation at single chromosome resolution in human cells	2276
PB3161. Targeted transcriptome sequencing enables exponential scaling of combinatorial barcoding.	2277
PB3162. Tempo: an unsupervised Bayesian algorithm for circadian phase inference in single-cell transcriptomics.....	2278
PB3163. Temporal motif discovery in biological interaction networks	2279
PB3164. The impact of ageing on the transcriptome profile of human germline and somatic cells... ..	2280
PB3165. The impact of insert length on variant calling quality in whole genome sequencing.	2281
PB3166. Tissue specific transcript and protein isoforms in the human neural retina	2282
PB3167. Transcriptome analyses of congenital heart disease tissue from participants with Trisomy 21.	2283
PB3168. Transcriptome sequencing of RNA isolated from small volumes of blood stabilized in Tempus solution: a technical assessment of different extraction methods and DNase treatment.....	2284
PB3169. Transcriptome signature of sodium intake in and links to cardiovascular traits	2285
PB3170. Transcriptomic analysis of circulating endothelial colony-forming cells in patients with sickle cell anemia and ischemic stroke.....	2286
PB3171. Transcriptomic analysis reveals long-term effect of SARS-CoV-2 infection on host genetic regulation	2287
PB3172. Ultrasensitive Duplex Sequencing for Diverse Applications on the Singular Genomics G4 Platform.....	2288
PB3173. Ultra-sensitive TCR/BCR clonotyping and immunophenotyping.	2289

ASHG 2022 Annual Meeting Poster Abstracts

PB3174. Undergraduate research assistant	2290
PB3175. Untargeted metabolomics profiling in patients with and without epilepsy.....	2291
PB3176. Using an NGS readout for high-throughput proteome-wide analysis in large population health studies.	2292
PB3177. Using combinatorial barcoding to simultaneously profile the transcriptome and immune repertoire of 1 million T cells.	2293
PB3178. Using single cell CRISPR/dCas9-based regulatory element screening to dissect complex genetic loci.....	2294
PB3179. Using single nuclei multiomics and spatial transcriptomics to identify drivers of early pulmonary fibrosis	2295
PB3180. Utilization of Agena massARRAY to improve quality assessment of saliva samples for clinical genome sequencing	2296
PB3181. Utilization of multi-omic, multiplexed cell atlases and benchmark analysis demonstrates purity of product in the manufacture of induced pluripotent stem cell derived natural killer cells	2297
PB3182. Utilizing paraformaldehyde fixation to expand opportunities for single-cell RNA-sequencing.	2298
PB3183. Utilizing REVEAL SingleCell for single cell spatial transcriptomics data storage, analysis, and visualization	2299
PB3184. Visualize RNA biomarkers to spatially interrogate complex tissues using the RNAscope™ HiPlex v2 in situ hybridization assay.....	2300
PB3185. WES, with its limitations, continues to provide an attractive mode for genomic data collection	2301
PB3186. Whole genome sequencing of low input tagmentation-based libraries results in high quality somatic variant calling comparable to ligation-based PCR-free libraries.....	2302
PB3187. Why long-read sequencing is poised to become key to clinical genetics: an episodic ataxia study.....	2303
PB3188. A comprehensive analysis of the reusability of public omics data across 2.8 million research publications.....	2304
PB3189. A comprehensive database of Long-chain fatty acid oxidation disorder gene variants.....	2305
PB3190. A comprehensive Japanese genetic variation database TogoVar.....	2306
PB3191. A globally representative panel of human genomes: Resource and tutorials to support analyses of diverse populations	2307
PB3192*. A novel reference architecture for multi-party federation: enabling joint analysis of large-scale clinical-genomic data across distributed Trusted Research Environments.....	2308
PB3193. A Private, universal registry for patient engagement in biomedical research - Xia-Gibbs Syndrome as a model.....	2309
PB3194. A reference panel composed of Alzheimer's Disease enriched single-nucleotide and structure variants for structure variants discovery on SNP array data.....	2310

ASHG 2022 Annual Meeting Poster Abstracts

PB3195. A systematic assessment of the completeness of TCR databases across <i>Mus musculus</i> strains.	2311
PB3196. ALS/FTD Compute: an open, centralized repository of genomic data for ALS/FTD research	2312
PB3197. An improved platform for sharing genomic information of underrepresented populations: BIPMed 2.0.....	2313
PB3198*. Assessing the completeness of immunogenetics databases across diverse populations	2314
PB3200. Blockchain and Artificial Intelligence-Enabled Stratified Trial System (BESTS) - A patient driven data sharing platform that leverages genomic and health data to accelerate clinical trial recruitment for precision therapies	2315
PB3201. Breast cancer related genes: a genome-first approach of the Geisinger and UK Biobank cohorts.....	2316
PB3202. Brotman Baty Institute Clinical Variant Database (BBI-CVD).....	2317
PB3203. Burden of Mendelian disorders in a large middle eastern biobank.....	2318
PB3204. Clinical Variants for Brain Gene Curation: A Powerful and Under-Utilized Resource.	2319
PB3205. Colorado Biobank Portal is a cloud platform for the interactive exploration of genome-wide association studies results	2320
PB3206. CombVar: A pipeline to combine variant impact from prioritizers, integrators, and annotations	2321
PB3207. Concordance in variant calls from short-read sequencing data of admixed individuals using different reference genomes.....	2322
PB3208. Distance-based panel generation optimizes gene selection for targeted gene panel design.	2323
PB3209. Dynamically querying thousands of genomes to identify genetically matched cohorts with Intel® Optane™ Persistent Memory.	2324
PB3210. Early development of a locus specific database for <i>GUSB</i> , the gene associated with Mucopolysaccharidosis type VII: hints of a higher predicted prevalence.	2325
PB3211. Efficient querying of genomic reference databases with <i>gget</i>	2326
PB3212. Efficient reinterpretation of rare disease cases using Exomiser.....	2327
PB3213. Genomic database query tracking to understand the prevalence of rare diseases: <i>ASAH1</i> in the VarSome database.....	2328
PB3214. Global landscape of primary omics data generation and its secondary analysis across 193 countries and territories.....	2329
PB3215. Growing a research ecosystem in the cloud: Early insights from the All of Us Research Program.....	2330
PB3216. HLA-SPREAD: a natural language processing based resource for curating HLA association from PubMed abstracts	2331
PB3217*. How UK Biobank is democratising access to large-scale genomic and phenotypic data for discovery science	2332

ASHG 2022 Annual Meeting Poster Abstracts

PB3218. Identification and characterization of genomic variations from population scale whole genome sequencing in south India.....	2333
PB3219. Identifying Variants of Interest in Samples with Whole Genome Sequencing Data using Kids First Variant WorkBench.....	2334
PB3220. Indiana Biobank: a resource of linked electronic health records, omics data, genetic data, and biospecimens for the research community.....	2335
PB3221. JBrowse 2: a modular genome browser for visualizing synteny and structural variants	2336
PB3222. Lifebit delivers a federated Trusted Research Environment to allow researchers seamless access to 35,000 whole genomes in a novel COVID-19 cohort.....	2337
PB3223. MANE Select and beyond: expanding the joint NCBI and EMBL-EBI transcript set	2338
PB3224. Metadata retrieval from sequence databases with <i>ffq</i>	2339
PB3225. MetaMatchMaker: An artificial Intelligence tool for data discovery and meta-data harmonization	2340
PB3226. MGeND: Integrated database of clinical and genomic information to encourage precision medicine in Japan.....	2341
PB3227. National Center Biobank Network (NCBN) and genomic information for NCBN bioresources.	2342
PB3228. NCBI dbGaP FHIR API Provides Access to Thousands of Studies to Improve Data Integration and Interrogability for Biomedical Research.....	2343
PB3229*. NIA Genetics of Alzheimer’s Disease Data Storage Site (NIAGADS): 2022 Update.	2344
PB3230. North Carolina Macular Dystrophy (NCMD/MCDR1): Analysis of our entire database, a model disease of non-coding mutations.....	2345
PB3231. Participant-provided case-level data from GenomeConnect can impact variant classification	2346
PB3232. Partly Cloudy: analysis of blood lipids across All of Us and UK Biobank demonstrates the promise and limitations of cross-biobank analyses in the cloud.....	2347
PB3233. Path to independence: Overview of challenges and opportunities of computational data-driven research in biology.....	2348
PB3234. PhenX presents Bone and Joint updates and expansion of Social Determinants of Health .	2349
PB3236. PreSiBO: Database system for genome guided and data driven drug repurposing.....	2350
PB3237. Rare Disease Phenotyping: What is Next?.....	2351
PB3238. Regeneron Genetics Center (RGC) Cohort Identification Application.....	2352
PB3239. Role of Python APIs, Big data and Cloud computing in understanding the Genetic basis of a Disease	2353
PB3240. STR Truth Set for Evaluating Genome-wide STR Calls	2354
PB3241. Study design using the PhenX COVID-19 Collection and Variable Compare Tool	2355
PB3242. The Alzheimer’s Disease Sequencing Project - Follow Up Study ADSP-FUS: APOE genotype status and demographic characteristics across datasets.....	2356

ASHG 2022 Annual Meeting Poster Abstracts

PB3243. The BaseJumper™ research platform for single cell multiomic analysis and interactive visualization	2357
PB3244*. The cellxgene suite is an online analytical platform and the largest repository of standardized single-cell data.	2358
PB3245. The ClinGen Allele Registry, Linked Data Hub, and Evidence Repository Support Curation Efforts Through Registration of Variants, Aggregation of Supporting Evidence, and Publication of Pathogenicity Classifications	2359
PB3246. The importance of assessment of gene-disease association in the clinical practice.	2360
PB3247. The International Mouse Phenotyping Consortium: A unique resource underpinning the study of human disease.....	2361
PB3248. The NINDS Human Genetics Resource Center: 20 Years of Impact	2362
PB3249*. The One-Sided Matchmaking Platform (OSMP) - Accelerating genomic matchmaking and gene discovery through robust data sharing.....	2363
PB3250. The Osteogenesis Imperfecta Variant Database: current state.....	2364
PB3251. TT-Mars: structural variants assessment based on haplotype-resolved assemblies.....	2365
PB3252. Using RGD to analyze genes associated with obesity.	2366
PB3253*. Variant reclassification based on the allele frequency information from exome sequencing data of 20,455 patients enriched with under-represented populations.....	2367
PB3254. VariantMatcher a tool to enable connections amongst individuals with interest in a specific variant	2368
PB3255. VuTR: Visualising the consequences of high-impact 5' UTR variants	2369
PB3256. Web Tools and Resources for Generating and Accessing ACMG Criteria Specifications for Variant Classification by ClinGen Variant Curation Expert Panels	2370
PB3257*. Web-CSEA: Web-based Cell-type-Specific Enrichment Analysis of Genes.....	2371
PB3258. 8 Billion Polygenic Risk Score distributions: building unique reference populations based on genetic ancestry to increase the performance of PRS	2372
PB3259. A Cox proportional hazards whole genome sequencing study of ischemic stroke risk in a Brazilian sickle cell population.....	2373
PB3260. A fast Bayesian screen to identify pleiotropic loci and describe pleiotropic profiles.	2374
PB3261. A fast variational inference approach for powerful genome-wide association of quantitative traits.	2375
PB3262. A framework to improve the interpretation and prediction of variant effect sizes using non-linear functional models.....	2376
PB3263. A gene-based association test of interactions for maternal-fetal genotypes identifies genes associated with non-syndromic congenital heart defects.	2377
PB3264. A generalized framework to decorrelate family or population structure that can easily incorporate epistasis, GxE interaction or multiple-variant analysis in GWAS and NGS studies	2378
PB3265*. A genetic correlation disease-disease network (gcDDN) for the improved identification of novel phenotype relationships.....	2379

ASHG 2022 Annual Meeting Poster Abstracts

PB3266. A genome-wide search for parent-of-origin effects in 4,505 unrelated individuals identifies the <i>8q21.3</i> locus as a novel genetic regulator of Factor V plasma levels.	2380
PB3267. A locus near <i>ICOS</i> modifies risk for <i>Mycobacterium avium</i> complex airway infection in people with cystic fibrosis.	2381
PB3268. A machine-learning approach for disease prediction improves GWAS and polygenic scores	2382
PB3269. A multi-omics risk score approach to identify individual risk for incident type 2 diabetes	2383
PB3270*. A network based approach for fine-scale population structure inference in genetic datasets	2384
PB3271*. A novel approach identified eight gene x alcohol or gene x smoking interactions that contribute to serum lipids.....	2385
PB3272. A novel computational deconvolution method to estimate sample-level cell-type-specific gene expression profile.....	2386
PB3273. A Novel Gene-Based Test for Sequencing Studies Based on a Bayesian Variable Selection of Rare Variants	2387
PB3274. A novel haplotype based approach allows for conducting ancestry agnostic genome wide association studies.....	2388
PB3275. A novel method for annotation-agnostic differential transcript usage analysis	2389
PB3276. A novel method for multiple phenotype association studies based on genotype and phenotype network	2390
PB3277. A pipeline for predicting pathogenicity of monogenic cardiovascular disease genetic variants.	2391
PB3278. A polygenic framework nominating causal SNPs, their target genes, and cell-types of action	2392
PB3279. A polygenic score method including recessive model for improving odds ratio.	2393
PB3280. A precise, combinatorial approach for the estimation of <i>de novo</i> mutation rate from short-read WGS data.	2394
PB3281. A Rare Missense Variant of Large Effect is Associated with Cataract in Puerto Ricans	2395
PB3282. A robust whole-genome Mendelian randomization approach for improved estimation and inference of causal effects.....	2396
PB3283. A scalable framework for robust linear mixed model association testing.....	2397
PB3284. A scalable pipeline for robust relatedness inference with application to a study of 1,144,542 individuals.....	2398
PB3285. A scalable statistical framework for genome-wide interaction testing harnessing cross-trait correlations with an application to Alzheimer's Disease	2399
PB3286. A simple distance-based model predicts effector genes at GWAS loci	2400
PB3287. A statistical framework for integrative genetic association analysis.	2401
PB3288. A statistical framework for local genetic correlation across populations using GWAS summary data.	2402

ASHG 2022 Annual Meeting Poster Abstracts

PB3289. A transcriptomic signature of late-life depression.	2403
PB3290. A two-stage mixed effects model to analyze longitudinal zero-inflated microbiome count data	2404
PB3291*. A virtual transcriptome approach identifies drug repurposing candidates for Alzheimer’s disease.	2405
PB3292*. Accurate and cost-effective imputation of genotypes at whole genome level with Sparse Denoising Autoencoders.	2406
PB3293. Accurate and efficient estimation of local heritability using summary statistics and linkage disequilibrium	2407
PB3294. Accurate <i>in silico</i> confirmation of rare copy number variant calls from exome sequencing data using transfer learning.....	2408
PB3295. Accurate variant calling of the camouflaged LPA KIV-2 VNTR in whole exome sequencing data.....	2409
PB3296. Adjusting for genetic confounders leads to reliable detection of causal genes from transcriptome-wide association studies.....	2410
PB3297. Admixture mapping for phenome-wide discovery in a diverse health system biobank.....	2411
PB3298. Allele frequency differences of causal variants explain the bulk of low portability of cross- ancestry PRS	2412
PB3299. Allelic imbalance of chromatin accessibility in cancer identifies candidate causal risk variants and their mechanisms.....	2413
PB3300. Alpha-2A adrenergic receptor (<i>ADRA2A</i>) modulates susceptibility to Raynaud’s disease.	2414
PB3301. AlphaCluster: Coevolutionary driven residue-residue interaction models enable quantifiable clustering analysis of <i>de novo</i> variants to enhance predictions of pathogenicity.....	2415
PB3302. An accelerated, massively scalable variant calling and filtering workflow for the detection of de novo mutations in whole genome sequencing data.	2416
PB3303. An accurate and efficient causal gene network inference method that handles many confounding variables.	2417
PB3304. An allelic series rare variant association test for candidate gene discovery.....	2418
PB3305. An alternative approach to the estimation of biological age	2419
PB3306. An evaluation of ancestry specific polygenic risk scores (PRS) for uterine fibroids across populations.....	2420
PB3307. An Exomiser-based R pipeline for gene burden testing aids novel gene discovery in rare Mendelian diseases.	2421
PB3308. An experimentally-based functional gene embedding	2422
PB3309. Analysis of follow up data in large biobank cohorts: a review of methodology.....	2423
PB3310. Analysis of vitamin D receptor-binding variants in susceptibility to pediatric-onset multiple sclerosis.....	2424
PB3311. Ancestry-specific eQTL associations in HIS/NHW Alzheimer's disease cohorts	2426
PB3312. Ancestry-Specific Polygenic Risk Scores for Telomere Length in TOPMed.....	2427

ASHG 2022 Annual Meeting Poster Abstracts

PB3313. Application of the colorectal cancer polygenic risk score to the VA CSP #380 “Prospective Evaluation of Risk Factors for Large (≥ 1 cm) Colonic Adenomas in Asymptomatic Subjects” longitudinal colonoscopy screening cohort.....	2428
PB3314. Applying human phenomics to electronic health records provides a framework for understanding skin-aging related phenotypes	2429
PB3315. Assessing the safety of lipid-modifying medications among East Asian adolescents using genetics: Evidence from Hong Kong’s “Children of 1997” birth cohort.....	2430
PB3316. Assessing the value of non-coding variants in 500,000 whole-genome sequences in the discovery of gene-phenotype associations.....	2431
PB3317. Assessment of correlated multi-SNP Mendelian Randomization method using brain eQTLs to discover schizophrenia genes as a case study	2432
PB3318. Association Between Brain Structure and Alcohol Use Behaviors in Adults: A Mendelian Randomization and Multiomics Study.....	2433
PB3319*. Association of genetically predicted iron homeostasis biomarkers in type 2 diabetes and glycemic traits: A two-sample Mendelian Randomization study	2434
PB3320*. Association testing in admixed and multiethnic data sets nominate genetic modifiers of age at onset of Alzheimer’s disease.....	2435
PB3321*. Asymmetric non-portability across populations of different ancestry.	2436
PB3322. Auxiliary classifier generative adversarial network for phenotypic conditional genomic data synthesis.....	2437
PB3323. Being overweight in late life can be a causal factor contributing to longevity: a Mendelian Randomization study.	2438
PB3324. Bi-ancestral phenome-wide association of complement component 4 haplotypes in 550,000 individuals.....	2439
PB3325. Boosting SNP discoverability: How genetic architecture impacts the power of multi-trait GWAS.....	2440
PB3327. Broad- and narrow-sense heritability-contributing regions are shared across age, genetic sex and geographical cohorts.	2441
PB3328. BTS: Bayesian Tissue Significance scores for functional evaluation of GWAS results	2442
PB3329. Building ancestry-informed tools for more inclusive genomics and improved clinical interpretation in admixed populations.....	2443
PB3330. Burden testing of imputed rare variants to inform therapeutic hypotheses.....	2444
PB3331. Calcium signaling in Alzheimer’s disease: Insights from <i>PTK2B</i> and <i>miR146a</i> interaction analyses.	2445
PB3332. Calibrated and adaptive prediction interval improves polygenic score (PGS) performance in diverse populations	2446
PB3333. CARMA: Novel Bayesian model for fine-mapping with high-dimensional functional data	2447
PB3334. Causal inference for the relationship between transcription and DNA methylation in breast cancer.	2448

ASHG 2022 Annual Meeting Poster Abstracts

PB3335. Cell composition inference and identification of layer-specific transcriptional profiles with POLARIS.....	2449
PB3336. Characterizing copy number variations and their association with plasma proteins using short and long-read sequencing	2450
PB3337. Characterizing features affecting Local Ancestry Inference in Latin American populations	2451
PB3339. Colorectal cancer polygenic risk score is inversely associated with eye phenotypes	2452
PB3340. Combining family history with a polygenic risk score enables highly effective risk stratification for glaucoma	2453
PB3341. Common and Rare Variants Genetic Association Analysis of Late-Onset Alzheimer's Disease in 500,000 Exomes from UK Biobank.....	2454
PB3342. COMPADRE: a robust, cloud-optimized genetic relatedness platform.....	2455
PB3343. Comparing methods to adjust for fine-scale population structure in rare variant analyses..	2456
PB3344. Comparing the relationships of genetically proxied PCSK9 inhibition with mood disorders, cognition, and dementia between men and women: a drug-target Mendelian randomization study ..	2457
PB3345. Comparison of trans-ancestry meta- and mega-analyses of cis-eQTLs in whole blood and LCLs	2458
PB3346. Constructing and benchmarking omnibus polygenic risk scores using GWAS summary statistics.....	2459
PB3347. Controlled Discovery and Localization of Causal Variants via Bayesian Linear Programming	2460
PB3348. Copy-number variants as modulators of common disease susceptibility.....	2461
PB3349. Credible set determination for multi-ancestry fine-mapping.	2462
PB3350. Cross-trait meta-analysis reveals shared genetic architecture between PCOS and chronic inflammation markers	2463
PB3351. <i>De novo</i> mutation hotspots in homologous protein domains identify function-altering mutations in neurodevelopmental disorders.	2464
PB3352. Deep Learning-based Phenotype Imputation on Biobank-scale Data Increases Genetic Discoveries.....	2465
PB3353. Deep Representational Clustering for Genetic Endotype Discovery from Clinical Data	2466
PB3354. Deep sequencing of DNA from urine of kidney allograft recipients to estimate donor/recipient-specific DNA fractions.....	2467
PB3355. DeepPerVar: a multimodal deep learning framework for functional interpretation of genetic variants in personal genome.....	2468
PB3356*. Deep-PheWAS.R: an innovative R package for phenome wide association of novel phenotypes.	2469
PB3357*. Defining longitudinal disease trajectories in 146,000 individuals with hypertension from Penn Medicine Electronic Health Records	2470

ASHG 2022 Annual Meeting Poster Abstracts

PB3358*. Deprivation index as a predictor in PheWAS recapitulates associations between socio-economic status and phenotypes captured by ICD9 codes.	2471
PB3359. Detecting haplotype association using synthetic genetic variants.....	2472
PB3360*. Detecting shared and sex-specific causal genetic effects on complex traits	2473
PB3361. Detecting Somatic Mosaicism at Tandem Repeats	2474
PB3362. Detecting variable-number tandem repeat (VNTR) variation using Oxford Nanopore sequencing data at population scale.	2475
PB3363. <i>Detection of underdiagnosis of complex diseases due to underrecognition of high polygenic risk</i>	2476
PB3364. Developing an African American population-based transcriptome prediction model from the GENE-FORCAST cohort	2477
PB3365. Development and validation of an RNA-seq-based transcriptomic risk score for asthma...	2478
PB3366. Differential gene expression analysis based on linear mixed model corrects false positive inflation for studying quantitative trait	2479
PB3367. Differentially Private Release of Genomic Summary Statistics Using Factorization Mechanisms	2480
PB3368*. Differentiate association and causal effects of cardiovascular disease risk factors on Alzheimer's disease and its age-at-onset.	2481
PB3369. Discovery of clinically relevant gene-phenotype associations from 454,712 whole exomes in the UK Biobank	2482
PB3370. Discovery of rare variants associated with alcohol problems improved by leveraging machine learning phenotype prediction and empirical functional variant weighting.....	2483
PB3371. Disentangling the common genetic architecture and causality of rheumatoid arthritis and systemic lupus erythematosus with COVID-19 outcomes: genome-wide cross trait analysis and bi-directional Mendelian randomization study.....	2484
PB3372. Disentangling the genetic overlap and causal relationship between primary open-angle glaucoma, brain morphology, and major neurodegenerative disorders.	2485
PB3373*. Distinct explanations underlie gene-environment interactions in the UK Biobank	2486
PB3374. Diversity is key for POAG: the proportion of variance for primary open-angle glaucoma explained by covariates and genetic risk scores varies by ancestral group in the Million Veteran Program.....	2487
PB3375. DNAREcords:An extensible sparse format for petabyte scale genomics analysis.....	2489
PB3376. DosaCNV: A deep multiple instance learning framework for jointly predicting copy number variation pathogenicity and gene dosage sensitivity.....	2490
PB3377. DrFARM: Identification and inference for master regulator variants in multi-trait GWAS	2491
PB3378. Effect of type 2 diabetes and its genetic susceptibility on severity and mortality of COVID-19 in UK Biobank	2492
PB3379. Efficient multivariable Mendelian randomization for confounder adjustment using public GWAS databases	2493

ASHG 2022 Annual Meeting Poster Abstracts

PB3380. Efficient tests of marginal epistasis reveal the impact of interactions on complex traits.....	2494
PB3381*. Elucidate trans-ancestry cross trait genetic architecture for tobacco and alcohol use phenotypes using 3 million Individuals.	2495
PB3382. Estimating heritability explained by local ancestry and stratification bias in admixture mapping from summary statistics	2496
PB3383. Estimating indirect maternal genetic effects on children’s autism spectrum disorder risk..	2498
PB3384. EVai’s “Suggested Diagnosis” feature: a new AI-based method to increase diagnostic yield in Rare Disease Patients	2499
PB3385. Evaluating differential expression of imputed miRNA expression in the subgenual anterior cingulate cortex.....	2500
PB3386. Evaluating human genetic evidence for existing drug targets in the UK Biobank	2501
PB3387*. Evaluating strategies to infer continental ancestry using single-cell RNA-seq datasets: an application to the Human Cell Atlas	2502
PB3388. Evaluating the accuracy genotype imputation in selected African populations.....	2503
PB3389. Evaluating the impact of parallel mutations on the accuracy of statistical phasing.....	2504
PB3390. Evaluation of GENESIS, SAIGE and REGENIE for genome-wide association study of binary traits in correlated data.....	2505
PB3391. Evaluation of germline variants in pediatric low grade glioma using non-negative matrix factorization	2506
PB3392. Exome sequencing of 628,388 individuals identifies common and rare variant associations with clonal hematopoiesis phenotypes.....	2507
PB3393. Exploiting the mediating role of the metabolome to unravel transcript-to-phenotype associations	2508
PB3394. Exploring germline genetics of <i>in situ</i> and invasive cutaneous melanoma.....	2509
PB3395. Exploring the genetic, socio-economic and environmental determinants of drug adherence	2510
PB3396*. Expression and Splice QTLs found in COVID-19 patients and controls show differential colocalization and heterogeneity between infectious and non-infectious states.	2511
PB3397. Extending Genome-Wide Association Studies to admixed cohorts with high degrees of relatedness.....	2512
PB3398. Extremely sparse models of linkage disequilibrium in diverse GWAS.....	2513
PB3399. Fine-mapping one chromosome at a time: Software and an initial application to UK Biobank	2514
PB3401*. Functional Annotations-Informed Whole Genome Sequence Analysis Identifies Novel Rare Variants for AD in the Alzheimer’s Disease Sequencing Project.....	2515
PB3402. Genealogy-wide association improves detection of rare and low-frequency variants in under-sequenced populations.	2516
PB3403. Gene-by-gene and gene-by-sex interactions identified in the UK Biobank using Explainable Boosting Machines.....	2517

ASHG 2022 Annual Meeting Poster Abstracts

PB3404. Gene-sodium intake interactions reveal new blood pressure loci in a multi-ancestry genome wide association study in UK Biobank and CHARGE.....	2518
PB3405. Genetic association models are robust to common population kinship estimation biases...	2519
PB3406. Genetic determinants of a metabolic model for age.....	2520
PB3407. Genetic Determinants of Circulating Glycine and Cardiovascular Disease: Causal Relationship or Red Herring?	2521
PB3408. Genetic evidence of assortative mating in the Japanese population.	2522
PB3409. Genetic variants associated with serum IGF-1 levels in the Long Life Family Study.....	2523
PB3410. Genetically predicted counts of white blood cells are associated with lower risk of obesity: a Mendelian randomization study.....	2524
PB3411. Genetically-guided phenotype imputation of partial or missing traits in biobank data increases power for genetic discovery.....	2525
PB3412. Genome wide association study of asthma in 1587 French Canadian subjects.	2526
PB3413. Genomes, exomes, and imputation: Comparing technologies for variant association and discovery in large-scale genetic studies	2527
PB3414. Genome-wide analysis of dental caries variability identifies genotype-by-environment interactions.....	2528
PB3415. Genome-wide association studies of brain imaging endophenotypes derived from unsupervised learning identifies genes relevant to brain structure	2529
PB3416*. Genome-wide association studies of metabolic traits in Samoans.....	2530
PB3417. Genome-wide association study of obstructive sleep apnea in the Million Veteran Program uncovers heterogeneity by sex and genetic ancestry.....	2531
PB3418. Genome-wide association study of the human metabolome in diverse ancestries identifies an association of the <i>OPLAH</i> locus with 5-oxoproline (pyroglutamic acid) in individuals with African ancestry.	2532
PB3419. Genome-wide classification of epigenetic signal reveals regions of enriched heritability in complex immune traits.....	2533
PB3420. Genome-wide imputed differential expression enrichment (GIDEE) analysis identifies trait-relevant tissues	2534
PB3421*. Genome-wide study on 72,298 Korean individuals in Korean biobank data for 76 traits identifies hundreds of novel loci.....	2535
PB3423. GrafPop: A tool to quickly infer subject ancestry from multiple large genotype data sets..	2536
PB3424*. GWAS meta-analysis of DXA-derived bone mineral density identifies 42 novel loci and uncovers biological pathways not previously identified in larger studies of ultrasound-derived bone mineral density.....	2537
PB3425. GWAS of breast cancer family history reveals evidence for indirect genetic effects on maternal cancer risk.	2538
PB3426. GWAS of pericarditis derived from a natural language processing model on self-reported free text data identifies a genome-wide significant association on chromosome <i>2q14.1</i>	2539

ASHG 2022 Annual Meeting Poster Abstracts

PB3427. GWAS of serum ALT and AST using whole genome sequencing data for 119,009 UK Biobank participants yields novel associations with rare variants.....	2540
PB3428. Hand grip strength and its underlying genetics is associated with widespread protein changes revealed by population-based plasma proteomics.....	2541
PB3429. Haplotype-based fine-mapping of variant associations for complex traits	2542
PB3430. Hierarchical scanning strategy reveals signal regions in 260 genes implicated in regulation of gene expression on chromosome 19q	2543
PB3431. HLA Analysis Tool Kit v2.0	2544
PB3432. Hybrid autoencoder for robust ancestry inference in the presence of data artifacts and relatedness.....	2545
PB3433. HybridGWAIS-Web: A fast and secure epistasis detection web service.....	2546
PB3434*. Identification of age-related genetic and molecular determinants and their link to disease.	2547
PB3435. Identification of key biomarkers associated with obesity using multi-omics data integration	2548
PB3436. Identification of rare genetic coding variants associated with Alzheimer's disease.....	2549
PB3437. Identifying circulating proteins as biomarkers for age at menarche and age at natural menopause: insights from a Mendelian randomization study.....	2550
PB3438. Identifying common genetic susceptibility underlying comorbid phenotypes using Binomial Regression.....	2551
PB3439. Identifying composite biomarkers in multi-omics datasets of osteoporosis for drug discovery	2552
PB3440. Identifying functionally-related SNPs and genes from phenome-wide genetic effect correlations using GWAS summary statistics.....	2553
PB3441. Identifying non-common variant sources of clinical comorbidities through integration of electronic health data and genetics for schizophrenia.....	2554
PB3442. Identifying the conserved versus divergent structural sequences of the human pangenome.....	2555
PB3443. Identity-by-descent strategies uncover cryptic relatedness, deliver novel pedigrees, and facilitate gene discovery in amyotrophic lateral sclerosis.....	2556
PB3444. Imaging derived phenotypes of the knee reveal novel genetic and clinical risk factors associated with knee osteoarthritis severity.	2557
PB3445*. Impact of heterogeneity-by-ancestry on GWAS in admixed populations.....	2558
PB3446. Implementing Mendelian Randomization to Assess Foetal Risk from Intrauterine Prescriptive Drugs for the Treatment of Diabetes, Hypertension and Thyroidism in Pregnancy.....	2559
PB3447. Improved breast cancer risk stratification by integration of a cross-ancestry polygenic model with clinical risk factors.....	2560
PB3448. Improved genotyping of rare variants from AppliedBiosystems™ Axiom™ microarrays, using Support Vector Machine (SVM) prediction models.	2561
PB3449. Improved Prediction Performance of Polygenic Risk Score with Gene Expression Data ...	2562

ASHG 2022 Annual Meeting Poster Abstracts

PB3450. Improved risk prediction using functionally calibrated polygenic risk scores in admixed populations.....	2563
PB3451. Improvements in performance of polygenic risk scores for Alzheimer's disease in the Midwestern Amish using founder population-specific weights.....	2564
PB3452. Improving Cross-Population Polygenic Risk Scores with a Tree-Guided Deep Learning Method.....	2565
PB3453. Improving fine-mapping by modeling infinitesimal effects.....	2566
PB3454. Improving the accuracy and interpretation of polygenic risk score through modeling the pathways of disease and multiple risk factors.....	2567
PB3455. Impseer: Machine learning prediction of imputation quality.....	2568
PB3456. Imputation of heritable gene expression reveals transcriptomic associations with neuroimaging phenotypes.....	2569
PB3457. In silico GWAS: the rapid and accurate (pre)computation of genetic disease associations using only population-level data.....	2570
PB3458. Incorporating functional annotation with bilevel continuous shrinkage for polygenic risk prediction.....	2571
PB3459. Incorporating functional annotations of rare variants in interpretable neural networks identify novel Alzheimer's disease risk loci.....	2572
PB3460. Incorporating related individuals in genome-wide association studies to reduce biases from familial effects and population stratification in Mendelian randomization.....	2573
PB3461. Inference for set-based effects in genome-wide association studies with multiple interval-censored outcomes.....	2574
PB3462. Insights into the comorbidity between type 2 diabetes and osteoarthritis: a genetics view.....	2575
PB3463. Integrating external controls by regression calibration for genome-wide association study.....	2576
PB3464. Integrating rare and common variant association data to improve therapeutic target discovery.....	2577
PB3465. Integration of gene expression data in Bayesian association analysis of rare variants.....	2578
PB3466. Integration of International Mouse Phenotype Consortium (IMPC) and UK Biobank data to identify genes associated with blood molecular phenotypes.....	2579
PB3467. Integrative approaches to unravel sex- and age-associated gene signatures and networks.....	2580
PB3468. Interrogating causality of epidemiological risk factors on adolescent idiopathic scoliosis: a two-sample Mendelian randomization study.....	2581
PB3469. Inverted genomic regions between reference genome builds in humans impact imputation accuracy and decrease the power of association testing.....	2582
PB3470. Investigating miRNA role in the neuropathology of Major Depression in a large postmortem brain sample.....	2583
PB3471. Investigating the effect of body size between menarche and first birth on breast cancer risk in later life: A lifecourse Mendelian randomization study.....	2584

ASHG 2022 Annual Meeting Poster Abstracts

PB3472. Investigating the impact of the rare pathogenic variants linked to Intellectual Disability genes on the cognitive ability in adults: exome analysis of the UK Biobank cohort.....	2585
PB3473. Investigating the potential impact of PCSK9-inhibitors on mood disorders using eQTL-based Mendelian randomization	2586
PB3474. Investigating the relationship between immune disease genetic risk and Alzheimer's Disease incidence.	2587
PB3475. Investigation of Genetic Variants and Causal Biomarkers Associated with Brain Aging ...	2588
PB3476. Investigations of the potential causal effects of the immune response to varicella-zoster virus on multiple health conditions: a Mendelian randomization phenome-wide association study	2589
PB3477. Is the portability of polygenic scores a matter of genetic ancestry alone?	2590
PB3478. Isoform level transcriptome wide association study of prostate cancer risk reveals large number of transcript specific associations	2591
PB3479. Joint analysis of GWAS and multi-omics QTL summary statistics reveals a large fraction of GWAS signals shared with molecular phenotypes	2592
PB3480*. Latent factor analysis reveals genetic components across GWAS traits.....	2593
PB3481. Leveraging cohorts from diverse ancestries to uncover generalizable genotype driven fetal- and adult-specific brain transcriptomic mechanisms in psychiatric disorders.	2594
PB3483. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries	2595
PB3484. Leveraging Molecular-Diagnosis Workflow to Identify Modifier Genes in Diseases.....	2596
PB3485. Leveraging multiple reference panels by stacked regression TWAS identifies 6 novel independent risk genes for Parkinson's disease	2597
PB3486. Leveraging multiple traits to detect shared non-additive genetic variation in genome-wide association studies.....	2598
PB3487. Leveraging polygenic risk scores of biomarkers to identify obesity endotypes.	2599
PB3488. Leveraging the use of meta-analysis summary statistics to improve gene expression prediction models.	2600
PB3489. Leveraging tissue-wide functional regulatory information improves portability of trans-ancestral polygenic risk score applications	2601
PB3490*. Linking the joint genetic structure of neuroanatomical phenotypes with psychiatric disorders.	2602
PB3491. Local ancestry at the MHC region is associated with disease heterogeneity in a multi-ethnic lupus cohort.....	2603
PB3492. Loci for Cognitive Preservation in the Midwestern Amish	2604
PB3493. Longitudinal, multi-modal EHR prediction of incident CKD and ESRD for risk stratification and genetic studies	2605
PB3494*. Low-coverage sequencing imputation with 150,119 UK Biobank reference samples	2606
PB3495. Low-coverage sequencing induces uncertainty in polygenic scoring	2607

ASHG 2022 Annual Meeting Poster Abstracts

PB3496. LT-Free: A novel method for leveraging family history in genetic association studies of arbitrarily complex diseases.....	2608
PB3497. LTPI: A generalized liability threshold model combining dichotomous and continuous phenotypes in EHR increases the association power	2609
PB3498. Machine learning approach for identifying diabetes nephropathy and non-diabetic chronic kidney disease cases and controls using urinary high-throughput transcriptomic data.	2610
PB3499. Machine Learning Approaches to Genome-wide association studies.....	2611
PB3500. Machine learning based methods for predicting guide RNA effects on cell fitness and gene expression in CRISPR epigenomic experiments.....	2612
PB3501. Machine learning guided transcriptomic analyses reveal a novel 9-gene module associated with schizophrenia risk	2613
PB3502. Machine learning model discriminates false from true CNVs with a 98% accuracy.....	2614
PB3503. MagicalRsq: Machine learning based genotype imputation quality calibration.	2615
PB3504*. MANOCCA : An innovative powerful test to detect predictors of the gut microbiome variability	2616
PB3505. MaSk-LMM: a matrix sketching-based fast and scalable linear mixed model for association studies in large biobanks	2617
PB3506. MATS: A novel multi-ancestry transcriptome-wide association study to account for heterogeneity in the effects of cis-regulated gene expression on complex traits	2618
PB3507. mBAT-combo: a gene-based association test to decipher masking effects	2619
PB3508. Mendelian randomization and colocalization of amino acids in the UK Biobank NMR metabolomic profiles	2620
PB3509. Mendelian randomization implicates GDF15 blood levels as a causal factor in inflammatory disease.	2621
PB3510. Mendelian randomization indicates sex-specific causal effects of estradiol levels on kidney function.	2622
PB3511*. MENDEL-modified segregation analysis of 168 <i>TP53</i> -positive families estimates age-specific risks for cancer types beyond the established spectrum for Li-Fraumeni syndrome.....	2623
PB3512. Metabolite prediction models in UK Biobank: A metabolome-wide association study (MWAS)	2625
PB3513. Meta-imputation combining the East Asian and multi-ethnic reference panels enhanced the imputation performance of rare variants for East Asian population	2626
PB3514. Methods and software for empirical Bayes multivariate multipletesting and effect size estimation.....	2627
PB3515. MethylPipeR: Multi-omic Prediction of Complex Traits.....	2628
PB3516. Metric projection—AUCs, correlations, and heritability—for biobank scale data: using sparse prediction in the UK Biobank to plug diversity gaps in future biobanks.....	2629
PB3517. miRNA Metabolome - wide association study: A multi-omics integrative approach to Asthma	2630

ASHG 2022 Annual Meeting Poster Abstracts

PB3519. MTClass: Identification and annotation of multi-tissue cis-eQTLs using machine learning.	2631
PB3520. Multi trait GWAS for diverse ancestry : Mapping the knowledge gap.....	2632
PB3521. Multi-ancestry and sex-specific phenome-wide association meta-analysis of genetically-predicted height.....	2633
PB3522. Multi-ancestry polygenic risk scores for venous thromboembolism	2635
PB3523. MultiSuSiE: Multi-population fine-mapping under the sum of single effects model	2636
PB3524. Nearest Neighbor Simulated Annealing Cohort Selection for Improved PCA Genome Matching.	2637
PB3525. Network-based cross-phenotype risk scoring models for non-linear compositing multiple disease risks using biobank-scaled PheWAS data	2638
PB3526. New paradigm for uncovering the clinical consequences of genetic variation.	2639
PB3527. New perspectives in the genetics of persistent opioid use: reexamining candidate gene studies and presenting new results from the Michigan Genomics Initiative.....	2640
PB3528. Next generation phenotyping of DEGCAGS syndrome.	2641
PB3529. NGS Variant Detection as a Sequence-to-Sequence Modeling Problem.....	2642
PB3530. Non-additive outlier effects of gene expression on anthropometric traits	2643
PB3531. Novel Approach to Age Estimation of Rare Variants Enables Prioritization of Functional Variants	2644
PB3532. Omics profiling in overweight and obese children in the California Bay Area during childhood development and the effects of a multi-modal health intervention program tailored to a majority Latinx community	2645
PB3533. Omnibus proteome-wide association study (PWAS) method identified 28 novel risk genes for Alzheimer's disease.....	2646
PB3534*. On the polygenic trait model in population-based human genetics studies: What is random and what is fixed.	2647
PB3535. One-sample and two-sample mendelian randomization studies consistently showed that an increase in adiposity raises serum uric acid level in the East Asian populations.....	2648
PB3536. OTTERS: A powerful TWAS framework leveraging summary-level reference data	2649
PB3537. PathWAS: Combining transcriptomics with proteomics to predict pathway functionality and associations with complex traits	2650
PB3538. Pathway analysis identifies novel non-synonymous variants contributing to extreme vascular outcomes in Williams-Beuren syndrome.....	2651
PB3539. Pathway analysis using expression factors reveals biological processes underlying genetics of complex traits.....	2653
PB3540*. Performance and accuracy of imputation panels for genetic association studies in sub-Saharan African populations.....	2654
PB3541. Personality traits are consistently associated with blood mitochondrial DNA copy number estimated from genome sequences in two genetic cohort studies.....	2655

ASHG 2022 Annual Meeting Poster Abstracts

PB3542. Phenome-wide Mendelian randomization study of plasma triglycerides and 2,600 disease traits.	2656
PB3543. Phenotype integration improves power and preserves specificity in biobank-based genetic studies of MDD.....	2657
PB3544. Phenotypic and genetic relationships between physical activity, sleep, and brain health....	2658
PB3545*. Phenotypic subtyping via contrastive learning reveals heterogeneity in the genetic architecture of traits.	2659
PB3546. PheWAS-based clustering of Mendelian Randomization instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment.	2661
PB3547. PLEIOVAR - Assessing Pleiotropic Effects of SNPs on Traits in Genome-wide Association Studies.....	2662
PB3548*. Polygenic prediction across populations is influenced by ancestry, genetic architecture, and methodology	2663
PB3549. Polygenic risk scores for coronary heart disease in diverse populations using population-specific optimization.	2664
PB3550. Polygenic scores enable discovery of widespread genetic interactions associated with quantitative traits in the UK Biobank.	2665
PB3551. Polygenic scores for insulin resistance are associated with brain volumes in the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program.	2666
PB3552. Polygenic transcriptome risk scores can translate genetic results between species.	2668
PB3553. PopMLvis: A Tool for Analysis and Visualization of Population Structure in Genome-Wide Association Studies.....	2669
PB3554. Population diversity in global reference genome cohorts	2670
PB3555*. Power Window: a power-based sliding window method to identify the rare and novel variants underlying gene-based association signals.....	2671
PB3556. Practical screening for differences in predictive value of polygenic scores across the phenotypic range.....	2672
PB3557. Predicting cancer risk from germline next-generation sequencing data using a novel context-based variant aggregation approach.	2673
PB3558*. Predicting ExWAS results from GWAS data: A shorter path to causal genes.	2674
PB3559. Predicting how our genetic variants regulate splicing and impact human phenotypes.....	2675
PB3560. Prediction of atrial fibrillation and stroke using machine learning models.	2676
PB3561. Prediction of congenital heart disease subgroups associated with copy number variants using a multinomial machine learning-based classifier.....	2677
PB3562. Prioritization of causal genes from genome-wide association studies by Bayesian data integration across loci.	2678
PB3563. Prospective analysis of disease incidence and progression with genetic, clinical and lifestyle risk factors.....	2679
PB3564. Protecting privacy in polygenic risk score calculation by homomorphic encryption	2680

ASHG 2022 Annual Meeting Poster Abstracts

PB3565. Proteome-Wide Association Analysis in the Women’s Health Initiative Study.....	2681
PB3566. Publicly Available Privacy-preserving Benchmarks for Polygenic Prediction.....	2682
PB3567. PUMICE-S: Prediction Using Models Informed by Chromatin conformations, Epigenomics and Summary Statistics.....	2683
PB3568. Quantifying portable genetic effects and improving cross-ancestry genetic prediction with GWAS summary statistics	2684
PB3569. Quantifying the causal impact of modifiable risk factors and biomarkers on total healthcare burden via Mendelian randomization	2685
PB3570*. Quantifying the effects of high-dimensional cross-trait assortative mating on complex trait genetic architectures.....	2686
PB3571. Rapid identification of clusters of multi-level colocalized traits within hundreds of candidate traits and in the presence of multiple causal variants.....	2687
PB3572. Rare variation analyses of over 6,900 ALS cases in a multi-ethnic population.....	2688
PB3573. Rare-variant aggregation analysis reveals new genes involved with IgG glycosylation.	2689
PB3574. Rare-variant association studies: When are aggregation tests more powerful than single-variant tests?.....	2690
PB3575*. Reaching 95.0 years of age: The genetic association in the elderly Croatian population..	2691
PB3576*. REGENIE v3: more efficient analysis of rare genetic variation with an extended set of gene-based tests.	2692
PB3577. Reliability of genotyping arrays for detection of rare variants: design & performance of a bespoke array focusing on obesity & diabetes.....	2693
PB3578. sCLC: a Novel Statistical Method for Association Studies of Multiple Phenotypes and Genetic Variants based on GWAS Summary Statistics.....	2694
PB3579. SDPRX: A statistical method for cross-population prediction of complex traits	2695
PB3580. Sensitive detection of within-species contamination from low-pass whole genome sequencing data.....	2696
PB3582. Sex-stratified vs. sex-combined analysis in the presence of genetic effect heterogeneity...	2697
PB3583. Shared genetic etiology and causality between COVID-19 and venous thromboembolism: evidence from genome-wide cross trait analysis and bi-directional Mendelian randomization study	2698
PB3584. Shared genetic etiology between Alzheimer’s disease and stroke: A large scale genome-wide cross-trait analysis.....	2699
PB3585. Shared Genetics Drive Mate Selection.....	2700
PB3586. Short tandem repeat expansions are present in up to 25% of sporadic amyotrophic lateral sclerosis and frontotemporal dementia patients	2701
PB3587. Simple phenotype transformations help elucidate genetic architecture in the tails of biomarker distributions.....	2702
PB3588. Single cell RNA-seq data analysis of the Huntington’s disease cerebellum using the interactive webserver, ICARUS.....	2703
PB3589. Single-cell transcriptomic landscape in Alzheimer’s disease	2704

ASHG 2022 Annual Meeting Poster Abstracts

PB3590. Smoking-informed methylation QTLs in human nucleus accumbens	2705
PB3591. Specific HLA risk alleles associated with the highly variable region of T cell receptors (TCRs) in Leukemia, Melanoma and COVID-19 patients.	2706
PB3592. ssCTPR: summary statistic based cross-trait penalized regression.....	2707
PB3593*. Subcontinental Admixture in Individuals with European Ancestry and Implications for Genetic Epidemiology Studies.....	2708
PB3594*. Supervised modeling on CRISPR gene regulatory screening and cell-type-specific gene regulatory landscapes to prioritize GWAS Targeted Genes	2709
PB3595. Supervised multiset sparse partial least squares discriminant analysis in multi-omics data integration.	2710
PB3596. Systematic Integration of Multi-omics Data for the Study of Coronary Artery Disease and Subclinical Atherosclerosis.....	2711
PB3597. The effect of CA and D vitamins on the course of COVID 19 from the plasma osmolality perspective in a Turkish Cohort.	2712
PB3598. The effect of mental diseases on chronic kidney disease: a bidirectional Mendelian randomization study.....	2713
PB3599. The effects of demographic-based selection bias on GWAS results in the UK Biobank	2714
PB3600. The hidden factor: the importance of accounting for covariate effects in power and sample size computation when analyzing a binary trait.	2715
PB3601*. The impact of 22q11.2 copy number variants on human traits in the general population.	2716
PB3602. The impact of genetic variation on hearing acuity variability.....	2717
PB3603. The impact of rare regulatory variation from epigenetics to protein	2718
PB3604*. The largest deep-coverage whole genome seq meta-analysis on osteoporosis identified novel GWAS loci.....	2719
PB3605. The nature of nurture is not in the parents' genes: estimating indirect genetic effects from parents using imputed grandparental genotypes	2721
PB3606. The relationship between kinship and heritability estimation accuracy.....	2722
PB3607. The role of sleep in human brain and heart health: an investigation using 40,000 brain and cardiac magnetic resonance images from the UK Biobank	2723
PB3608*. The unusual suspects: genome-wide significant sex-difference in minor allele frequency in gnomAD v3.1.2 whole genome sequence via a novel population-aware retrospective regression	2724
PB3609. The use of population-based common controls in GWAS of infectious diseases can result in biased association signals	2725
PB3610*. Tissue-specific impacts of aging and genetics on gene expression patterns in humans. ...	2726
PB3611. Tools for efficient manipulation and meta-analysis of summary statistics from multi-exposure, genome-wide, gene-environment interaction studies.....	2727
PB3613*. Tradeoff between prediction accuracy and transferability in the design of polygenic risk scores.	2728

ASHG 2022 Annual Meeting Poster Abstracts

PB3614. Transcript-aware gene burden analyses in the UK Biobank reveal isoform-specific compositions and functions.....	2729
PB3615. Transcriptome and GWAS identified disease-specific eosinophilia-pathways and their upstream regulators in primary biliary cholangitis.	2730
PB3616*. Transcriptomics-based drug-repurposing screen discovers potential novel therapies for primary open-angle glaucoma.....	2731
PB3617. Transferability of polygenic scores for anthropometric traits to the Native Hawaiian population	2732
PB3618. Transformer generative adversarial networks and variational autoencoder for causal analysis of genetic variation in the presence of unobserved confounding.....	2733
PB3619. TWAS of the X Chromosome in Neurodegeneration	2734
PB3620. Uncovering ancestry-specific differences in genetically driven transcriptomic dysregulation in schizophrenia and bipolar disorder.	2735
PB3621. Unpicking the Gordian knot: Mendelian randomisation identifies priority groups for prophylactic EBV vaccination.	2736
PB3622. Unravelling the genetic contributions to gestational duration in Autism Spectrum Disorder individuals: results from the SPARK sample.	2737
PB3623. Unsupervised representation learning significantly improves genomic discovery for lung function and respiratory disease prediction.	2738
PB3624. Using external reference panel and meta-analysis summary statistics for rare-variant aggregation tests.....	2739
PB3625. Using genetic associations from shared clustered genetic predictors to infer causation.....	2740
PB3627. Validation of large-scale Hispanic S-PrediXcan findings by independent Hispanic RNA-Seq data analysis confirms novel lipid-associated genes.....	2741
PB3628. Variants in <i>CRY2</i> and <i>MTNR1B</i> influence the diurnal variability of blood glucose levels..	2742
PB3629. Whole-exome sequencing in 545,835 individuals implicates 27 genes in eosinophil biology	2743
PB3630*. Winner's curse in rare variants association analysis: bias depends on effect direction and the pooled methods used.....	2744
PB3631. Machine learning-based phenotyping significantly improves power for genetic discovery in NAFLD and NASH.....	2745
PB3633. A unifying statistical framework to discover disease genes from GWAS	2746
PB3634. Local ancestry stratified GWAS identifies pharmacogenomic variants associated with metformin glycemic response in African American individuals with type-2 diabetes	2747

Cancer Posters - Wednesday

PB1001. 3D genomics with Arima Hi-C sequencing enables detection of clinically relevant gene fusions in pediatric cancer samples.

Authors:

K. Sikkink¹, M. Farooqi², A. Ahmed³, D. Reid¹, A. Schmitt¹; ¹Arima Genomics, San Diego, CA, ²Children's Mercy Hosp. Kansas City, Kansas City, MO, ³Seattle Childrens Hosp., Seattle, WA

Abstract Body:

Gene fusions are known to have oncogenic effects in various cancers and serve as disease biomarkers. Obtaining gene-level resolution of these large structural variants is not possible with karyotyping and can be challenging using fluorescence in situ hybridization (FISH) or RNA sequencing due to low transcript abundance or low-quality RNA. Here, we used a novel method, Arima-HiC sequencing, on 7 archived pediatric cancer samples to determine the technique's effectiveness in detecting gene fusions. We first adapted the Arima Hi-C method for use with formalin-fixed paraffin-embedded (FFPE) tissue. We then selected five archived pediatric alveolar rhabdomyosarcoma (ARMS) tumors (FFPE archival period range: 8-12 years)—known to be fusion-positive via prior clinical testing. Briefly, FFPE tissue scrolls were dewaxed, and the tissue rehydrated, underwent chromatin fragmentation, end-labeling, and proximity ligation prior to DNA purification per the Arima-HiC FFPE protocol. Purified DNA was next prepared as a short-read sequencing library and sequenced on a HiSeq. The raw reads were aligned and deduplicated, and structural variants were called using the HiC-Breakfinder software. We additionally selected 2 non-FFPE pediatric leukemia cases with cryopreserved blasts (archival period range: 1-3 years), for Arima-HiC sequencing (as above). These cases had undergone standard of care cytogenetic (karyotyping, FISH) and molecular (targeted cancer NGS sequencing panel) testing clinically and a genetic driver / known gene fusion had not been identified. Using Arima-HiC sequencing in the first ARMS cohort, we identified either a PAX3-FOXO1 (n=3) or PAX7-FOXO1 (n=2) gene fusion in each of our 5 cases, consistent with the original diagnostic cytogenetic finding. HiC data was additionally able to provide previously unknown partner gene information (PAX3 or PAX7) in 4 of the 5 cases where the partner was previously not known. We then analyzed the cryopreserved blasts from the two leukemia cases without previously detectable genetic drivers or gene fusions, a precursor B-cell acute lymphoblastic leukemia (B-ALL) in a 3-year-old male, and acute myeloid leukemia (AML) in a 2-year-old female. In the first case, we detected an EP300-ZNF384 gene fusion, which is a known, but rare, gene fusion seen in B-ALL. In the second case, we detected a rare KMT2A-MLLT10 gene fusion. These findings are clinically important as both fusions are prognostically relevant. In summary, this study demonstrates how Arima-HiC sequencing provides molecular diagnostic value in archived pediatric solid and liquid tumor specimens via the identification of clinically relevant gene fusions.

Cancer Posters - Thursday

PB1002*. 4C guided cis- and trans-interaction networks associated with the *RCCD1* gene at the 15p26.1 breast and ovarian cancer risk locus

Authors:

S. Chen¹, F. Segato Dezem², S. Dhungana¹, K. Ayaluri¹, B. Davis¹, S. Kar³, S. Gayther², **J. Plummer²**; ¹Cedars-Sinai Med. Ctr., Los Angeles, CA, ²Cedars Sinai Med. Ctr., Los Angeles, CA, ³Univ. of Cambridge, Cambridge, United Kingdom

Abstract Body:

Breast and ovarian cancers share common etiologies and similar genetics, such as mutations in *BRCA1/2* which are responsible for most multi-case breast and ovarian cancer in families. Previous meta-analysis of breast and ovarian combined GWAS revealed several shared risk regions including a region on 15q26. The *P53* interacting gene *RCCD1* in this region is a candidate susceptibility gene for both cancers. A colocalization analysis of breast and ovarian cancer case-control genetic association studies in over 145,000 and 146,000 controls fine mapped the shared association in this region to 17 pleiotropic credible causal risk variants ($P_{\text{breast}} < 1.16 \times 10^{-14}$ and $P_{\text{ovary}} < 7.50 \times 10^{-7}$). These variants had a strong *eQTL* associated with the *increased* expression of *RCCD1* in normal breast and ovarian tissues ($P_{\text{breast}} < 1.7 \times 10^{-21}$ and $P_{\text{ovary}} < 8.3 \times 10^{-6}$). To confirm *RCCD1* as the target gene, we utilized circular chromosome conformation capture (4C) to identify the physical interactions between this susceptibility gene and risk SNP. 4C analysis of *RCCD1* in breast and ovarian cancer cells identified similar patterns of cis-interaction and significant binding site enrichment for the *BRCA2* interacting gene *EMSY* ($P_{\text{adjusted}} = 9.24 \times 10^{-6}$). The 4C analysis pinpointed a single 2kB *RCCD1* that identified 9 shared cis- and 3 shared trans-interacting regions in four breast and ovarian cancer cell lines. *RCCD1* trans-interacting regions mapped to previously identified genome wide significant ($P < 5 \times 10^{-8}$) breast cancer risk loci (1p34.2 and 3p14.1) and to the pleiotropic breast-ovarian cancer risk locus at chromosome 9q34.2. Stable overexpression of *RCCD1* in breast and ovarian cancer precursor cells identified 13 and 11 differentially expressed genes (DEGs) respectively associated with breast and ovarian cancer risk at genome-wide significance ($\text{PMAGMA} < 2.6 \times 10^{-6}$ after Bonferroni correction). Eighty-two DEGs shared between breast and ovarian cancer were strongly enriched in *TP53* ($P = 9.9 \times 10^{-4}$), *Hippo* ($P = 2.51 \times 10^{-3}$) and *TNF* signaling ($P = 4.7 \times 10^{-3}$) pathways. This study describes a functional framework to identify and understand the biology of cis- and trans- interaction networks associated with common variant risk alleles identified by GWAS.

Cancer Posters - Wednesday

PB1003. A B-ALL pediatric patient with IGH rearrangement

Authors:

A. Zhao¹, W. Su¹, C. Stewart¹, K. Eastwood², M. T. Guardiola², J. Nahoul¹, C. A. Tirado³; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²Baylor Scott & White, Temple, TX, ³Baylor Scott & White Hlth., Temple, TX

Abstract Body:

B-cell acute lymphoblastic leukemia (B-ALL) can afflict both adult and pediatric patients B-cell acute lymphoblastic leukemia (B-ALL) can afflict both adult and pediatric patients, and this disease is characterized by a build-up of B lymphoblasts. Here we present a case of a 25-year-old male patient with a history of B-ALL. Ninety percent of the bone marrow revealed pancytopenia with sheets of B lymphoblasts consistent with the diagnostic of B-ALL for acute pre-B lymphoblastic leukemia. The immunophenotype also presented predominant immature precursor B lymphoid cells positive for CD19, CD10, CD34, CD58, CD38, CD9, and TdT. Chromosome analysis of the bone marrow showed a complex karyotype described as 45~47,XY,i(8)(q10),der(10)add(10)(p11.1)add(10)(q23),-20,+1~2mar[cp3]/46,XY[36]. While *IGH* rearrangements were cryptic cytogenetically, DNA FISH analysis showed evidence of the *IGH* (14q32.2) gene rearrangement in 96.5% of the nuclei examined. These results were described as nuc ish(IGHx2)(5'IGH sep 3'IGHx1)[187/200],(5'IGH,3'IGH)x1~4(5'IGH con 3'IGHx0~2)[6/200]. The remaining probes were normal: nuc ish(PBX1, D4Z1,PDGFRB,CDKN2A,D9Z3,ABL1,D10Z1,KMT2A,ETV6,TCF3, RUNX1,BCR)x2[200]. Metaphase FISH showed the *IGH* signal on the derivative chromosome 8. In light of these results the karyotype was characterized as 45~47,XY,i(8)(q10),der(10)add(10)(p11.1)add(10)(q23),-20,+1~2mar[cp3].ish i(8)(q10) IgH+. *IgH* abnormalities are rare in B-ALL and are associated with a poor prognosis. Further studies using the MYC/IGH DC,DF probe from Abbott showed a gain of *IGH* signal in 7.5% of the nuclei examined: nuc ish(MYCx2,IGHx3)[15/200]. This is most likely the same abnormality that was seen in the patient's previous FISH study where we used instead an IGH break apart probe. But there was a significant decrease in the number of abnormal nuclei representing residual disease. However, up to 2022, the patient presented no evidence of persistent or residual disease characterized by the immunophenotype of a small population of immature B lymphoid cells, no monotypic B lymphoid or aberrant T lymphoid populations, and no phenotypic evidence of significant myeloid immaturity.

Cancer Posters - Thursday

PB1004. A Case of A Hyperdiploid Complex Karyotype in a Patient with a Myelodysplastic Syndrome

Authors:

H. Robinson¹, A. Lozada², N. Elhajjaoui¹, J. Sung², K. Eastwood³, M. Guardiola³, C. A. Tirado³; ¹Univ. of California Los Angeles, Los Angeles, CA, ²Univ. of California, Los Angeles, Los Angeles, CA, ³Baylor Scott & White Hlth., Dept. of Pathology, Temple TX, USA., Temple, TX

Abstract Body:

We present a case study of a 73-year old female with a history of pancytopenia. Chromosomal analysis of the bone marrow revealed an abnormal karyotype including gain of chromosome 1, 4, 6, 8, 9, 19, and 20 in addition to loss of chromosome 11, 13, 15, 16, 17, and 22. Also, additional material of unknown origin found on 3q, 5p, 9p, 11p, 13p, 14p, and 15p, and two copies of 19p, a deletion of 8q, and rings and markers were present. This was characterized as:

75~77,XXX,+1,der(1;6)(p10;p10),add(3)(q27),+4,add(5)(p15.1),+6,+8,del(8)(q24.1),+add(9)(p24),-11,add(11)(p13),-13,add(13)(p10),add(14)(p11.2),-15,add(15)(p11.2),-16,-17,+19,add(19)(p13.3)x2,+20,-22,+0~4r,+4~10mar[cp11]/46,XX[8]. The cytogenetic analysis correlates with the concurrent FISH study which was positive for additional signals of *EVII* (3q26.2), *TAS2R1* (5p15.31), *EGR1* (5q31.2), *RELN* (7q22), *TES* (7q31), *RUNX1T1* (8q21.3), *ABL1* (9q34), *KMT2A* (11q23), *PML* (15q24.1), *CBFB* (16q22), *RARA* (17q21), *PTPRT* (20q12), *MYBL2* (20q13.12), *RUNX1* (21q22.12) and *BCR* (22q11.2). These results were described as nuc ish(EVI1x3~4)[74/100],(TAS2R1x3~6,EGR1x3~5)[91/100],(RELNx3~7,TESx3~4)[88/100],(RUNX1T1x2~4,RUNX1x3~4)[90/100],(ABL1x2~4,BCRx2~4)[84/100],(KMT2Ax3~6)[72/100],(PMLx2~4,RARAx2~5)[83/100],(CBFBx3~4)[66/100],(PTPRTx3~6,MYBL2x3~6)[84/100]. The present findings are suggestive of a diagnosis of myelodysplastic syndrome, unspecified (MDS-U). The presence of pancytopenia within the context of complex chromosomal structural and numerical abnormalities are usually associated with a poor prognosis in MDS.

Cancer Posters - Wednesday

PB1005. A cross-ancestry meta-analysis identifies 451 susceptibility loci for prostate cancer and yields a genetic risk score effective in multiple populations.

Authors:

A. Wang¹, R. Madduri², J. Shen³, A. Rodriguez², E. Saunders⁴, B. Darst⁵, X. Sheng¹, M. Gaziano⁶, G. Giles⁷, H. Nakagawa⁸, F. Wiklund⁹, S. Chanock¹⁰, T. Dadaev¹¹, S. Berndt¹², T. Edwards¹³, J. Witte¹⁴, A. Justice¹⁵, R. Eeles¹⁶, Z. Kote-Jarai¹⁷, D. Conti¹⁸, C. Haiman¹, Million Veterans Program, PRACTICAL Consortium; ¹Univ. of Southern California, Los Angeles, CA, ²Argonne Natl. Lab., Lemont, IL, ³Univ. of Southern California, Los Angeles, CA, ⁴Inst. of Cancer Res., Sutton, Surrey, United Kingdom, ⁵Fred Hutchinson Cancer Res. Ctr., Seattle, WA, ⁶VA Boston Hlth.care System, Jamaica Plain, MA, ⁷Cancer Council Victoria, Melbourne, Australia, ⁸RIKEN IMS, Yokohama, Japan, ⁹Karolinska Inst.t, Stockholm, Sweden, ¹⁰Natl. Cancer Inst, Rockville, MD, ¹¹Inst. of Cancer Res., London, United Kingdom, ¹²Natl. Cancer Inst., Rockville, MD, ¹³Vanderbilt Univ. Med. Ctr., Nashville, TN, ¹⁴Stanford Univ., Burlingame, CA, ¹⁵Yale Sch. of Med., New Haven, CT, ¹⁶The Inst. of Cancer Res., Sutton, United Kingdom, ¹⁷The Inst. of Cancer Res., London, United Kingdom, ¹⁸Univ of Southern California, Los Angeles, CA

Abstract Body:

Previous genome-wide association studies (GWAS) identified 269 susceptibility loci for prostate cancer (PCa). To further our understanding of the genetic architecture of PCa, we combined the previous largest-available multi-ancestry PCa GWAS with 10 additional independent studies (UK Biobank, FinnGen Study, BioMe, BioVu, eMerge, PLCO, MVP, MDAnderson, CA UG, and NCI-MD) resulting in a total of 156,319 PCa cases and 788,443 controls from men of European (122,188/604,640), African (19,391/61,608), Asian (10,839/95,790), and Hispanic (3,931/26,405) ancestry. In total, 42,428,922 variants were examined for association with PCa risk using logistic regression adjusting for age, sub-study, and principal components. Per-allele ORs and SEs from individual studies were combined by a fixed-effects inverse-variance weighted meta-analysis. To identify and refine independent index risk loci for PCa, we implemented a forward-selection approach using multi-population Joint Analysis of Marginal summary statistics (mJAM) to identify index SNPs and any additional independent signals within each region via conditional analysis. In total, we identified 451 susceptibility loci for prostate cancer, 187 of which were novel. We constructed a genetic risk score (GRS) by summing risk allelic dosages weighted by the conditional per-allele log-ORs based on all 451 variants, and compared the performance to the GRSs based on past marker sets (n=100, 181, 269). The percentage of cases has increased for each population within higher GRS categories (e.g. from 40.5% in GRS₁₀₀ to 51.2% in GRS₄₅₁) and decreased within lower GRS categories (e.g. from 7.5% in GRS₁₀₀ to 4.4% in GRS₄₅₁). Correspondingly, the OR per GRS SD has also increased over time and for the current GRS₄₅₁ the estimated OR per SD ranged from 2.12 (95% CI: 2.08-2.17) for men of African ancestry to 2.47 (95% CI:2.44-2.50) for men of European ancestry. Moreover, in African ancestry, per SD of GRS₄₅₁ was significantly associated with higher risk of aggressive disease than of non-aggressive disease (OR 2.20 vs. 2.05, p_{het}=2.4x10⁻³), while no significant difference in disease aggressiveness was observed for other groups. Our finding of novel risk variants improves the ability of GRS in risk estimation and stratification across ancestral groups.

Cancer Posters - Thursday

PB1006. A genome-wide association study to identify predictive markers for the risk of nivolumab-induced immune-related adverse events

Authors:

H. Zembutsu¹, C. Udagawa², S. Tsuchida³; ¹Natl. Cancer Ctr., Res. Inst., Tokyo, Japan, ²Natl. cancer center, Tokyo, Japan, ³Tsuchida Hosp., Sapporo, Japan

Abstract Body:

In this study, we sought to identify the potential genetic variants that could predict the risk of nivolumab-induced immune-related adverse events (irAEs) in patients with cancer. We enrolled 622 Japanese patients who had been treated with nivolumab, and carried out a genome-wide association study (GWAS) on 315 cases (with irAEs) and 86 control subjects (without irAEs). The associations for 507 single nucleotide polymorphisms (SNPs) showing $P < 0.001$ were further investigated using an independent cohort of 205 cases and 16 control subjects as a replication study. In the combined analysis of the GWAS and replication studies, possible associations were found for a total of 90 SNPs. Although we did not identify any SNPs to be significantly associated with nivolumab-induced irAEs, the SNP most strongly associated with nivolumab-induced irAEs was rs469490 on chromosome 21q21.3 ($P_{\text{combined}} = 2.97 \times 10^{-7}$, odds ratio (OR) = 5.15). In a subgroup analysis for the identification of genetic markers for nivolumab-induced hypothyroidism, rs8023690 on chromosome 15q26.2 was possibly associated with nivolumab-induced hypothyroidism ($P = 2.58 \times 10^{-7}$, OR = 4.09). We consider this study to be an important hypothesis-generating study to guide future studies in larger and/or other ethnic cohorts. These findings may also help to contribute to the prediction of irAE risk for better prognosis and quality of life among patients with cancer.

Cancer Posters - Wednesday

PB1007. A genome-wide gene expression signature of Breast cancer in the Moroccan women

Authors:

S. Ayoubi¹, P. Ostano², M. Mello-Grand², I. Gregnanin², M. Elkarroumi³, S. Nadifi¹, G. Chiorino²; ¹Genetics molecular pathology Lab., Med. Sch., Univ. Hassan II of Casablanca, Casablanca, Morocco, ²Genomics Lab., Fondazione Edo ed Elvo Tempia, Biella, Italy, ³Mohammed VI Ctr. foCancer Treatment, Univ. Hosp. Ctr. Ibn Rochd, Casablanca, Morocco

Abstract Body:

Background Breast cancer (BC) is the commonest form of cancer in North African countries. Thus, the early diagnosis is crucial for deciding the course of treatment and saving lives. Long non-coding RNAs (lncRNAs) are non-coding RNA class, emerged as key regulators of several biological processes. It has been reported that aberrant expression of lncRNAs was involved in BC. No studies have specifically focused on lncRNAs in BC in developing countries like Morocco. The prime aim of the present study was to establish a unique “lncRNAs-mRNAs” expression signature specific of BC in the Moroccan women. **Material and Methods** We analyzed lncRNAs-mRNAs expression profiles in breast tumors and normal breast tissues by Agilent whole genome microarrays. Then, we screened out the differentially expressed (DE) genes. We performed functional analysis of DE-lncRNAs-mRNAs using public databases. RT-QPCR was employed to validate the previous results. **Results and discussion** 190 dysregulated lncRNAs and 2500 aberrantly expressed mRNAs were identified. 90 lncRNAs and 1500 mRNAs were down-regulated in breast tumors compared to normal tissues. While 100 lncRNAs and 1000 mRNAs were up-regulated genes. THNSL1-2 lncRNA was strongly up-regulated in all molecular subtypes of BC; while “MIR497HG lncRNA” was commonly under-expressed. Most of the genes were related to cancer-associated behaviors, such as the p53 signaling pathway and cell cycle. The RT-QPCR results confirmed the microarray’s data. **Conclusion** Our study is the first to analyze lncRNAs-mRNAs expression profiles of BC in Moroccan women; it provides useful information for exploring novel biomarkers in order to improve BC diagnosis and therapy in Morocco.

Cancer Posters - Thursday

PB1008. A novel immunohistochemical marker helps in the diagnosis of Lynch syndrome: detection of deficient DNA mismatch repair pathway proteins crypts within non-tumor colonic mucosa.

Authors:

F. Coulet¹, S. Breton², N. Basset¹, M. Svrcek²; ¹APHP.Sorbonne Univ.-Genetics Dept., Paris, France, ²APHP.Sorbonne Univ.-Anatomo-pathology Dept., Paris, France

Abstract Body:

Lynch syndrome (LS) is the most common form of hereditary colorectal cancer (CRC), and is characterized by a germline monoallelic pathogenic variant in one of the genes of DNA mismatch repair pathway (MMR system). Tumor tissues of these patients show a microsatellite instability (MSI) phenotype and/or loss of MMR protein expression (dMMR). Few studies suggest the existence of morphologically normal colonic crypts losing the expression of one of the MMR proteins in patients with SL. These crypts, called "MMR-deficient colonic crypts", would be pathognomonic of LS. However, few data in the literature explain their precise frequency and the optimal way to detect them. The objectives of this work were (i) to specify the characteristics of dMMR crypts in patients with LS and to propose an optimal detection protocol; (ii) to investigate whether dMMR crypts detection could be a useful tool in case of patients with Lynch-like syndrome (LLS); or (ii) to add argument allowing classification of uncertain significance genetic variants (VUS). All patients were selected retrospectively, according to available tissues, among dMMR/MSI CRC patients from Sorbonne University hospitals. The study included 15 patients with SL (5 MLH1; 7 MSH2; 3 MSH6); 7 patients with LLS and 7 patients with VUS in MMR genes. For each patient, 10 IHC slides were stained with the antibody against the tumor-lost MMR protein on 1 or 2 blocks of normal colonic mucosa adjacent to and distant from the CRC. For each block, the number of dMMR crypts was counted and a crypt ratio (number of dMMR crypts/total number of crypts analyzed) was determined; dMMR crypts were identified in 80% of patients with SL, and were identified in 1 patient with LLS, becoming a strong candidate for further molecular exploration. Finally, dMMR crypts were identified in 4 patients with VUS, argument in favor of the pathogenicity of those variants. Our study confirmed the presence of dMMR crypts in 80% of patients with LS and established an immunohistochemical protocol for their optimal detection. Furthermore, our data suggest that this tool is relevant in front of interpretation difficulties such as LLS or VUS. The presence of dMMR crypts could help in the diagnosis of authentic SL even if it would only have a positive value.

Cancer Posters - Wednesday

PB1009. A polygenic risk score for prostate cancer risk prediction.

Authors:

M. Shi¹, K. Schaffer¹, J. P. Shelley², M. Bagheri¹, J. Tosoian¹, J. Mosley³; ¹VUMC, NASHVILLE, TN, ²Vanderbilt Univ. Sch. of Med., Nashville, TN, ³Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Background: Prostate cancer is the second leading cause of cancer death among males. Screening protocols for prostate cancer frequently result in men undergoing a prostate biopsy that finds either no cancer or clinically insignificant (Grade Group [GG]=1) cancer. A polygenic risk score (PRS), which measures an individual's burden of common genetic risk variants, could improve risk stratification for high grade cancer. We compared the performance of a prostate cancer PRS against the Prostate Biopsy Collaborative Group (PBCG) clinical risk calculator to risk-stratify individuals with high-grade (GG>1) cancers. **Methods:** A retrospective cohort of 655 Black and White race men who underwent a first prostate biopsy were identified in Vanderbilt University Medical Center's BioVU resource. The prostate cancer PRS was based on a previously validated predictor comprising 269 SNPs identified in a trans-ancestry GWAS. To determine whether the PRS improved risk prediction for high grade (GG>1) cancer, as compared to the PBCG risk calculator, multivariable logistic regression model was used to test the association between PRS and the outcome of either any cancer or a GG>1 cancer with and without adjusted for the PBCG predictor. Prediction for the presence of a GG>1 cancer on prostate biopsy was evaluated based on discrimination (c-statistic), calibration and reclassification (Integrated Discrimination Improvement [IDI]). **Results:** Of 655 participants, 80 (8.2%) were of African ancestries, the median age was 63 (interquartile range [IQR], 57 - 69) years and the median PSA value was 5.3 (IQR, 4.3 - 6.8) ng/ml. The PRS is associated with a finding of any prostate cancer (age-adjusted odds-ratio, 2.0 [95% CI, 1.6 - 2.4]) and high-risk cancer (age-adjusted odds-ratio, 1.5[95% CI, 1.3 - 1.9]). The PRS, when added to the PBCG predictor, improved discrimination for identifying any cancer (c-statistic, 0.71 vs 0.67, difference= 0.042 [95% CI: 0.014, 0.073]), but not for GG>1 cancer (0.74 vs 0.74, difference = 0.002 [-0.011 - 0.012]). The IDI for any cancer was significant for any cancer (0.048 [0.015 - 0.078]) but was not significant for GG>1 cancer (0.009 [-0.008 - 0.022]). **Conclusions:** A prostate cancer PRS improved the prediction of any prostate cancer on biopsy, but did not improve discrimination, calibration, or reclassification for aggressive cancer, as compared to the PBCG risk predictor. This prostate cancer PRS may not help guide the decision-making to undergo a prostate biopsy.

Cancer Posters - Wednesday

PB1010. A ten-year-long experience of *BAP1*-Tumor Predisposition Syndrome diagnostics

Authors:

M. Sculco¹, M. La Vecchia¹, A. Aspesi¹, M. Clavenna¹, M. Salvo², G. Borgonovi¹, A. Pittaro³, W. Gianluca⁴, F. Napoli⁵, A. Listi⁵, F. Grosso⁶, R. Libener⁷, A. Maconi⁷, O. Rena⁸, R. Boldorini², D. Giachino⁹, P. Bironzo⁵, A. Maffè¹⁰, G. Ali¹¹, L. Elefanti¹², C. Menin¹², L. Righi⁵, C. Tampieri¹³, G. Scagliotti⁵, D. Ferrante¹⁴, E. Migliore¹⁵, C. Magnani¹⁴, D. Mirabelli¹⁵, G. Matullo⁹, I. Dianzani¹; ¹Dept. of Hlth.Sci., Università del Piemonte Orientale, Novara, Italy, ²Dept. of Hlth.Sci., Università del Piemonte Orientale; Unit of Pathology, AOU Maggiore Della Carità Hosp., Novara, Italy, ³Pathology Unit, AOU Città della Salute e della Scienza, Torino, Italy, ⁴Dept. of Med. Sci., Università di Torino, Torino, Italy, ⁵Dept. of Oncology, Università di Torino at San Luigi Hosp., Torino, Italy, ⁶6 Mesothelioma Unit, AO SS. Antonio e Biagio e Cesare Arrigo, Alessandria, Italy, ⁷Dept. of Integrated activities Res. and Innovation, AO SS. Antonio e Biagio e Cesare Arrigo, Alessandria, Italy, ⁸Thoracic Surgery Unit, AOU Maggiore della Carità, Novara, Italy, ⁹Med. Genetics Unit, Dept. of Clinical and Biological Sci., Università di Torino, AOU S. Luigi Gonzaga, Torino, Italy, ¹⁰Genetics and Molecular Biology Unit, Santa Croce e Carle Hosp., Cuneo, Italy, ¹¹Unit of Pathological Anatomy, Univ. Hosp. of Pisa, Pisa, Italy, ¹²Immunology and Diagnostics Molecular Oncology Unit, Veneto Inst. of Oncology IOV-IRCCS, Padova, Italy, ¹³Dept. of Med. Sci., Pathology Unit, AOU Città della Salute e della Scienza, Torino, Italy, ¹⁴Dept. of Translational Med., Unit of Med. Statistics, Università del Piemonte Orientale and Cancer Epidemiology, CPO Piemonte, Novara, Italy, ¹⁵Unit of Cancer Epidemiology, Città della Salute e della Scienza, Univ.-Hosp. and Ctr. for Cancer Prevention (CPO), Torino, Italy

Abstract Body:

Germline mutations in the tumor suppressor gene BRCA1-Associated Protein-1 (*BAP1*) gene are responsible for the *BAP1*-tumor predisposition syndrome (*BAP1*-TPDS), characterized by susceptibility to several tumor types, including mesothelioma, melanoma, renal cell carcinoma, and basal cell carcinoma. We report our ten-year experience in the molecular diagnosis of *BAP1*-TPDS. We sequenced by Sanger sequencing or targeted next-generation sequencing (tNGS) (using a custom panel designed to target: *BAP1*, *CDKN2A*, *CDK4*, *POT1* and *MITF*) the germline DNA samples from 101 individuals with suspected *BAP1*-TPDS. Moreover, a panel of 246 unrelated and unselected patients with mesothelioma were sequenced by tNGS using a custom panel of 107 cancer predisposing genes (including *BAP1*). We validated pathogenic variants (PVs) by assessing *BAP1* somatic loss in matching tumor specimens using immunohistochemistry (IHC), microsatellite analysis, and fluorescence *in situ* hybridization (FISH). We also performed multiplex ligation-dependent probe amplification (MLPA) on DNA extracted from fresh tumor samples. When feasible, cascade genetic testing was also performed. Moreover, *BAP1*-TPDS tumors were studied by tNGS (161 genes). Overall, we identified seven patients (7/101, 6.9%) carrying six different germline *BAP1* pathogenic variants (PVs), including one novel variant, among individuals with suspected *BAP1*-TPDS. Cascade testing identified other seven *BAP1* PV carriers. No somatic alteration besides *BAP1* was identified in the five tumors tested suggesting that *BAP1* is a strong driver of carcinogenesis. We also observed that certain tumors arisen in PV carriers do not show loss of the wild-type allele, suggesting a sporadic origin of the tumor or a functional role of heterozygous *BAP1* in carcinogenesis. Interestingly, we did not identify any germline PV in *BAP1* in the unrelated and unselected patient group (0/246, 0%), but we found that about 9% (23/246) of them carried germline PVs in other tumor suppressor genes, most of them involved in the DNA repair pathways. Altogether, our findings have important clinical implications in shaping personalized treatments.

Cancer Posters - Thursday

PB1011. Accelerated Tumor Only Variant Analysis Utilizing a GPU Framework.

Authors:

P. Vats¹, A. Sethia², H. Clifford³, M. Samadi⁴; ¹NVIDIA, Santa clara, CA, ²Nvidia, Santa Clara, CA, ³NVIDIA, Santa Clara, CA, ⁴NVIDIA, San Jose, CA

Abstract Body:

In the last decade, Next Generation Sequencing (NGS) cancer genetics studies have grown exponentially. Due to the rapid decline in sequencing costs and the advancement in sequencing technologies, large scale consortium-based cancer genomics studies have been made possible. Increased accessibility of these large-scale genomic studies has provided access to massive amounts of NGS data, which has created a major challenge for existing bioinformatics software to process and analyze in a timely manner. With recent advancements in oncology genomics workflows, clinically relevant information can now be obtained by sequencing oncology gene panels however it is often limited by tumor only sequencing. In this study, we employ a GPU (Graphics Processing Unit) accelerated framework to perform tumor-only variant analysis on a whole genome dataset from three publicly available studies. Tumor-only variant analysis workflows are not limited to whole genomes, but more frequently used in exome and large cancer gene panels studies. The accelerated framework provides significant speedup over the two most widely used bioinformatics software and presents an opportunity to perform consensus variant calling on tumor-only samples from NGS. In this benchmarking study, we applied two state of art CPU (Central Processing Unit) based variant callers Mutect2 and LoFreq, along with their GPU-accelerated counterparts on three different publicly available datasets. All three WGS datasets (dataset1: Tumor41x WGS dataset2: Tumor 46x WGS, and dataset3: Tumor 37x WGS) were processed through the original CPU and corresponding accelerated variant callers to evaluate runtime and performance accuracy. The accelerated framework provided an average 10x speedup for LoFreq and around 53x speedup for Mutect2 without compromising the results. For LoFreq, we were able to reproduce identical Single Nucleotide Variant (SNV) calls on all the three datasets (COLO829: 7830816, SEQC2: 14602852 and LINST: 8636026 SNV variants) in comparison to baseline variant callers (100% sensitivity and 100% precision, respectively). For Mutect2 we achieved 99.9% sensitivity and 99.9% precision on all three datasets. The results in this study highlight the strength of a GPU accelerated framework and displayed how large-scale tumor only variant calling can be performed in a timely manner, while maintaining accuracy.

Cancer Posters - Wednesday

PB1012. AmpliconClassifier detects the mechanisms of focal genome amplifications in cancer

Authors:

J. Luebeck, V. Bafna; UC San Diego, La Jolla, CA

Abstract Body:

Focal somatic copy-number amplifications (fSCNA) are an important class of targeted genomic amplifications found frequently in cancer genomes. fSCNA arise frequently from genome instability and provide a basis for the overexpression of oncogenes through rapid expansion of copy number. Importantly, fSCNAs may appear in chromosomal or extrachromosomal (ecDNA) forms, depending on the mechanisms by which they occur - and also dictating how they respond to selective pressures. Recent studies highlighted the prognostic importance of distinguishing ecDNA amplification from other events due to its association with worsened patient survival and more rapid tumor evolution (Kim et al., Nat. Gen., 2020). Importantly, due to the megabase scale of fSCNA events, and their complex, heterogeneous structures (Deshpande et al., Nat. Comms., 2019), unambiguous genome assembly to identify the mechanism of amplification remains challenging without ultra-long reads. However, it is still possible to identify the signatures of such events from individual structural variants (SVs) and SV chains. Consequently, even in datasets composed only of NGS data, the frequencies of events such as ecDNA and breakage-fusion-bridges (BFBs) may be determined.

We present AmpliconClassifier (AC), which examines signatures of genomic structural rearrangements to detect fSCNA events from whole-genome NGS data. AC utilizes copy number-aware genome breakpoint graphs generated by the structural variant analysis tool AmpliconArchitect. AC detects signatures of ecDNA and BFB by examining possible substructures identified in the breakpoint graph, identifying SV chains which match to known biological mechanisms of amplification. Given a collection of copy number-weighted paths, AC computes the fraction of amplification which ends up decomposed into cyclic paths to assist in the discovery of ecDNA. It further distinguishes BFB by searching for the presence of palindromic BFB subsequences and foldback SVs. AC also quantifies the structural similarity of fSCNA to detect clonal amplifications and quantify structural evolution.

We applied AC to a cytogenetically validated selection of 81 genes across 43 cancer cell lines, achieving 87% sensitivity in detecting ecDNA. Subsequently, we applied AC to 329 cancer cell lines from the Cancer Cell Line Encyclopedia, revealing ecDNA in 165 cell lines (50%), BFB in 62 cell lines (19%), and multiple independent ecDNA species in 85 cell lines (26%). The classification identifies important functional differences between samples carrying ecDNA, BFB, and other modes of oncogene amplification, pointing to the value of AC as a tool for profiling complex SVs.

Cancer Posters - Thursday

PB1013. Amplification of *RUNX1* in a patient with AML

Authors:

R. Hurtado¹, S. Tello-Vera², J. Juarez-Ynoñan³, R. O. Llontop¹, K. Eastwood⁴, M. Guardiola⁴, A. Lozada⁵, C. Tirado⁶; ¹Univ. Natl. Pedro Ruiz Gallo, Chiclayo, Peru, Peru, ²Hosp. Natl. Alanzor Aguinaga Asenjo, Chiclayo, Peru, Peru, ³Hosp. Natl. Alanzor Aguinaga Asenjo, Chiclayo, Peru, ⁴Baylor Scott & White, Temple, TX, ⁵Univ. of California, Los Angeles, Los Angeles, CA, ⁶Baylor Scott & White Hlth., Temple, TX

Abstract Body:

Acute myeloid leukemia (AML) is a heterogeneous disease, characterized by clonal expansion of undifferentiated myeloid precursors, leading to alterations in hematopoiesis and bone marrow failure. Characteristic chromosomal abnormalities in AML are translocations t(8;21), inv(16), t(15;17), t(9;22) as well as mutations of genes that regulate proliferation and survival (*FLT 3*, *PTPN 11*, *ETV 6 / PDGFB*), or genes responsible for differentiation and apoptosis (*RUNX-1 / RUNX 1 T 1*, *PML / RARA*, *KMT2A*, *CEBPA* and *CBFB*). Amplification of *RUNX1* is a rare event in AML. Herein we described a 60-year-old patient that was admitted to the hospital due to a clinical picture of symptoms of acute anemia, thrombocytopenia, leukocytosis, and profuse nasal bleeding. He reported that he had been taking Celecoxid for muscle cramps and denied comorbidities. A complete abdominal ultrasound was performed, which showed hepatomegaly, splenomegaly, and gallstones. The blood cell count indicated the presence of 72% blasts, 17% lymphocytes, 5% segmented, 4% myelocytes, and 2% monocytes. In the bone marrow aspirate, increased cellularity, infiltration by medium to large cells, ample cytoplasm, basophilic, granular, lax chromatin nucleus with 1 to 2 nucleoli, compatible with blasts of myeloid lineage, these blasts represent 97% of cellularity, neutrophils, and precursors represent <3% of the assessed cellularity. The flow cytometry study showed a population of precursors, MPO^{neg} / +, CD34 +, CD19^{neg} / + d, CD117 +, CD38^{neg} / +, HLA-DR ++, CD13^{neg} / +, CD33^{neg}, CD15^{neg}, D56^{neg}, CD123 +, CD7^{neg}, CD11b^{neg}, CD64^{neg}, CD41^{neg}, which represented 68% of the pathological cellularity. These findings are compatible with AML. Chromosome analysis showed an abnormal karyotype characterized as **47,XY, add(9)(p22),+21,i(21)(q10)x2[13]/44,ídem,-5,-6,-8[10]**. FISH Studies using the TEL/AML1 probe showed five signals for *RUNX1* in all cells examined and described as **nuc ish(RUNXT1x2,RUNX1x1~5)[200]**. Amplification of *RUNX1* is a rare event in AML with only few cases reported in the literature (mainly therapy related AML) and it is usually associated with poor prognosis.

Cancer Posters - Wednesday

PB1014. An active learning framework improves tumor variant interpretation.

Authors:

J. Capra¹, A. Blee², B. Li², T. Pecan³, Z. Nagel³, W. Chazin²; ¹Univ. of California San Francisco, San Francisco, CA, ²Vanderbilt Univ., Nashville, TN, ³Harvard Univ., Cambridge, MA

Abstract Body:

Accurate tools for interpreting variants of unknown significance (VUS) are essential for cancer precision medicine to reach its full potential to guide clinical decision-making. Tumor genetic variants are often critical in determining tumor progression and response to therapy. Thus, nonrecurrent tumor VUS pose a significant challenge to the implementation of precision medicine. However, current protein variant predictors are not accurate or specific enough for clinical use. The performance of most variant prediction tools is limited by the difficulty of acquiring sufficient training data to make testable mechanistic predictions.

To overcome these limitations, we developed an iterative “active learning” approach to predict the impact of tumor VUS in cancer relevant proteins from available biochemical, evolutionary, and functional annotations. Active learning augments traditional machine learning with iterative guided functional validation to refine the learned model. During each round, VUS that are most challenging to classify for the current model are functionally evaluated and then reincorporated with the phenotype label for subsequent iterations of algorithm training.

We first demonstrate the potential of active learning to improve variant interpretation on comprehensive deep mutational scanning (DMS) datasets for four cancer-relevant proteins. Given the consistent improvement observed, we then explore the utility of the approach to guide interpretation of tumor VUS in the nucleotide excision repair (NER) protein Xeroderma Pigmentosum Complementation Group A (XPA), a potential biomarker for cancer therapy sensitivity. We leverage a quantitative high-throughput cell-based NER activity assay to iteratively validate XPA VUS selected by the active learning strategy. In all cases, active learning yielded a significant improvement in variant effect predictions over traditional learning.

These analyses demonstrate that active learning is well suited to guide experimental validation and significantly improve interpretation of VUS in cancer patient genomes, a critical need in cancer research and precision medicine. Our new prediction model provides a foundation for deeper investigation of the correlation between genetic variants in NER genes and cisplatin sensitivity. More broadly, the active learning approach we introduce has potential for tumor VUS in other biomarkers of interest.

Cancer Posters - Thursday

PB1015. An AML patient showing an abnormal hyperdiploid karyotype.

Authors:

B. Hamid¹, C. Tran¹, H. Mendelsohn¹, C. Tirado²; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²Baylor Scott & White Hlth., Temple, TX

Abstract Body:

Acute myeloid leukemia (AML) is the most common acute leukemia in adults. It usually presents with an increase in myeloid blast cells and a decrease in mature cells and a reduction in mature red blood cells, granulocytes and platelets. Herein, we present a 71-year-old male with a history of non-Hodgkin lymphoma, fever, and pancytopenia. Cytogenetic studies indicated a deletion of 5q in 7 of the 20 metaphase cells examined as well as additions of chromosomes X, 2, 4, 8, 13, 19, 21 and 22 in 3 of the 3 metaphase cells with a 5q deletion, while the remaining thirteen cells show a normal karyotype(46,XY,del(5)(q13q33[4]/50~59,idem,+X,+2,+4,+8,+13,+19,+21,+22,+22[cp3]/46,XY[13]). Additionally, immunohistochemistry indicated small populations of CD34, CD117, CD3, PAX5, CD138 positive cells. Scattered small non-caseating granulomas are seen. Their significance remains unclear as they are typically associated with systemic infection, autoimmune disease, malignancy, and as a drug reaction, but staining indicated no acid fast organisms identified or fungal forms. Deletion 5q and hyperdiploidy are both common abnormalities for myeloid malignancies, such as MDS and AML, and the patient's complex karyotype involving both abnormalities is associated with genomic instability and a poor prognosis. However, the patient unexpectedly presented with a non-diagnostic immunophenotype with a mixture of T and B lymphocytes, T lymphocyte predominance, small polyclonal B lymphoid population, and no phenotype evidence of significant myeloid immaturity.

Cancer Posters - Wednesday

PB1016. An integrative genomic transcriptomic and proteomic investigation to characterize differences between isogenic radioresistant prostate cancer cell lines.

Authors:

R. Haas¹, S. Khan², G. Frame³, W. Zhao¹, B. Carlin¹, T. Yamaguchi¹, Y. Z. Bugh¹, J. Livingstone¹, C. Zhu¹, R. Hugh-White¹, S. Tao¹, A. Macklin², V. Ignatchenko², N. Kurganovs², D. Vesprini³, A. Loblaw³, M. R. Downes^{3,4}, T. Kislinger^{3,2}, P. C. Boutros^{1,3}, S. Liu^{3,4}; ¹Univ. of California, Los Angeles, CA, ²Univ. Hlth.Network, Toronto, ON, Canada, ³Univ. of Toronto, Toronto, ON, Canada, ⁴Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada

Abstract Body:

Relapse of Prostate Cancer (PC) following radiotherapy is a major clinical concern. Hypo-Fractionated (HF) radiotherapy, with high radiation doses per treatment, has recently become clinically preferable over Conventionally Fractionated (CF) due to better outcomes and substantial logistical benefits. Radioresistant phenotypes can emerge following both CF and HF treatments, but it is unknown whether or not similar mechanisms underlie resistance to these different therapeutic regimens. Here we investigated the genomic, transcriptomic, and proteomic differences between isogenic HF and CF radioresistant PC cells, that were originated from radiosensitive DU145 parental cells. We discovered that CF radioresistant cells gained twice the number of somatic single-nucleotide variation (sSNV) than HF. Nevertheless, the gained mutations irrespective of the treatment schedule converged on mutational signatures associated with DNA repair. The RNA abundance profiles of driver-genes and cancer-hallmark genes in CF cells were distinct from HF and parental profiles, which were relatively similar. The differences in protein abundance display cell-fraction-dependent clusters, foremost of which are elevated levels of DNA repair and P53 regulatory proteins in CF cell nuclei. Collectively, we observed a far more aggressive phenotype in CF cells compared to HF across all tested molecular levels. Finally, our top potential therapeutic targets for radioresistance were associated with clinical features in a cohort of ~300 PC patients, corroborating their involvement in shaping aggressive phenotypes. Our study provides a platform for the development of therapies for radio-recurrent prostate cancer. Ongoing proteogenomic integration will help understand the relationships amongst radioresistance-associated changes at the DNA, RNA, and protein levels.

Cancer Posters - Thursday

PB1017. Analysis of second-hit events in Tuberous Sclerosis Complex normal tissues and lesions.

Authors:

K. Klonowska¹, J. M. Grevelink², K. Giannikou^{1,3}, E. Aronica⁴, B. A. Ogorek¹, Z. T. Herbert⁵, A. R. Thorner⁶, T. N. Darling⁷, J. Moss⁸, D. J. Kwiatkowski¹; ¹Cancer Genetics Lab., Div. of Pulmonary and Critical Care Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, ²Boston Dermatology and Laser Ctr., Massachusetts Gen. Hosp., Boston, MA, ³Div. of Hematology and Oncology, Cancer and Blood Disease Inst., Children's Hosp. Los Angeles, Los Angeles, CA, ⁴Dept. of (Neuro)Pathology, Amsterdam UMC, Univ. of Amsterdam, Amsterdam, Netherlands, ⁵Molecular Biology Core Facilities, Dana-Farber Cancer Inst., Boston, MA, ⁶Ctr. for Cancer Genomics, Dana-Farber Cancer Inst., Boston, MA, ⁷Dept. of Dermatology, Uniformed Services Univ., Bethesda, MD, ⁸Pulmonary Branch, NHLBI, NIH, Bethesda, MD

Abstract Body:

Background: Tuberous sclerosis complex (TSC) is due to inactivating mutations in *TSC2* or *TSC1*, and is characterized by tumor development in multiple tissues, including facial angiofibroma (FAF), and cortical tubers. Several TSC tumors are known to develop through a two-hit mechanism. Here we explored frequency of point mutation second-hit events in TSC. **Methods:** We developed a Multiplex High-sensitivity PCR Assay (MHPA) (Klonowska et al., JCI 2022), enabling mutation detection at extremely low (<0.1%) variant allele frequencies (VAF) in *TSC1*, *TSC2*, and *TP53*. **Results:** *TSC2*-MHPA analysis of 81 samples, including 24 FAF biopsies from TSC patients with *TSC2* mosaicism, led to the discovery that UV-induced second-hit mutation causing *TSC2* inactivation was pervasive in TSC facial skin, with an average of 4.8 mutations per 2-mm biopsy at median VAF 0.08%, generating more than 150,000 incipient facial tumors (subclinical 'micro-FAFs') in the average TSC subject (Klonowska et al., JCI 2022). *TSC1*-MHPA analysis of 6 FAF biopsies from patients with *TSC1* mosaicism showed that UV mutations are much less common in *TSC1* (average 0.8 mutations) than in *TSC2* (average 4.8 mutations, $p=0.006$), consistent with the overall lower severity of FAF involvement in TSC patients with *TSC1* mutations than in those with *TSC2*. *TSC2*-MHPA analysis of TSC brain tubers identified 9 nonsynonymous SNV/indel somatic mutations in 6 of 13 (46%) tubers at median VAF 0.06%. No clear UV signature mutations were identified in any of the brain samples, in contrast to previously analyzed TSC normal skin and FAF. **Conclusions:** Our *TSC2*-MHPA analysis of TSC FAF revealed that TSC facial skin can be viewed as harboring a patchwork of clonal fibroblast proliferations (micro-FAFs) with indolent growth, a small proportion of which develop into clinically observable FAF. Such lesions are much less common in *TSC1* than in *TSC2* patients. *TSC2*-MHPA analysis of TSC tubers identified low frequency second-hit *TSC2* mutations in some lesions, that will be studied further and validated. **Funding:** The study was funded by the FY2020 Tuberous Sclerosis Alliance Postdoctoral Fellowship Award; the Engles Family Fund; and the NIH, National Heart, Lung, and Blood Institute (U01HL131022-04 and Intramural Research Program).

Cancer Posters - Wednesday

PB1018. Analysis of the spatial transcriptomic profiles in tumor and tumor microenvironment of colorectal cancer in Alaska Native patients.

Authors:

M. Thomas¹, H. Yin¹, D. Redwood², A. L. Koehne¹, L. Zhang³, J. J. Tiesinga⁴, T. A. Harrison¹, L. Hsu¹, C. I. Li¹, T. K. Thomas⁴, J. R. Huyghe¹, U. Peters¹; ¹Fred Hutchinson Cancer Ctr., Seattle, WA, ²Alaska Native Tribal Hlth.Co, Anchorage, AK, ³NanoString Technologies, Inc, Seattle, WA, ⁴Alaska Native Med. Ctr., Anchorage, AK

Abstract Body:

Tribal Health Organizations recognize the high incidence and mortality rates of colorectal cancer (CRC) among Alaska Native people and are undertaking initiatives to address it including improved screening and engaging in research. Studying the transcriptomic profiles in tumor and tumor microenvironment (TME) in Alaska Native CRC patients will provide new insights into the disease and identify predictors for survival. Therefore, we performed a comprehensive spatial tissue analysis using the Whole Transcriptome Atlas (WTA) panel on the NanoString GeoMx® Digital Spatial Profiler (DSP). WTA panel enables high-plex data profiling of 18,676 protein-coding RNA targets simultaneously combined with spatial information from any area of interest (AOI) from a single tissue section using next generation sequencing (NGS) readout. In this study, a tissue microarray (TMA) comprising a total of 48 cores from 13 CRC patients, was profiled for 96 AOI by separating tumor and TME tissue compartments based on segmentation on the pan-cytokeratin antibody morphology marker and targeted normal epithelia tissue. Each AOI was spatially profiled for up to 18,676 genes. Using the GeoMx NGS Pipeline software, Illumina FASTQ sequencing files were automatically processed to GeoMx readable digital counts (DCC) and input back into the Data Analysis Suite for analysis. The DCC were mapped back to each AOI, generating a map of transcript activity within each AOI. The R BioConductor package GeoMxWorkflows was used to perform the analysis. In this study, NanoString GeoMx WTA panel offered by Nanostring's TAP lab was utilized to perform spatially resolved expression profiling of CRC samples to understand the genes and pathways involved in CRC cancer. Sequencing quality was inspected for sufficient saturations (>50%) and data was normalized to third quartiles (Q3) to account for cellularity and area of interest (AOI) sizes. In the downstream analysis, 11,080 genes normalized by 3rd quartile, which were expressed above limit of quantification in at least 15% of AOIs. Differential expression analysis helped identify several genes that are significantly expressed between regions and enrichment analysis identified pathways associated with TME of CRC such as extracellular matrix, CTLA4 and cytokine-cytokine receptor interaction pathways. Spatial deconvolution showed abundant immune cell types in TME. This study will be used to help understand the underlying biological mechanisms of CRC tumorigenesis and better understand the role of immune response in the Alaska Native population.

Cancer Posters - Wednesday

PB1019. ArCCH: Improving the Performance of Clonal Hematopoiesis Variant Calling and Interpretation using a Consensus Based Approach

Authors:

I. Chan¹, A. Panchot¹, E. Schmidt¹, B. J. Wiley¹, J. Liu¹, L. Moukarzel², A. Schmidt¹, Y. Zhang³, B. Lajoie⁴, K. Blease⁴, J. Zhao⁴, S. Kruglyak⁴, C. E. Mason⁵, D. C. Link¹, K. H. Stopsack³, K. L. Bolton¹; ¹Washington Univ. in Saint Louis - Sch. of Med., Saint Louis, MO, ²Mem. Sloan Kettering Cancer Ctr., New York, NY, ³Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ⁴Element BioSci.s, San Diego, CA, ⁵Weill Cornell Med. Coll., New York, NY

Abstract Body:

The acquisition of somatic mutations in hematopoiesis stem and progenitor stem cells with resultant clonal expansion, clonal hematopoiesis (CH), is linked to a higher risk of hematologic malignancies and other adverse health outcomes. The prevalence of CH is dependent on sequencing technique. With ultra-high depth sequencing, CH can be detected in the most adults. CH variant calling is challenging due to difficulty in distinguishing low frequency CH mutations from sequencing artifacts. Here we evaluated the performance of three somatic variant callers (Mutect2, Lofreq, and Vardict) and an ensemble approach using deep targeted sequencing datasets generated from AML tumor dilutions and independent sequencing data with orthogonal validation. We sequenced a tumor dilution series from six AML patients and 27 technical controls using a targeted panel with unique molecular indexes including nine common CH genes (DNMT3A, TET2, ASXL1, TP53, PPM1D, JAK2, SF3B1, SRSF2) with a mean unique coverage of 18,268x. For higher VAF mutations (>5%), Mutect2, Lofreq, and Vardict showed reasonable sensitivity and positive predictive values (PPV) but poor performance at lower VAFs. We applied additional filters based on sequencing quality, depth, regional complexity, and an empiric estimate of sequencing error at the position of a given called variant using our technical controls. Taking the consensus of three callers and including our additional filters, we not only retained sensitivity and PPV for high VAF variants (1.00 sensitivity, 1.00 PPV) but improved the PPV with retained sensitivity for variants below 1% (0.91 sensitivity, 0.83 PPV). Applying our consensus approach to a cohort of 31 individuals sequenced using the same targeted panel with variants validated using an orthogonal approach, we saw excellent sensitivity (0.99 [>5%], 0.85 [1-5%], and 0.82 [0.1-1%]) and PPV (1.00 [>5%], 1.00 [1-5%], 0.95 [0.1-1%]). In addition, we will present results of a second independent sequencing dataset using the AVITIÔ System in order to benchmark the performance of our approach using a non-Illumina sequencing technology. In conclusion, we developed ArCCH, an advanced artifact filtering and consensus CH calling pipeline that includes custom annotation features to facilitate pathogenicity determination that is available as a workflow package. Our results show that a consensus approach with advanced error correction filters can substantially improve performance of CH variant calling beyond single variant callers.

Cancer Posters - Thursday

PB1020. Artemis disease-associated variants and their effect on VDJ recombination and genome instability.

Authors:

M. Gavilan, W. Lu, J. Sekiguchi; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Human patients with Severe combined immunodeficiency disease (SCID) have been previously identified to carry inactivating mutations in the *DCLRE1C* gene. *DCLRE1C* encodes the ARTEMIS DNA nuclease. ARTEMIS participates in V(D)J recombination, the lymphocyte-specific DNA rearrangement that assembles the variable region genes of both T and B cell receptors. The variable regions of T and B cell receptors recognize antigens from the environment and start the mechanism of defense. ARTEMIS also mediates the repair of DNA double strand breaks by non-homologous end-joining (NHEJ) in all cells and tissues. Many missense mutations and variants in *DCLRE1C* have been found to be associated with a wide range of immune system disorders, including late onset immunodeficiency, autoimmunity, autoinflammation, and in some patients, cancer predisposition. However, the impact of the majority of these mutations on ARTEMIS functions has not been defined. This study focuses on two mutations identified in patients that presented a variety of phenotypes. The first mutation, p.T71P, was identified in an immunodeficient patient with residual B and T cell production and no radiosensitivity. The second mutation, p.P171R, was identified in a patient with a progressive immune dysfunction, low levels of B and T cells, a mild radiosensitivity, DSB repair defects and carcinoma. The relationships between these specific *DCLRE1C* mutations on ARTEMIS activities and patient phenotypes are poorly understood. My research takes advantage of an inducible, chromosomally-integrated V(D)J recombination reporter construct in an *ARTEMIS* null pre-B cell line to define the impact of these disease associated mutations on V(D)J recombination, DNA repair and maintenance of genome stability.

Cancer Posters - Wednesday

PB1021. Assessing the relationship between coding variation and cancer risk in known cancer GWAS susceptibility regions

Authors:

A. Hammermeister Suger¹, T. Jia¹, A. Fohner¹, A. I. Phipps^{1,2}, T. A. Harrison¹, S. Lindstrom^{1,2}; ¹Univ. of Washington, Seattle, WA, ²Fred Hutchinson Cancer Ctr., Seattle, WA

Abstract Body:

Purpose: Expanding our understanding of the genetic architecture of cancer could provide insights into cancer biology to help mitigate the health effects of increasing global cancer incidence. Prior work primarily focused on the genetic basis of susceptibility to individual cancer types but elucidating the shared genetic etiology of multiple cancer types could provide novel insights into cancer etiology and development. **Methods:** Our analysis included 195,507 subjects from UK Biobank who had whole-exome sequencing data. We identified 41,670 cases diagnosed with at least one cancer (in situ or invasive cancer) using ICD9/10 codes representing more than 45 distinct tumor sites. We investigated pleiotropic cancer associations for coding variants in 7,421 autosomal genes located within 1 Mb of a previously identified cancer GWAS SNP. We included variants predicted to change protein function, including splice, frameshift, missense, and predicted loss-of-function variants in gene-based variant set analyses using a generalized linear mixed-effects model framework to robustly control for population stratification and cryptic relatedness. **Results:** In gene-based analyses, we identified a significant association between coding variation in *BRCA2* and cancer diagnosis ($p = 1.09 \times 10^{-6}$), and a suggestive association for *BRCA1* ($p = 6.08 \times 10^{-6}$). We also identified a significant single variant association ($p = 1.83 \times 10^{-18}$, OR = 1.12) with cancer diagnosis for the *MC1R* missense variant rs1805007, which has previously reported associations with skin cancer risk. **Conclusions:** Our results highlight the potential for investigating genetic contributions to cross-cancer risk using large, population-based samples with sequencing data. These results also support prior evidence that deleterious *BRCA1*, *BRCA2*, and *MC1R* mutations influence risk across a variety of cancer types.

Cancer Posters - Thursday

PB1022*. Assessment of genomic instability in aggressive forms of breast cancer in Mali

Authors:

C. Ongoiba, O. Samassekou, G. Landouré, M. Traoré, M. Keita; USSTTB, Bamako, Mali

Abstract Body:

Introduction: In Mali, breast cancer is the first and the deadliest cancer in women. More than 90% of breast cancer patients are diagnosed at stage III, and the median survival rate of patients barely reaches 2 years after diagnosis, underscoring the high malignancy and the dire prognosis of this cancer. Data have indicated that the level of malignancy is correlated to the degree of genomic instability whose assessment at diagnosis might pave the way for finding a predictive biomarker, lacking for some aggressive forms of breast cancer such as the triple-negative. Currently, challenges exist to accurately assess genomic instability from tissue sections which is the most widely used clinical material for diagnosis. In this study, we developed a new approach to accurately assess genomic instability in breast cancer at the cellular level. Then, we hypothesize that from this new approach, we can accurately categorize aggressive forms of breast cancer.

Methodology: We enrolled breast cancer patients at different stages, and used serial tissue sections to determine genomic instability by assessing the telomeric nuclear organization (3D telomeric fluorescence in situ hybridization, the 3D-microscopy, and the 3D-telomere analysis) and centromere copies (centromere copy number for six different centromeric loci). Thereafter, we superimposed the digital images of telomeres, centromeres, and the tissue section to generate integrated images and data which allow us to assess cellular heterogeneity and genomic instability. Finally, we used bioinformatics and statistical approaches to analyze the cellular and genomic data and correlate them to the clinical parameters of the patients.

Results: We found that the alteration of the telomeric nuclear organization and changes in centromeric copy number was associated with tumor stages. By combining the telomeric, centromeric, and classical pathological data, we were able to determine cellular heterogeneity as well as the degree of genomic instability from different tumoral regions. Most importantly, we detected high genomic instability in cells presenting low nuclear grade or in a high proportion of cells from samples of patients diagnosed at an early stage. Finally, we detected genomic instability from patient samples that were considered by classical pathology assessment as non-aggressive tumors.

Conclusion: we developed a new approach to accurately assess genomic instability in breast cancer, and this approach might help discover a predictive biomarker for aggressive forms. Most importantly, this new approach can be rapidly translated into clinics for the benefit of breast cancer patients.

Cancer Posters - Wednesday

PB1023. Association of the rs12587 variant in the KRAS gene in patients with breast cancer and control group of the Mexican population

Authors:

A. Garibaldi Rios^{1,2}, L. E. Figuera^{1,2}, A. Rivera Camaras^{1,2}, M. G. Márquez Rosales², M. T. Magaña Torres², G. M. Zúñiga González³, D. P. Pacheco Verduzco², B. C. Gómez Meda⁴, E. J. Maciel Cruz¹, H. Moreno Alcocer², M. P. Gallegos Arreola²; ¹Doctorado en Genética Humana, CUCS, UDG., Guadalajara, México., Mexico, ²División de Genética, CIBO, IMSS., Guadalajara, México., Mexico, ³División de Med. Molecular, CIBO, IMSS., Guadalajara, México., Mexico, ⁴Departamento de Biología Molecular y Genómica, Inst. de Genética Humana "Dr. Enrique Corona Rivera", CUCS, UDG., Guadalajara, México., Mexico

Abstract Body:

Background: Variant of the *KRAS* gene have been shown to be associated with cancer. However, their association with breast cancer (BC) has been inconsistent. **Objective:** This study aimed to determine the frequency with which the rs12587 variant of the *KRAS* gene is associated with BC in patients of the Mexican population. **Patients and methods:** The rs12587 A>C variant was determined by PCR real-time in 325 healthy Mexican subjects and 281 BC patients. **Results:** The rs12587 variant was associated with BC susceptibility when the BC patients and the control group were compared for the *AA* genotype (WT) it was observed in 72% of patients with BC vs 82% of controls, comparing them with each other, they showed significant differences [OR 0.51 (Confidence intervals CI95% 0.35-0.76), p=0.001], *AC* genotype (heterozygous) variant rs12587 was observed in 26% of the patients compared with 17% of controls [odds ratio OR 1.81, Confidence intervals CI95% 1.22 - 2.68, p=0.004) and the *CC* polymorphic genotype it was not significant. **Conclusions:** Our results conclude that variant rs12587 was associated with BC susceptibility in BC patients in the Mexican population.

Cancer Posters - Thursday

PB1024*. Associations of genetic ancestry to the somatic mutational landscape from tumor profiling data of 100,000 cancer patients

Authors:

F. De La Vega, B. Rhead, Y. Pouliot, J. Guinney; Tempus Labs, Inc., Chicago, IL

Abstract Body:

The incidence and mortality of cancer vary widely across race and ethnicity. Such variation is influenced by socioeconomic factors, environmental exposures, and genetic background. Individuals of non-European descent are underrepresented in cancer genomic studies, which limits a comprehensive understanding of disparities in the diagnosis, prognosis, and treatment of cancer among these populations. Furthermore, the social constructs of race and ethnicity are far from precise categories to understand the biological underpinnings of such differences. In this study, we use a large real-world data (RWD) patient cohort to examine associations of genetic ancestry with somatic alterations in cancer driver genes. We inferred genetic ancestry from approximately 100,000 de-identified records of patients with diverse histology who underwent tumor genomic profiling with the Tempus xT 648-gene next-generation sequencing (NGS) assay. We selected 654 ancestry informative markers that overlap the capture regions of the assay to infer global ancestry proportions at the continental level: Africa (AFR), America (AMR), Europe (EUR), East Asia (EAS), and South Asia (SAS). While most patients are of European descent (72%), our cohort includes 8 to 12-fold more patients with substantial (>50%) non-European ancestry than TCGA. Logistic regression was used to examine associations between continental ancestry proportions and presence of nonsynonymous somatic mutations and copy number alterations (CNA) in cancer genes, controlling for assay version, gender and age. P-values were adjusted for multiple testing by the Benjamini-Hochberg method to control the false discovery rate at 5%. We identify 7 significant associations with small somatic mutations and 15 with CNAs with non-European ancestries (all $p < 0.0001$). Among others, we found associations between small somatic mutations in *CTNNB1* with EAS ancestry (OR=1.44; odds per 20% increase in given ancestry proportion), *EGFR* with EAS (OR=1.49) and AMR (OR=1.78) ancestries in lung cancer, and *ASXL1* with AMR ancestry in brain cancer (OR=2.48). Furthermore, we identified several associations between ancestry and CNAs: *MTAP* with AMR (OR=1.45) and *EGFR* with SAS (OR=1.46) in lung cancer, among others. Finally, we observed a reduction in actionable mutations (OncoKB levels 1, 2, and R1) with AFR ancestry in *BRAF* (OR=0.73) and *EGFR* (0.77) in colorectal and lung cancers, correspondingly. Our results support the use of genetic ancestry inference on RWD to improve upon the use of race and ethnicity to better understand the impact of ancestry on mutational processes that influence cancer incidence, progression, and outcomes.

Cancer Posters - Wednesday

PB1025. B cell repertoire analysis in prostate cancer identifies gene expression pattern differences in FC gamma receptors.

Authors:

R. Reddy¹, D. Van Booven², A. Vedenko¹, H. Arora^{1,2,3}; ¹Desai Sethi Urology Inst., Miller Sch. of Med., Univ. of Miami, Miami, FL, ²John P Hussman Inst. for Human Genomics, Miller Sch. of Med., Univ. of Miami, Miami, FL, ³The Interdisciplinary Stem Cell Inst., Miller Sch. of Med., Univ. of Miami, Miami, FL

Abstract Body:

INTRODUCTION: The progression of prostate cancer (PCa), from a benign androgen (AR) dependent stage to AR independent castration and neuroendocrine stages, is often complemented by therapeutic resistance. Such resistance necessitates methods that can aid in comprehending the etiology of progression to better patient outcomes. Therapeutics such as immune checkpoint inhibition focus on restoring antibody diversity to destroy tumor cells that elude immune recognition, but treatment efficacy is limited to a subset of patients. This limitation arises mainly due to molecular heterogeneity and tumor adaptability. Therefore, this study assesses the immune modulatory patterns correlated with the stages of PCa progression. **METHODS:** First, the RNAseq raw counts (n=499) from the PRAD project of The Cancer Genome Atlas (TCGA) were acquired. Next, a general principal component analysis (PCA) was conducted, and then using Gleason scores (GS) as a measure of tumor severity, patients were stratified into subclasses. Subsequently, the software packages TRUST and MiXCR were utilized to evaluate immunoglobulin (Ig) profiles. Both total CDR3 sequences, taken from Ig heavy chains, and amino acid (AA) length were examined and compared with mutation rates and switching events. Lastly, all stratifications were compared to known tumor severity by GS. **RESULTS:** No patterns of Ig expression correlatable with PCa subclass were observed. However, Ig profiles significantly differed ($p=0.002$) between high grade PCa (GS9) and low grade PCa (GS6), with an average of 780 and 300 unique CDR3 sequences, respectively. A significant increase in somatic mutations was also noted (GS9: 0.75 vs. GS6: 0.33, $p=2.2e-5$). Furthermore, a significant correlation existed when looking at differential expression in 6 of the 7 genes in the FC gamma receptor gene family (mean $p=0.00005$). Finally, with regards to the Shannon entropy index, lower diversity correlated with lower tumor severity, suggesting a significant change in the immune milieu ($p=0.03$). **CONCLUSIONS:** Overall, the findings indicate that there are immunomodulatory patterns associated with stages of PCa, and their correlation with components such as mutations, GS, and subclasses of PCa can help further elucidate the underpinnings of PCa progression and develop personalized treatment.

Cancer Posters - Wednesday

PB1026. Breast cancer genetics in non-Ashkenazi breast cancer patients - carrier rates and variant classification in a genetically diverse population

Authors:

S. Lieberman¹, O. Murik², F. Zahdeh², R. Beeri¹, P. Renbaum¹, M. Barzily¹, T. Klopstock¹, A. Tomer¹, D. A. Zeevi², T. Mann², O. Olsha³, T. Hadar^{3,4}, M. Carmon³, H. Goldvaser^{5,4}, S. Paluch-Shimon^{5,4}, E. Friedman^{6,7}, A. Lahad^{8,9}, R. Bernstein-Molho^{10,7}, **E. Levy-Lahad**^{1,4}; ¹Shaare Zedek Med. Ctr., Med. Genetics Inst., Jerusalem, Israel, ²Shaare Zedek Med. Ctr., Translational Genomics Lab, Med. Genetics Inst., Jerusalem, Israel, ³Shaare Zedek Med. Ctr., Dept. of Surgery, Jerusalem, Israel, ⁴Hebrew Univ., the Faculty of Med., Jerusalem, Israel, ⁵Shaare Zedek Med. Ctr., Dept. of Med. Oncology, Jerusalem, Israel, ⁶Sheba Med. Ctr., The Meiray Ctr. High Risk Clinic, Tel Hashomer, Israel, ⁷Tel Aviv Univ., Sackler Sch. of Med., Tel Aviv, Israel, ⁸Hebrew Univ., Dept. of Family Med., Jerusalem, Israel, ⁹Clalit Hlth.Services, Jerusalem, Israel, ¹⁰Sheba Med. Ctr., Dept. of Med. Oncology, Tel Hashomer, Israel

Abstract Body:

Background: There is limited information on the landscape of inherited breast cancer (BC) in non-Ashkenazi Jews. This heterogeneous, little tested population can offer insights into novel variants in known BC-predisposition genes. Methods: In 2015-2021 consecutively diagnosed non-Ashkenazi BC patients underwent multigene panel testing (MGPT). MGPT was also offered to a control group of unaffected non-Ashkenazi participants. Results: Genetic testing was performed in 743 affected and 813 unaffected women. *BRCA1/BRCA2* pathogenic variants (PV) were identified in 26 (3.5%) affected vs. 5 (0.6%) unaffected women ($P < .0002$). Only 3 *BRCA1* PVs were recurrent. PVs in other genes were found in 33 (4.6%) affected vs. 8 (1%) unaffected women. This included PVs in *ATM*, *CHEK2*, *BRIP1*, *TP53*, *MRE11*, *FANCM*, *NBN*, *PALB2* and *PTEN*. Rare variant classification: 9490 rare variants ($< .01$ MAF) were observed in 28 genes: 695 (7.3%) exonic, 8789 (92.7%) non-coding. 541 (78%) exonic variants were previously in ClinVar. Using frequencies from this cohort enabled downgrading of 15 ClinVar variants of uncertain significance (VUS) to Benign, based on MAF > 0.01 among unaffected low family history (FH) women. MAF by sub-ethnicity downgraded another 28 VUSs. Among non-coding variants, 68 were significantly more common in affected with significant FH, suggesting possible pathogenicity. Conversely, 52 non-coding variants were classified as benign based on higher frequency in controls. Conclusion: Non-Ashkenazi women exhibit low rates of PVs in known breast cancer genes, even in affected women with substantial FH. Genetic analysis in diverse populations can contribute to variant classification, especially of non-coding variants whose interpretation by standard tools is limited.

Cancer Posters - Wednesday

PB1027*. Breast cancer polygenic risk score as a tool to risk stratify participants in the Healthy Nevada Project.

Authors:

C. Hajek¹, G. Elhanan², K. Schiabor Barrett¹, E. Cirulli³, N. Washington¹, J. Lu¹, A. Bolze¹, J. Grzymalski²; ¹Helix, San Mateo, CA, ²DRI, Reno, NV, ³Helix, Lakeside, CA

Abstract Body:

Background: Providers use clinical risk assessments to identify a woman at high lifetime risk for breast cancer who could thus benefit from enhanced screening measures such as breast MRI or evaluation in high risk breast cancer clinics. However, the application and usage of these assessments is inconsistent and based on factors such as the provider familiarity, or the patient's insurance. In Nevada, a new law requires payors to cover counseling and BRCA1/2 testing for women who meet certain family health history criteria. In parallel, polygenic risk scores (PRS) are emerging as a clinically relevant genomic predictor of breast cancer risk. In population genomics programs, all of these tools could be used to provide a more consistent approach to identifying women at increased risk for breast cancer.

Objectives: Assess the impact of using (i) BRCA1/2 pathogenic variants, (ii) family history and/or (iii) a polygenic risk score to risk stratify patients in the overall population in a US health system.

Methods: The analysis was based on genetic information and Renown Health electronic health records for 21,427 women participants in the Healthy Nevada Project.

Results: All three tools independently identified women at higher risk of breast cancer. Women with a BRCA1/2 pathogenic variant (n=128) had a 7.1x increased risk of breast cancer compared to those without a BRCA1/2 pathogenic variant. Women who reported a positive family history (n=1,431) had a 5x increased risk of breast cancer compared to those without a known family history. And women with a polygenic risk in the top 10% (n=2,118) had a 2x increased risk compared to women with a polygenic risk in the lower 90%.

However, each tool alone missed women at high risk. For example, only 49 (38%) of the women with a BRCA1/2 pathogenic variant reported a family history. A combination of these tools may be the best way to risk stratify patients. We showed that polygenic risk and family history could be combined. Women with a positive family history, negative for BRCA1/2 and in the top 10% of the PRS (n=167) had a 1.5x increased chance of developing breast cancer compared to women with a positive family history, negative for BRCA1/2 and in the bottom 90% of the PRS (n=1,264). The combined positive family history and high PRS group risk for breast cancer was similar to women with a BRCA1/2 pathogenic variant (27% vs. 33.6%).

Overall these results show that genetic screening for BRCA1 and BRCA2 pathogenic variants, collecting family history information, and calculating a polygenic risk score based on common variants are complementary tools that can be used together to stratify patients who would benefit from enhanced screening.

Cancer Posters - Thursday

PB1028. Can impairment of kidney function cause kidney cancer? A bidirectional Mendelian randomization study.

Authors:

Y. Yang¹, Y. Lin^{2,3}, T. Fu¹, G. Liao⁴, B. Liao⁵, J. Huang¹; ¹Med. Device Regulatory Res. and Evaluation Ctr., West China Hosp., Sichuan Univ., Chengdu, China, ²West China Hosp., Sichuan Univ., Chengdu, China, ³Program in Genetic Epidemiology and Statistical Genetics, Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ⁴State Key Lab. of Oral Diseases, Natl. Clinical Res. Ctr. for Oral Diseases, West China Hosp. of Stomatology, Sichuan Univ., Chengdu, China, ⁵Dept. of Urology, Inst. of Urology (Lab. of Reconstructive Urology), West China Hosp., Sichuan Univ., Chengdu, China

Abstract Body:

Impairment of kidney function was reported to have a bidirectional relationship with kidney cancer in many observational epidemiology studies. However, whether these two kinds of traits were causally linked was still unclear. In the present study, two-sample Mendelian randomization (MR) methods were adopted to investigate the bidirectional causal relation between kidney cancer and kidney function biomarkers (creatinine-based estimated glomerular filtration rate (eGFR_{crea}), cystatin C-based estimated glomerular filtration rate (eGFR_{cys}), blood urea nitrogen (BUN), serum urate, and urinary albumin-to-creatinine ratio (UACR)). We used genetic instruments for these 5 kidney function traits, from a range between 288,649 and 1,004,040 European ancestry individuals, and kidney cancers from 408,786 European ancestry individuals (N_{case}:1338). To help down weight the potential bias of weak instruments, pleiotropy, and extreme outliers, we applied two state-of-art MR methods, contamination mixture (ConMix) and Robust Adjusted Profile Score (RAPS) in our main analysis. Various sensitivity analyses including other MR methods were also conducted to support the main results. In the forward, MR the eGFR_{crea} was found to play a negative role in the risk of kidney cancer (OR=0.007, 95%CI: 2.6×10⁻⁴ to 0.569, P=0.041). After adjusting for body composition and diabetes, genetically determined urate and UACR may potentially have a causal relationship with kidney cancer (P<0.05). Conversely, genetically determined kidney cancer was found to be causally associated with a decreased level of eGFR_{cys} (OR=-9.56×10⁻⁴, 95%CI: -9.56×10⁻⁴ to -9.56×10⁻⁴, P=0.003). Meanwhile, kidney cancer could increase the level of BUN with marginal significance (P = 0.070). Our results highlighted the bidirectional causal effect of kidney dysfunction on kidney cancer. Clinically, eGFR_{crea} and UACR could both be important predictors for the risk of kidney cancer, especially when patients were comorbid with obesity or diabetes. The level of eGFR_{cys} should be carefully monitored once diagnosed with kidney cancer, to identify whether the kidney function was harmed, regardless of the nephrectomy performed.

Cancer Posters - Thursday

PB1030. Cancer subclone detection based on DNA copy number in single cell and spatial omic sequencing data.

Authors:

C-Y. Wu¹, A. Sathe², J. Rong¹, P. Hess¹, B. Lau², S. Grimes², H. Ji², N. Zhang³; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Stanford Univ., Palo Alto, CA, ³Univ Pennsylvania, Philadelphia, PA

Abstract Body:

Cancer results from somatic mutations such as copy number alterations (CNAs) that continue to accumulate during disease progression. These mutations can lead to functional heterogeneity within tumors and can influence the efficacy of cancer therapy. Therefore, studying the functional characteristics and spatial distribution of genetically distinct subclones is crucial to the understanding of tumor evolution and the design of cancer treatment. Here, we present Clonalscope, a method for subclone detection using copy number profiles that can be applied to spatial transcriptomics (ST) data and data from single-cell sequencing platforms such as scRNA-seq and scATAC-seq. Clonalscope implements a nested Chinese restaurant process, which mimics the tumor evolutionary process, to identify *de novo* subclones within one or multiple samples from the same patient. Clonalscope incorporates prior information from paired whole-genome or whole-exome sequencing (WGS/WES) data to achieve more reliable subclone detection and malignant cell labeling. On scRNA-seq and scATAC-seq data from four gastrointestinal tumor samples, Clonalscope successfully labeled malignant cells and identified genetically different subclones, which were validated in detail using matched scDNA-seq data. On ST data from a squamous cell carcinoma and two invasive ductal carcinoma samples, Clonalscope was able to label malignant spots, trace subclones between associated datasets, and identify spatially segregated subclones expressing genes associated with drug resistance and survival. Using Clonalscope, we detected subclones based on (allele-specific) copy number profiles in spatial and single cell tumor sequencing data, enabling the investigation of the interplay between genome, transcriptome, and spatial environment during tumor evolution.

Cancer Posters - Wednesday

PB1031. Cancer-driving mutations are enriched in genic regions intolerant to germline variation

Authors:

R. Dhindsa¹, D. Vitsios², D. Matelska², J. Mitchell³, Z. Zou², J. Armenia², F. Hu², Q. Wang⁴, B. Sidders², A. Harper², S. Petrovski⁵; ¹Baylor Coll. of Med., Houston, TX, ²AstraZeneca, Cambridge, United Kingdom, ³AstraZeneca, Melbourn, United Kingdom, ⁴AstraZeneca, Chapel Hill, NC, ⁵AstraZeneca, Fulbourn, United Kingdom

Abstract Body:

Large reference datasets of protein-coding variation in human populations have allowed us to determine which genes and genic sub-regions are intolerant to germline genetic variation. There is also a growing number of genes implicated in severe Mendelian diseases that overlap with genes implicated in cancer. Despite their success in prioritizing germline variants, population genetics-based approaches have yet to be applied in the context of distinguishing between somatic cancer driving mutations and neutral “passenger” mutations. Here, we hypothesized that cancer-driving mutations might be enriched in genic sub-regions that are depleted of germline variation relative to somatic variation. We introduce a new metric, OncMTR, which uses 125,748 exomes in the gnomAD database to identify these genic sub-regions. We demonstrate that OncMTR can significantly predict driver mutations implicated in hematologic malignancies. Divergent OncMTR regions were enriched for cancer-relevant protein domains, and overlaying OncMTR scores on protein structures identified functionally important protein residues. OncMTR outperformed every population- and evolutionary-based scores, including CADD, ncER, LINSIGHT and phyloP in predicting leukemia-causing missense variants. Finally, we performed a rare variant, gene-based collapsing analysis on an independent set of 394,694 exomes from the UK Biobank and find that OncMTR dramatically improves genetic signals for hematologic malignancies, with up to 1,000-fold increased effect sizes for certain associations. Our web app enables easy visualization of OncMTR scores for each protein-coding gene (<http://oncmtr.public.cgr.astrazeneca.com>).

Cancer Posters - Thursday

PB1032. Case report: Olaparib use in metastatic lung adenocarcinoma with *BRCA2* pathogenic variant

Authors:

J. Soon, D. Tan Shao Weng, J. NGEOW, J. Chiang; Natl. Cancer Ctr. Singapore, Singapore, Singapore

Abstract Body:

Poly (ADP-ribose) polymerase (PARP) inhibitors are approved in malignancies associated with germline *BRCA1* or *BRCA2* pathogenic variants, such as breast, ovarian, prostate and pancreatic cancer. In cancers not associated with germline *BRCA1* or *BRCA2* pathogenic variants (PV), the therapeutic relevance of PARP inhibitors is less clear. We describe a 64 year old Chinese gentleman with a smoking history of forty pack years who presented with a skin lump over the epigastrium. Family history was significant as his sister had breast cancer in her forties and had surgery. Excision of the lump led to the diagnosis of a metastatic lung adenocarcinoma with an *EGFR* Exon 18 mutation, with c.2126A>C(p.Glu709Ala) and c.2156_2157delins CA(p.Gly719Ala). Afatinib was started with partial response. After four months, he progressed with symptomatic left femur and ulnar metastases. Afatinib was resumed after orthopaedic intervention and radiotherapy, continued till progression, with a disease control of eleven months. Liquid biopsy and repeat lung biopsy were negative for T790M mutation. The lung biopsy showed a PD-L1 TPS score of 70%. After two cycles of pembrolizumab, he progressed with symptomatic brain metastases, and whole brain radiotherapy was delivered. Next generation sequencing of the lung biopsy revealed a somatic *BRCA2* PV. Analysis of the tumor showed somatic *BRCA2* c.1411G>T(p.Glu471Ter) variant with variant frequency of 28.4%, *RBI* c.1807G>A(p.Ala603Thr) variant with variant frequency of 35.6%, and estimated tumour mutation burden of 14 Muts/Mb. The two *EFGR* mutations found on initial molecular testing were found. Three variants of unknown significance (VUS) in *ATR*, *CREBBP* and *SLX4* gene were found. Olaparib was started at 300mg twice a day. He achieved partial response four months later, and his progression free survival was extended by eight months. In view of a somatic *BRCA2* PV and positive family history of *BRCA2* associated cancer, germline testing was performed on his blood. This showed *EGFR* c.3353C>T(p.Ala1118Val) VUS. No germline *BRCA2* PV was found. Post olaparib, he received two cycles of pemetrexed carboplatin and then two cycles of docetaxel before demise. The G>T variant within *BRCA2* at genome coordinate Chr13:32332889 is predicted to result in an amino acid substitution of glutamic acid with a stop codon, causing a nonsense variant with loss of normal protein function. Loss of function variants in *BRCA2* are pathogenic which increases the risk of tumorigenesis and susceptibility to PARP inhibitors. This study shows that patients with metastatic lung adenocarcinoma with somatic *BRCA2* PV may respond to PARP inhibitors, expanding the indications of olaparib.

Cancer Posters - Wednesday

PB1033. Case-control likelihood ratio calculation for clinical classification of *BRCA1* and *BRCA2* variants of uncertain significance

Authors:

M. Zanti¹, D. O'Mahony¹, M. T. Parsons², H. Li³, J. Dennis⁴, B-J. Feng³, D. F. Easton⁴, F. J. Couch⁵, A. B. Spurdle², D. E. Goldgar³, K. Michailidou¹, Breast Cancer Association Consortium (BCAC), Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA); ¹Biostatistics Unit, The Cyprus Inst. of Neurology and Genetics, Nicosia, Cyprus, ²Population Hlth.Res. Program, QIMR Berghofer Med. Res. Inst., Brisbane, Australia, ³Huntsman Cancer Inst., Univ. of Utah, Salt Lake City, UT, ⁴Ctr. for Cancer Genetic Epidemiology, Univ. of Cambridge, Cambridge, United Kingdom, ⁵Mayo Clinic, Rochester, MN

Abstract Body:

Following the recommendations and guidelines of the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP), population data can be used to inform the interpretation of variants identified via clinical genetic testing. The specific criterion states that a Relative Risk (RR) or an Odds Ratio (OR) > 5.0 , with nominal significance, from case-control data can provide strong evidence in favor of pathogenicity (PS4). We detail a likelihood ratio-based method incorporating gene-specific age-related penetrance. We used this method to derive case-control likelihood ratios for rare variants in the *BRCA1* and *BRCA2* genes, using a large series of case-control studies (75,657 breast cancer cases and 52,987 healthy controls) genotyped as part of the Breast Cancer Association Consortium (BCAC) OncoArray project. We also performed power calculations using simulated case-control data. This likelihood ratio method provides more informative results compared to classical OR calculations. A larger number of variants reached evidence in favor of pathogenicity, and a substantial number of variants had evidence against pathogenicity - findings not explicitly captured in the ACMG/AMP guidelines. It is also more powerful, as shown by the analysis of simulated data. We provide user-friendly scripts for implementation of the method for rare variants in *BRCA1*, *BRCA2* and other high-risk genes with known penetrance.

Cancer Posters - Thursday

PB1034. Cell-free DNA fragmentation patterns predict the response to immunotherapy in head and neck squamous cell carcinomas

Authors:

H. Zheng¹, H. Fu¹, L. Wang¹, S. Gulati², D. Hildeman¹, T. Wise-Draper², Y. Liu¹; ¹Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, ²Univ. of Cincinnati, Cincinnati, OH

Abstract Body:

Background

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide, with a ~50% recurrence rate. Immunotherapy by nivolumab or pembrolizumab, which the FDA recently approved, could significantly improve the survival rate of metastatic or recurrent HNSCC patients. However, only ~18% of HNSCC patients benefit from immunotherapy. It is urgent to identify the clinical biomarkers for predicting HNSCC patients that can benefit from the immunotherapy treatment.

Methods

We developed a cell-free DNA (cfDNA) fragmentomics-based approach to predict the immunotherapy responses in head and neck cancer patients by using low-pass (~1X) plasma whole-genome sequencing (WGS). In a prospective cohort across six institutes, we collected 184 plasma samples up to three different time points from 92 HNSCC patients receiving pembrolizumab therapy. Based on the cfDNA fragmentation pattern, we built a machine learning model to predict the responses to the immunotherapy with the area under the curve (AUC) of 0.79 even before the treatment, which shows a higher performance than other reported state-of-art predictive biomarkers. Based on the response prediction by cfDNA fragmentomics, we performed a Kaplan-Meier survival analysis. Patients predicted to be responsive to the therapy showed significantly different disease-free survival (DFS) and overall survival (OS) curves that were predicted to be non-responsive (p-value < 0.0001, log rank test). However, PD-L1 expression measured by Immunohistochemistry (IHC) assay from the same patients showed lower performance on the prediction of responses to immunotherapy (AUC=0.7) and insignificant separation of DFS and OS curve between groups.

Conclusion

Our study illustrates the potential prognosis power of plasma cfDNA fragmentomics to monitor the HNSCC patients after the immunotherapy. Potentially, our findings can be used as a screening method to determine if a patient with HNSCC benefits from PD-1/PD-L1 immunotherapy prior to the treatment.

Cancer Posters - Wednesday

PB1035. Cell-intrinsic metabolism and cofactor balancing define the metabolomic impact of NAD⁺ supplementation in pancreatic cancer cell metabolism

Authors:

S. Milosavljevic^{1,2}, X. Li¹, S. H. Elsea¹; ¹Baylor Coll. of Med., Houston, TX, ²Harvard Med. Sch., Boston, MA

Abstract Body:

While NAD⁺ precursor supplementation has been purported to prevent cancer and extend life as an essential component of DNA repair and cell growth, multiple studies have shown contrasting data for this supplement use in cancer prevention and treatment. Cancer is often characterized by an oxidative redox shift, with decreased redox buffer protection and increased production of free radical reactive oxygen species. Furthermore, with aging comes a decline in the levels of cofactors and coenzymes that function in the metabolic pathways that moderate and buffer oxidative stress in healthy cells. Our data have shown that pancreatic cancer (PaCa) etiology may involve dysregulation of NADK and allied pathways in NAD⁺ metabolism to facilitate metabolic reprogramming by reducing reactive oxygen species and increasing the production of NADPH, supporting cancer cell growth and proliferation. However, these findings contradict the common practice of NAD⁺ supplementation in cancer prevention. We purport that cofactor balancing is critical for cellular homeostasis and that altered metabolic pathway function is dependent upon both diet-sourced cofactors and genetic variation. Our hypothesis is that dysregulation of NADK and its allied pathways causing either increased or reduced oxidative stress may manifest as cancer cell proliferation in a cell-intrinsic manner, potentially dependent on genetic variation involving NADK and other genes in these linked pathways. To investigate this theory, we used untargeted metabolomics to define the cellular and biochemical impact of NAD⁺ supplementation and depletion in both healthy and PaCa cells to assess the effects of either enhancing or impairing NAD⁺ biosynthesis and utilization. PaCa cell lines with different genetic backgrounds were treated with a panel of NAD⁺ precursors and inhibitors of NAD⁺ biosynthesis to specifically pinpoint key molecules and entry points with the greatest impact on cellular pathways and redox control in PaCa metabolism. Metabolomic data indicate multiple perturbations in purine and pyrimidine metabolism, endocannabinoid pathways, the folate and methionine cycle, redox control, and nicotinate and nicotinamide biosynthesis in PaCa cells, highlighting the broad cellular impact of targeting this essential cofactor. Further, identifying the cell-intrinsic metabolic pathways affected by altered NAD⁺ metabolism are essential as these pathways underlie an individual's risk for disease versus maintaining healthy aging in the context of PaCa. These data will elucidate optimal cofactor utilization for redox metabolic control/balance and will reveal key targets for PaCa treatment.

Cancer Posters - Thursday

PB1036. Characterization of molecular differences between normal weight and overweight/obese cancer patients

Authors:

F. Huang¹, P. Xu², Z. Yue¹, M. Gao¹, Z. Chong²; ¹Univ. of Alabama at Birmingham, Birmingham, AL, ²UAB, Birmingham, AL

Abstract Body:

Overweight and obesity refer to abnormal or excessive body fat accumulation and associated comorbidities that pose a high burden on human health. Epidemiological studies have demonstrated strong associations between overweight/obesity and multiple cancer types. However, the differences in molecular features between overweight/obese and normal-weight cancer have not been investigated systematically. To comprehensively characterize the molecular differences between normal-weight and overweight/obese cancers in a large cohort of cancer patients, we performed a comprehensive analysis using the data from The Cancer Genome Atlas (TCGA) and investigated overweight/obesity-biased characteristics including mutation patterns, gene expression, and immune features across multiple cancer types. To overcome the confounding effects, we applied a propensity score weighting algorithm to balance confounder factors including age, gender, race, and tumor stage that may introduce bias to the molecular features. Through rigorous, multidimensional analysis, we identified that the mutation patterns in overweight/obese tumors were distinct from normal-weight tumors and identified several obesity-biased mutated genes in several cancers. The overweight/obesity-biased mutational patterns reflected differences in endogenous oncogenic processes such as defective DNA mismatch repair signature. We also identified obesity-biased mutated genes CDH11, MDGA2, HUWE1, FAT4, and SYNE1 in esophageal carcinoma (ESCA); BCOR, ATM, CSMD1, PPP2R1A, MTOR in uterine corpus endometrial carcinoma (UCEC). The differential expression genes (DEGs) between overweight/obese and normal-weight cancers were enriched in several immune and hormone relative pathways. In addition, compared to normal-weight tumors, overweight tumors did not show higher immune cells fractions, while obese tumors had significantly higher non-activated T cells, mast cells, macrophages, and lymphocytes in multiple cancers. All these differences may indicate the underlining molecular mechanisms of overweight/obesity in cancers. Overall, our study provides a systematic molecular-level understanding of the molecular basis for body weight disparities in multiple cancer types and sheds light on future development of precision therapy for overweight/obese associated cancers.

Cancer Posters - Wednesday

PB1037. Characterization of nucleosome positions and chromatin interactions of regulatory elements comparing Hi-C, Micro-C, and promoter capture Micro-C.

Authors:

S. Rhie, B. Lee, Z. Wu; Univ. of Southern California, Los Angeles, CA

Abstract Body:

Three-dimensional chromatin organization is important for gene regulation. To study chromatin interactions of regulatory regions, we performed Hi-C, Micro-C, and promoter capture Micro-C in human prostate cancer cells. We found that Micro-C identified more high-resolution chromatin loops including ones around structural variants than Hi-C. By evaluating the effect of deep-sequencing, we revealed that more than 2 billion reads of Micro-C are needed to capture chromatin interactions at 1kb resolution. We also showed that promoter capture Micro-C identifies chromatin loops near promoters with a lower amount of sequencing reads. By integrating Micro-C with NOMe-seq and ChIP-seq, we found that promoters that are involved in chromatin loops have a more accessible region with lower levels of DNA methylation and more highly phased nucleosomes compared to promoters that are not involved in loops. This work provides a framework for understanding the chromatin loops among regulatory elements and nucleosome-depleted regions in human cancer cells.

Cancer Posters - Thursday

PB1038. Characterization of the *2q22* renal cancer susceptibility locus

Authors:

A. Souza¹, L. Costa¹, L. Jessop¹, T. Myers², K. Brown³, S. Chanock⁴, L. Colli⁵; ¹USP, Ribeirão Preto, Brazil, ²NIH, Bethesda, MD, ³Natl. Cancer Inst., Rockville, MD, ⁴Natl. Cancer Inst, Rockville, MD, ⁵Univ. of Sao Paulo, Ribeirão Preto, Brazil

Abstract Body:

Genome-wide association studies (GWAS) have identified approximately 13 different genomic regions associated with the risk of developing RCC, however, most of these regions have not yet been functionally characterized. Among these, we seek to characterize the region 2q22 marked by rs12105918 ($p = 1.80 \times 10^{-8}$; odds ratio 1.29) which presents the SNP rs72858496. This SNP has a chromosomal loop that allows physical interaction with the promoter region of the *ZEB2* gene and is located in an intron region. To determine whether rs72858496 acts as an enhancer element in the context of renal cancer and to understand the function of its target gene (*ZEB2*), we adopted two recent techniques that modulate gene expression, CRISPRa (CRISPR activation) for gene activation, and CRISPRi (CRISPR interference) for gene expression interference. For CRISPRi, we targeted a catalytically inactive cas9 (dCas9) fused to a KRAB-ZIM3 repressive domain and for CRISPRa, we use a dCas9 fused to an activation domain (VP64). Both techniques were performed by targeting the SNP region in UO-31 and Caki-1 cell lines, as well as the transcription start site (TSS) of the *ZEB2* gene in cell lines (UO-31, 786-O, ACHN, HK2). Our initial results point to rs72858496 as a potential functional SNP in the 2q22 region, decreasing *ZEB2* gene expression when the SNP region is blocked by the dCas9-KRAB-ZIM3 domain targeted by 3 out of 4 different gRNAs targeting rs72858496 in the UO-31 cell line. In conjunction with data obtained by electrophoretic mobility shift assay (EMSA), we note that this SNP has a role associated with the gene modulation of *ZEB2*. The generation of stable cell lines for the dCas9-VP64 and dCas9-ZIM3 domains proved to be sufficient to increase and reduce, respectively, the expression of this gene. Additional functional studies are now underway to understand the *ZEB2* molecular mechanisms linked to the 2q22 region that may contribute to a better understanding of RCC.

Cancer Posters - Wednesday

PB1039*. Chromatin accessibility of primary human cancers ties regional mutational processes and signatures with tissues of origin

Authors:

J. Reimand, O. Ocsenas; Ontario Inst Cancer Res., Toronto, ON, Canada

Abstract Body:

Cancer genomes are shaped by complex and spatially variable mutational processes whose functional and genetic determinants remain poorly understood. At the megabase scale, somatic mutation burden correlates with DNA replication timing and chromatin accessibility. However, these observations are based on the epigenomes of common cell lines while the contribution of cancer epigenomes remains uncharacterised. Here we model megabase-scale mutation frequencies in thousands of whole cancer genomes using a compendium of hundreds of genome-wide epigenetic and replication-timing profiles that represent primary human cancers, normal tissues, and cell lines. Using a machine learning framework, we show that the chromatin profiles of primary cancers, rather than the profiles of normal cells, are the major determinants of regional mutation burden in most cancer types. Chromatin profiles of matching disease and cell types are the strongest predictors of these mutational processes based on feature analysis of our models, indicating tissues of origin, cancer heterogeneity, and exposures to mutagenic agents. Mutational signatures also associate with chromatin accessibility: signatures of carcinogen exposure and unknown origin are best predicted by chromatin profiles while the associations with endogenous signatures appear weaker. Further, our models point out specific genomic regions where the highly enriched mutations remain unexplained by the diverse epigenetic profiles. These regions include known driver mutations, suggest novel intergenic regions with putative non-coding drivers, and indicate an over-representation of developmental and lineage-specific genes and pathways. This analysis of mutational processes suggests that most chromatin-associated somatic passenger mutations in cancer genomes follow the epigenetic landscapes of cancer cells, indicating that cancer epigenomes capture better the inter-tumor heterogeneity of mutational processes and suggest that many somatic mutations in cancer genomes occur after normal cells have transformed to cancer.

Cancer Posters - Thursday

PB1040. Chromosomal alteration profiling of lymph node metastases in non-small cell lung cancer

Authors:

J. Valdebenito^{1,2}, J. Wong², Y. Jakubek³, D. Chang², T. McDowell², G. A. Eapen^{4,1}, H. Kadara⁵, P. Scheet^{2,1}; ¹The Univ. of Texas MD Anderson Cancer Ctr. UTHlth.Graduate Sch. of BioMed. Sci., Houston, TX, ²Dept. of Epidemiology, The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, ³Coll. of Med., Dept. of Internal Med. at the Univ. of Kentucky, Lexington, KY, ⁴Dept. of Pulmonary Med., Div. of Internal Med., The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, ⁵Dept. of Translational Molecular Pathology, Div. of Pathology/Lab Med., The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract Body:

Goal: We sought to assess mosaic chromosomal alterations (mCAs) as a marker in lymph node involvement of a prospective cohort of lung cancer patients. **Background:** Lung cancer is the leading cause of cancer death worldwide, often presenting with metastatic disease. Lymph nodes are thought to be a gateway to distant metastasis and offer an opportunity to study the genomics of this transitional state. Mosaic chromosomal alterations (mCAs) are megabase-scale structural variants, such as deletions or duplications, known to drive the clonal expansion of mutated clones. Their presence (aneuploidy) is considered a hallmark of cancer and thus a potential biomarker for diagnosis or prognosis. **Methods:** We processed paired primary tumors and lymph node samples using a biospecimen collection from non-small cell lung cancer (NSCLC) patients undergoing endobronchial ultrasound transbronchial needle aspiration (EBUS/TBNA). EBUS and matched germline blood samples were analyzed using SNP array intensity data. To detect mCAs, we applied hapLOH, a sensitive haplotype-based method to detect copy number events. **Results:** We discovered 322 autosomal mCAs in 80 participants, most concentrated between ages 60-80 and in late TNM stages. Testing specific genomic regions where other mCAs were observed in the same patient demonstrated an increase of 58% mCAs in pathologically negative lymph nodes. In addition, smoking was significantly associated with the metastatic status ($p=0.054$) and recurrence of the disease ($p=0.012$). **Conclusions:** Preliminary analyses suggest chromosomal alterations may precede phenotypic and histological changes at the tissue level and are consistent with tobacco use as a leading risk factor. Not surprisingly, participants with a higher number of mCAs are older and at advanced stages of the disease. Studying the recurrence will require the maturation of this clinical cohort, but the discovery of mCA calls in both lymph node-negative and positive samples might suggest regions of interest for potential recurrence of the disease. In summary, our study helps demonstrate the utility of mCAs as a potential marker for NSCLC diagnosis and prognosis.

Cancer Posters - Wednesday

PB1041. Chromosome Arm Copy Number Profiling using a Clinical, Targeted NGS Panel

Authors:

J. Gascoyne, A. Olvera-Morales, J. Fowler, H. Alvarez, A. San Lucas, P. Scheet; MD Anderson Cancer Ctr., Houston, TX

Abstract Body:

This study assesses the performance of calling chromosomal arm-level copy number variation (caCNV) using “MDA MAPP”, a clinical NGS panel for cancer patients implemented at MD Anderson Cancer Center targeting 610 genes and 2MB of the genome at 1000x average read-depth. MDA MAPP profiles solid tumor FFPE samples with paired PBMC normals, providing characterization of mutations, gene-level somatic CNVs, and structural variations in addition to calling microsatellite-instability and estimating tumor mutation burden. However, this panel does not currently provide caCNV biomarkers which have clinical actionability in solid tumors. In this study, 20 pairs of solid tumor FFPE tissues and PBMC normals are profiled using both MDA MAPP and whole-exome sequencing, the latter of which is used to infer a caCNV “gold standard”. It has been widely shown that caCNV calling is accurate using WGS and exome sequencing, but the feasibility of calling caCNV on smaller targeted panels has not been demonstrated. In our approach, caCNV calling is implemented as a heuristic algorithm that incorporates both sequencing read-depth and haplotype-informed allelic imbalance, leveraging off-target reads. We use CNVkit to call segmental CNVs and perform post-processing to estimate an initial chromosome copy number. To attempt to further discriminate noise from true amplifications and deletions, we then assess evidence of chromosome-level allelic imbalance using hapLOHseq, an existing haplotype-based method to detect allelic imbalance, for caCNV near average CN thresholds at borderline levels (2.45 - 2.5 for amplifications and 1.5 - 1.55 for deletions). At amplification and deletion copy numbers of 2.5 and 1.5 respectively, a sensitivity of 0.917, a specificity of 0.977, a PPV of 0.846, and a NPV of 0.989 were observed. Although our results are promising, we have identified areas in which our approach can be improved: due to optimization of the platform for specific oncogenes, individual chromosome arms do not exhibit equal performance and thus performance characteristics may improve if we apply chromosome-specific thresholding; moreover, another limitation of this analysis is that copy-neutral LOH is not considered in the chromosomal event assessments and therefore inflate our calculated false positive rate. Overall, our analysis demonstrates the feasibility of using targeted panels for caCNV profiling. For future clinical testing using MDA MAPP and other targeted NGS panels, we plan to evaluate chromosome calling on specific arms of importance (such as 1p19q co-deletion in gliomas), and potentially augment the panel with strategic targets on the arms of interest.

Cancer Posters - Thursday

PB1042. Circulating miRNA signature for diagnostic prediction in breast cancer.

Authors:

S. Yerukala Sathipati¹, M-J. Tsai^{2,3}, S. Shukla¹, S-Y. Ho^{4,5,6}; ¹Marshfield Clinic Res. Inst., Marshfield, WI, ²Hinda and Arthur Marcus Inst. for Aging Res. at Hebrew Senior Life, Boston, MA, USA, Boston, MA, ³Dept. of Med., Beth Israel Deaconess Med. Ctr. and Harvard Med. Sch., Boston, MA, USA, Boston, MA, ⁴Inst. of Bioinformatics and systems biology, Natl. Yang Ming Chiao Tung Univ., Hsinchu, Taiwan, ⁵Dept. of Biological Sci. and Technology, Natl. Yang Ming Chiao Tung Univ., Hsinchu, Taiwan, ⁶Ctr. for Intelligent Drug Systems and Smart Bio-devices (IDS2B), Natl. Yang Ming Chiao Tung Univ., Hsinchu, Taiwan

Abstract Body:

Breast cancer (BC) is one of the most commonly diagnosed cancers worldwide. A better understanding of the molecular markers that can detect breast tumors at early stages may lead to improved survival and development of new therapeutic strategies. As key regulatory molecules in several biological processes, microRNAs (miRNAs) are potential biomarkers for cancer. The aim of this study was to develop a machine learning based evolutionary learning method called BSig to identify a BC specific circulating miRNA signature for diagnostic prediction in patients with BC. BSig established a compact set of miRNAs as potential markers from 1,280 patients with BC and 2,686 healthy samples retrieved from gene expression omnibus database (GSE73002) for the diagnostic prediction. BSig achieved a prediction performance: training accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve of 100%, 1.00, 1.00, and 1.0, respectively; and an independent test accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve of 99.90%, 1.00, 0.99, and 0.99, respectively. BSig identified 13 miRNAs, hsa-miR-658, hsa-miR-3648, hsa-miR-3656, hsa-miR-4476, hsa-miR-642a-5p, hsa-miR-5100, hsa-miR-5698, hsa-miR-6088, hsa-miR-6768-5p, hsa-miR-6800-5p, hsa-miR-6803-5p, hsa-miR-6807-5p, and hsa-miR-6836-3p, contributed significantly towards diagnostic prediction in BC. Bioinformatics analysis of the miRNA signature revealed their involvement in several biological and signaling pathways, including ErbB signaling pathway (hsa04012) that control proliferation, angiogenesis, metastasis and survival of breast cancer. The miRNA signature involved in some of significant Kyoto Encyclopedia of Genes and Genomes pathways, including prion diseases (hsa05020), renal cell carcinoma (hsa05211), glioma (hsa05214), and thyroid hormone signaling pathway (hsa04919). The miRNA signature significantly involved in top five gene ontologies annotations are organelle (GO:0043226), ion binding (GO:0043167), cellular nitrogen compound metabolic process (GO:0034641), biosynthesis process (GO:0009058) and nucleic acid binding transcription factor activity (GO:0001071). Furthermore, this circulating miRNA signature is associated with several major cancer/diseases and some miRNAs of the signature showed significant expression difference between BC and healthy groups. BSig may serve as the basis for cancer screening and therapeutic selection in BC.

Cancer Posters - Wednesday

PB1043*. *Cis*- and *trans*-eQTL TWAS of breast and ovarian cancer identify more than 100 risk genes in the BCAC and OCAC consortia.

Authors:

T. Head¹, A. Todor¹, J. Yang¹, J. Plummer², S. Gayther², S. Kar³, J. Schildkraut¹, M. P. Epstein¹; ¹Emory Univ., Atlanta, GA, ²Cedars Sinai Med. Ctr., Los Angeles, CA, ³Univ. of Bristol, Bristol, United Kingdom

Abstract Body:

Genome-wide association studies (GWAS) have identified numerous common risk variants for breast and ovarian cancer, with most GWAS-derived risk variants lying in non-coding regions of the genome. This has spurred considerable interest in transcriptome-wide association studies (TWAS) that aim to investigate how genetically regulated transcriptional activity plays a role in both the shared and distinct etiologies of these two complex diseases. However, the TWAS methods utilized to date have only considered the regulatory effects of variants located close to the target gene (*cis*-SNPs). With growing evidence for non-trivial distal regulatory effects of common variants (*trans*-SNPs) on gene expression, we performed TWAS of breast and ovarian cancer using a recently proposed Bayesian genome-wide method (BGW-TWAS) that incorporates both *cis*- and *trans*-expression quantitative trait loci (eQTL). We used whole genome sequencing and RNA sequencing data in breast and ovarian tissue from the Genotype-Tissue Expression Project V8 to train genome-wide gene expression imputation models with BGW-TWAS. We then used these imputation models to perform a TWAS on large-scale GWAS summary statistic data from the Breast Cancer and Ovarian Cancer Association Consortia (BCAC and OCAC) to identify genes whose genetically predicted expression levels are associated with risk of breast cancer overall, five breast cancer subtypes (luminal A-like, luminal-B like, luminal B/HER2-negative-like, HER2-enriched-like, triple-negative), non-mucinous epithelial ovarian cancer, and five ovarian cancer histotypes (high grade serous, low grade serous, mucinous, endometrioid, clear cell). We identified 104 unique significant genes after Bonferroni correction across breast cancer phenotypes and 9 across ovarian cancer phenotypes. These loci include both established GWAS risk genes such as *ANKLE1* at the 19p13 breast/ovarian cancer risk locus and several novel loci such as *ACAP3* (overall breast cancer risk) whose associations are predominantly driven by *trans*-eQTL effects. These *trans*-driven genes were not identified by a competing TWAS approach that considered *cis*-eQTL only. Results provide further insight into the complex genetic architecture underlying risk of breast and ovarian cancer subtypes in individuals of European ancestry. Efforts to validate expression prediction models of top genes using external data from the Cancer Genome Atlas are ongoing.

Cancer Posters - Thursday

PB1044. Classification of Central Nervous System Tumors with DNA methylation - The Brazilian Experience

Authors:

B. Wolff¹, J. Castro², Y. Gasparini¹, V. Almeida¹, G. Carvalho¹, L. Vieira¹, A. Nascimento¹, F. Costa¹, L. Kulikowski³; ¹Faculdade de Med. da Univ.e de São Paulo, São Paulo, Brazil, ²AC Camargo Câncer Ctr., São Paulo, Brazil, ³Univ.e de Sao Paulo, São Paulo, Brazil

Abstract Body:

The vast range of Central Nervous System (CNS) tumor entities makes the standardization of the histopathological classification of CNS tumors a challenge, leading to misleading decision in clinical practice, as well as the interpretation and validity of the results of clinical trials. The World Health Organization (WHO) establishes the criteria for classifying CNS tumors and recommends an integrated diagnosis, using clinical, morphological and molecular aspects, and the methylation profile is already recommended for the diagnosis of some tumor entities. **Objective:** We report a proof of concept of the usability and implantation of the diagnosis of CNS tumors through DNA methylation in the Brazilian diagnostic routine. **Methodology:** We analyzed seven samples of CNS tumors from paraffin material provided by the archive of the Department of Pathological Anatomy of the AC Camargo Cancer Center. We performed methylation assay using the Infinium HumanMethylation450 BeadChip (Illumina, Inc., San Diego, CA). The data obtained was submitted to the MolecularNeuropathology.org platform for classification based on methylation and Copy Number Variation (CNV) graph. The analysis was performed as proposed by Capper et al. (2018) evaluating a combination of tumor entities and classes. **Results:** The platform showed some relevant findings: (a) C1 sample as *High grade hemispheric glioma, H3 K27M-mutant*, and the classifier indicated a match with the class *Diffuse midline glioma, histone 3 K27-mutant* with a calibrated score of 0.99 and the MGMT gene promoter status as unmethylated; (b) C3 sample was classified as *anaplastic pleomorphic xanthoastrocytoma* with a calibrated score of 0.91 and promoter status of the gene as unmethylated, in agreement with the previous diagnosis; (c) C4 and C6 samples initially classified as *Astroblastoma, MNI-altered* and *Diffuse hemispheric glioma, H3 G34-mutant* were corroborated by the classifier with a calibrated score of 0.95 and 0.99 respectively. Sample C4 does not have the methylated MGMT gene promoter, while sample C6 does. Samples C2, C5 and C7 showed no match. **Conclusion:** Our experience applying methylation array and using the Heidelberg classification platform presented effective results to improve the analysis of CNS tumor entities and facilitate the diagnosis and treatment for patients.

Cancer Posters - Wednesday

PB1045. Clinical and genetic features of patients with multiple endocrine neoplasia type 1 in Korea

Authors:

B. Kim, J-S. Lee, M. KIM, M-W. Seong; Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract Body:

Introduction Multiple endocrine neoplasia type 1 (MEN1) is a hereditary syndrome characterized by the predisposition to endocrine tumors. It shows an autosomal dominant inheritance pattern and is caused by inactivating variants of the *MEN1* gene. Here, we analyzed the clinical and genetic features of MEN1 patients in Korea. **Methods** We enrolled patients who were diagnosed as MEN1 in Seoul National University Hospital between 2012 and 2022. We retrospectively reviewed the clinical and genetic characteristics of the patients from medical records. We adopted the recent proposition of TENGAN for the interpretation of MEN1 variants. **Results** Totally, 39 patients were diagnosed as MEN1 and primary hyperparathyroidism (69%) is the most common presentation. Twenty-six *MEN1* variants were identified and distributed along the *MEN1* gene. Twelve frameshift, 7 nonsense, 3 missense, 2 splice sites, 1 large deletion, and 1 intronic variants were detected. **Conclusion** This study broadens the understanding of clinical and genetic characteristics of MEN1 patients.

Cancer Posters - Thursday

PB1046. Clinical utility of genomic sequencing for hereditary cancer syndromes: A retrospective chart review

Authors:

R. Kodida¹, S. Shickh², C. Mighton², M. Clausen¹, J. Sam¹, D. Hirjikaka¹, E. Reble¹, E. Adi-Wauran², S. Krishnapillai², K. Schrader³, N. Baxter⁴, A. Laupacis², J. Lerner-Ellis⁵, R. Kim⁶, Y. Bombard²; ¹Li Ka Shing Knowledge Inst. of St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, ²Univ. of Toronto & Li Ka Shing Knowledge Inst. of St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, ³Univ. of British Columbia, Vancouver, BC, Canada, ⁴Univ. of Melbourne & Univ. of Toronto, Melbourne, Australia, ⁵Mount Sinai Hosp. & Univ. of Toronto, Toronto, ON, Canada, ⁶Univ. Hlth.Network, Mount Sinai Hosp., The Hosp. for Sick Children, Ontario Inst. for Cancer Res., Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Background Identification of patients with hereditary cancer syndromes (HCS) is important to detect families who can benefit from preventive surgeries and surveillance that reduce morbidity and mortality. However, most patients with query HCS who undergo standard panel tests receive uninformative results and are unable to access risk-reducing strategies. Genomic sequencing (GS) may reduce the diagnostic gap in this population. However, most studies evaluating GS utility have used small cohorts and phenotype-driven panels, and only reported yield of pathogenic/likely pathogenic (P/LP) variants. **Aim** Describe the yield of all types of cancer results for patients who received GS for HCS from an RCT. **Methods** An observational, retrospective chart review was conducted for cancer patients with previous uninformative panel results, who received GS from an RCT. Charts were reviewed by 5 team members to extract demographics, clinical history, genomic results and recommendations made by study clinicians and ordering clinicians. Descriptive statistics were used to describe demographics and clinical history. Proportions were calculated to compare frequencies of patients with different types of cancer results and recommendations made for cancer results. **Results** 276 patients were eligible and included. Participants were mostly female (240), of European descent (158) with breast cancer history (240). Twenty-five patients (9.1%) received ≥ 1 P/LP variant in a cancer gene, 246 patients (89%) had ≥ 1 VUS in a cancer gene and 27 (10%) were negative. Most P/LP variants (20/26) were in low/moderate risk genes and half (13/26) were considered secondary cancer findings. LP/P variants led to 94 recommendations in 21/25 patients. The most common recommendations were for genetic counseling (21), communication to relatives (21), and cascade testing (21), followed by clinical evaluation (9) and validation (7). Patients with preliminary evidence results (genes with limited literature and no management guidelines) did not receive recommendations for clinical evaluation from study clinicians; 2 did receive clinical evaluations at the request of the ordering clinician. 663 cancer VUS were identified: 253 were primary and the rest secondary. The mean number of VUS was 2.7/patient and was higher in patients of Non-European ancestry versus European (3.5 vs. 2.5, $p < 0.05$). **Conclusion** Although we identified modest clinical utility of GS for HCS, the variability in management highlights a need for guidelines to standardize management of patients with low/moderate risk results. To optimize use of GS, future research is needed to evaluate the clinical utility of GS as a first-tier test.

Cancer Posters - Wednesday

PB1047*. Clinicogenomic Neural Networks with Real World Evidence: Predicting Outcomes and Targeting Interventions at a Comprehensive Cancer Center.

Authors:

M. Flagg¹, C. Ramsey¹, K. Fisch², J. Califano¹; ¹Univ. of California San Diego, La Jolla, CA, ²UC San Diego, La Jolla, CA

Abstract Body:

Recent advancements in extraction of real-world data (RWD) from electronic health records (EHR) offer unprecedented precision in analysis of oncologic clinical events. Through bulk RWD extraction from a clinical data warehouse such as UCSD's Clinical Data Warehouse for Research (CDWR), the full operational records of oncologic care at a comprehensive cancer center can serve as inputs to machine learning (ML) models. However, genomic data is not routinely contained in the EHR; as such, we draw on the genomic data in the Registry for Molecular Oncology Research (ARMOR), all of which originates from genetic testing ordered in the course of ongoing care. Combining CDWR and ARMOR enables the creation of an integrated clinicogenomic dataset derived entirely from RWD to drive prediction of mortality, hospitalization, and other key outcomes. To this end, we developed a pipeline to extract RWD for a large population of patients receiving care at a comprehensive cancer center (n = 28,016), integrate the data with genomic results from ARMOR when available (n = 5,100), and train multiple off-the-shelf ML models such as random forest and conventional neural network architectures. The input features include a broad range of clinical data such as demographics, encounters, medications, tumor staging, referrals, and more. The feature set derivation highlights clinically meaningful values such as abnormal test results. We capture fluctuations in the density of data reporting frequency, representing temporal distinctions in patterns of care. ARMOR data on polymorphisms, mutational burden, copy number alterations, and fusion genes are incorporated, enabling association of genomic features with these clinical phenotypes. We iterate this pipeline across multiple cancer types, presenting the challenge of biases such as cancer type and differences in feature availability. To account for this issue, we employ a Few-Shot transfer learning approach derived from Ma et al.'s Transfer of Cellular Response Prediction (TCRP) to transfer predictions between cancer types. By combining clinicogenomic data with a novel ML approach, we produce robust predictions of adverse oncologic outcomes. On trials with a test population of 268 patients with 9739 clinicogenomic features, we effectively predicted inpatient admissions with an AUC of 0.82 with random forest and accuracy of 0.836 with a dense neural network. These can be immediately operationalized to guide population health interventions across our health system, along with presenting a scalable framework for extension of this approach to other health systems with a CDW.

Cancer Posters - Thursday

PB1048. CML with a typical BCR-ABL1 fusion in a 15-year-old male patient presenting with marked leukocytosis, thrombocytosis, left-shifted granulocyte maturation in the peripheral blood

Authors:

M. Solanki^{1,2}, A. Cheng^{1,2}, T. Basso^{1,2}, N. Madhu^{1,2}, K. Eastwood³, M. Guardiola³, C. A. Tirado^{3,2}; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²The Intl. Circle of Genetics Studies, Los Angeles, CA, ³Baylor Scott and White Hlth., Dept. of Pathology, Temple, TX

Abstract Body:

Chronic Myelogenous Leukemia (CML) is a slow-progressing bone marrow cancer that is typically characterized by acquired BCR-ABL gene mutations (caused by the presence of the Philadelphia chromosome) in immature myeloid cells. CML affects approximately 1 in 526 people, and makes up about 15% of all leukemia cases. The primary risk factors of CML are radiation exposure, old age, and gender, specifically in correlation to males. Here, we present a case of a 15-year-old male patient diagnosed with CML who was further assayed using conventional cytogenetic analysis, DNA fluorescence in situ hybridization (DNA FISH) analysis, and flow cytometric analysis to determine correlation with other diagnostic testing alongside being BCR/ABL1-positive. Conventional cytogenetic analysis on bone marrow aspirates using G-banding techniques following short-term cell culture without mitogens revealed the presence of an abnormal clone exhibiting a t(9;22)(q34;q11.2), which is a chromosome 9 and 22 translocation consistent with a Philadelphia chromosome positive form of CML. DNA FISH analysis was conducted on bone marrow samples using the LSI BCR/ABL Dual Color/Dual Fusion probe, and the findings indicated an abnormal chromosomal translocation between 9q34 (ABL1) and 22q11.2 (BCR) with a result description of nuc ish(ABL1,BCR)x3(ABL1 con BCRx2)[182/200]. Peripheral blood smear analysis found indications of normochromic, normocytic anemia, leukocytosis with left-shifted granulocytes, increased basophils, and prominent eosinophils along with marked thrombocytosis, which is consistent with patient's recent diagnosis of CML. Flow Cytometry revealed a non-diagnostic immunophenotype with: 1) Small population of T and B lymphocytes with T lymphoid predominance, 2) Predominant maturing granulocyte population, and 3) No phenotypic evidence of significant myeloid immaturity. Due to the chronic phase presentation of CML in this patient, it can be concluded that the patient has a favorable prognosis and could have promising remission results from standard CML therapies. Clinical correlation was indicated.

Cancer Posters - Wednesday

PB1049*. Co-delineating genomic and transcriptomic modes of resistance to MEK inhibition in individual triple negative breast cancer cells.

Authors:

D. Arvapalli¹, I. Gonzalez¹, T. V. Morozova¹, D. R. Goulet², G. L. Johnson², V. J. Weigman¹, J. A. A. West¹, G. L. Harton¹, J. Zawistowski¹; ¹BioSkryb Genomics, Inc., Durham, NC, ²Univ. of North Carolina Sch. of Med., Chapel Hill, NC

Abstract Body:

Triple negative breast cancer (TNBC) tumors are frequently driven by MAPK signaling and are initially susceptible to MEK inhibition, yet resistance invariably develops. This resistance occurs, in part, by transcriptional adaptation where signaling is bypassed to reactivate MAPK signaling through a different node, yet resistance can also develop or co-develop through genomic modification. We therefore exploited single-cell ResolveOME chemistry to simultaneously capture whole genome single nucleotide variation (SNV), copy number variation (CNV) and full-transcript RNAseq from the same individual cell to define multifaceted contributions to drug resistance and to probe single-cell heterogeneity of these contributions for lineage identification. We employed two TNBC models of resistance to the MEK inhibitor trametinib whereby transcriptional and epigenetic contributions to the resistance have been extensively characterized, yet in which comprehensive assessment of genomic contributions to resistance has not yet been performed. In an epithelial subpopulation of SUM-229PE cells, previous work demonstrated resistance to trametinib coincided with upregulated *KRAS* expression relative to treatment-naive parental cells and manifested as one of the most statistically significant differentially-expressed genes. The genomic arm of ResolveOME and BaseJumper analysis software unveiled a genomic lesion, a ~650 kilobase block of differential allelic identity between parental and trametinib-resistant cells encompassing the *KRAS* gene—indicating structural variation at the locus and a candidate genomic mechanism underlying *KRAS* transcript upregulation and consequential MEK inhibitor resistance. The transcriptomic arm of the ResolveOME workflow concurrently validated the upregulation of *KRAS* and previously known gene sets while defining single-cell heterogeneity within. We also subjected parental and trametinib-resistant SUM-159PT cells, a mesenchymal model of the claudin-low molecular subtype of TNBC, to ResolveOME profiling. We observed expression modulation of CEBP family of enhancer factors and SMARCB1 family chromatin remodeling factors, and current efforts are focused on linking resistant cell-specific SNVs proximal to these and other upregulated factors in both models as candidate regulatory variants in the paradigm of epithelial mesenchymal transition (EMT). The unification of genomic and transcriptomic data here uncovered DNA variation underlying the observed transcriptional modulation and provides a framework for ultimately seeking targetable variation in longitudinal primary patient samples.

Cancer Posters - Thursday

PB1050. Combined FISH and cytogenetic testing increases the accuracy of classification and risk stratification of multiple myeloma samples.

Authors:

J. Smith^{1,2}, W. Bi^{1,2}, N. Owen^{1,2}; ¹Baylor Coll. of Med., Houston, TX, ²Baylor Genetics Lab., Houston, TX

Abstract Body:

Laboratories face cost cutting mandates as test costs rise and reimbursements fall. At the same time, discovery of new cytogenetic abnormalities in hematologic diseases with diagnostic/prognostic significance requires new FISH probe validations and more intense scrutiny of chromosome preparations, both of which come with additional expense for the laboratory. Multiple myeloma (MM) was once considered one disease entity. Now, analysis of plasma cells extracted from bone marrow enables classification of MM into distinct subtypes based on cytogenetic anomalies. Knowledge of these subtypes has led to refined risk stratification and continues to lead to better therapeutics. Significant increase in diagnostic yield on plasma cell FISH and difficulties in culturing plasma cells have prompted many laboratories to offer FISH only testing for MM. We present cases showing the value of cytogenetics in correctly classifying MM in concert with plasma cell FISH. FISH from patient 1 detected gain of 1q, trisomies 7, 9, 15 and 17, and an IGH rearrangement. Without chromosomes, it cannot be determined if this represents a rare case of two co-occurring primary abnormalities or hyperdiploidy with a secondary IGH rearrangement. Complex karyotypes from an IL-4 stimulated culture showed a co-occurrence of two primary abnormalities: hyperdiploidy and a t(6;14) IGH-CCND3 rearrangement. CCND3, unlike many other primary IGH rearrangements, does not confer poor prognosis. Patient 2 showed gains of 1q, loss of 1p, trisomies 7, 9 and 15, monosomy 13, and loss of p53, by deletion and monosomy. FISH data were suggestive of hyperdiploid MM with secondary high-risk abnormalities. Chromosome analysis revealed a highly complex pseudodiploid karyotype containing an 8;22 MYC-IGL translocation and a jumping 1q translocation. Abnormal cells were detected in both unstimulated and IL4 stimulated cultures indicating a highly proliferative clone. Both jumping 1q and MYC-IGL are high risk with possible treatment implications. The clone was also found to not be hyperdiploid as the FISH suggested. FISH from patient 3 indicated hyperdiploid disease with an IGH rearrangement and gain of 1q. Similar to the first case, FISH cannot distinguish between primary and secondary IGH translocations in the setting of hyperdiploidy. Cytogenetics confirmed hyperdiploidy and identified a secondary t(8;14) MYC-IGH translocation in this case; MYC rearrangements in hyperdiploid MM are associated with high risk disease. Chromosomes on MM patients allow for concurrent genome-wide assessment of abnormalities that provide additional essential information beyond that obtained by FISH alone.

Cancer Posters - Wednesday

PB1051. Comprehensive Characterization of Functional Cancer Susceptibility at the 5p15.33
TERT/CLPTM1L Pan-Cancer Risk Locus

Authors:

A. O'Brien^{1,2}, H. Kong¹, K. E. Connelly¹, M. Xu¹, I. Collins¹, H. Patel¹, B. Papenburg¹, H. Sowards¹, K. Funderburk¹, J. Hoskins¹, J. Zhong¹, M. B. Patel¹, M. Ho¹, International Lung Cancer Consortium (ILCCO), Bladder Cancer Consortium, M. Iles³, M. H. Law⁴, M. T. Landi⁵, Melanoma Meta-Analysis Consortium, R. Stolzenberg-Solomon⁵, B. Wolpin⁶, The Pancreatic Cancer Cohort Consortium (PanScan), A. P. Klein⁷, D. Li⁸, G. M. Petersen⁹, The Pancreatic Cancer Case-Control Consortium (PanC4), N. Orr², S. Chanock⁵, S. Lindström¹⁰, J. Shi⁵, L. Prokunina-Olsson¹, J. Choi¹, K. Brown¹, L. Amundadottir¹; ¹Lab. of Translational Genomics, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, ²Queen's Univ. Belfast, Belfast, United Kingdom, ³Leeds Inst. for Data Analytics, Sch. of Med., Univ. of Leeds, Leeds, United Kingdom, ⁴QIMR Berghofer Med. Res. Inst., Brisbane, Australia, ⁵Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, ⁶Dana Farber Harvard Cancer Ctr., Boston, MA, ⁷Johns Hopkins Sch. of Med., Baltimore, MD, ⁸Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, ⁹Dept. of Hlth.Sci. Res., Mayo Clinic Coll. of Med., Rochester, MN, ¹⁰Dept. of Epidemiology, Univ. of Washington, Seattle, WA

Abstract Body:

Genome wide association studies (GWAS) have identified ten distinct associations at 5p15.33 in as many cancer types, all with notable pleiotropy. This region harbors two plausible target genes: *TERT* and *CLPTM1L*. To comprehensively evaluate known candidate causal variants (CCV) at 5p15.33, assign target genes and characterize functional pleiotropy between cancer types, we are conducting CRISPRi/a screening in eight cell lines from four types (pancreatic cancer, melanoma, bladder cancer and lung cancer) whose proliferation has been shown to be dependent on both *TERT* and *CLPTM1L* expression. Massively parallel reporter assay (MPRA) of fine mapped variants in the same cell lines will accompany CRISPR screening. Furthermore, we have used PacBio long read sequencing to document two novel 5p15.33 CCVs which may explain one of the risk associations. 7244 sgRNAs were tiled across chr5:1,233,287-1,365,002 (hg19), with each of the 178 CCVs identified via Bayesian and LD-based fine mapping targeted by at least three sgRNAs within a ± 100 bp window. CRISPRi screening of the MIA PaCa-2 pancreatic cancer cell line revealed significant sgRNA depletion (FDR <0.01) within ± 100 bp of 71 fine mapped SNPs. 17 of these variants also exhibit significant (FDR <0.01) allele-specific cis-regulation in MPRA in the same cells, suggesting these as strong CCVs. Forthcoming CRISPR and MPRA data for seven additional cell lines aims to characterize the pleiotropy observed in the region.

PacBio sequencing has been used to identify two CCVs which could explain one of the ten 5p15.33 signals. Two highly polymorphic variable number tandem repeats intronic to *CLPTM1L* (denoted here VNTR1 & VNTR2) show segregation of longer VNTR copy number on the risk haplotype and are inherited in a Mendelian pattern in HapMap trios. Both VNTRs are enriched for biologically relevant transcription factor binding motifs and histone marks in relevant cell lines. Resequencing of 944 pancreatic cancer cases and 878 controls revealed that single repeat unit increases in VNTR length are a significant predictor of case/control status (VNTR1; $P = 1.20 \times 10^{-6}$, VNTR2; $P = 1.99 \times 10^{-7}$) in a logistic regression model (OR of longest diploid repeat length relative to the median: VNTR1: 2.46, VNTR2: 2.39). Imputation of VNTR copy number into GWAS and eQTL datasets to assess whether these variants are associated with disease risk of other cancers, as well as expression of *cis*-genes at this locus is ongoing.

These findings suggest that upon completion, our work will mark a significant progression in the

ASHG 2022 Annual Meeting Poster Abstracts

fundamental understanding of the mechanisms mediating cancer predisposition at the 5p15.33 pan-cancer risk locus.

Cancer Posters - Thursday

PB1052. Comprehensive CRISPR-based dissection of a complex common and familial melanoma risk locus on chromosome band 9p21.

Authors:

R. Hennessey¹, H. Sowards¹, M. Xu¹, R. Thakur¹, J. H. W. Wong², J. Barbour², R. Chari¹, Melanoma Meta-Analysis Consortium, A. Pritchard³, N. Hayward⁴, A. M. Goldstein¹, X. R. Yang¹, M. H. Law⁴, K. Brown¹; ¹Natl. Cancer Inst., Rockville, MD, ²The Univ. of Hong Kong, Pok Fu Lam, Hong Kong, ³Univ. of the Highlands and Islands, Inverness, United Kingdom, ⁴QIMR Berghofer Med. Res. Inst., Brisbane, Australia

Abstract Body:

The 9p21 locus has been established as a common melanoma risk locus through multiple GWAS (lead SNP rs871024, $P=3.32 \times 10^{-65}$, OR=1.18), where the region of association includes multiple independent signals not explained by protein-coding sequence variants. While this locus harbors a well-established, high-risk melanoma gene, *CDKN2A*, the association signal spans several genes which are also frequently deleted somatically in tumors, suggesting multiple genes may play a role in melanoma risk. We used both iterative conditional analysis and Bayesian fine-mapping to comprehensively identify candidate causal variants (CCVs) and identified six signals and 98 CCVs. Independent of GWAS analysis, exome sequencing of high-risk melanoma families identified three rare variants within ten base pairs of each other in the first intron of *CDKN2B* that abrogate CTCF binding. To determine the function of both the common and familial variants, we designed pooled CRISPRi/a gRNA libraries targeting all CCVs, familial variants, and genes at the locus, as well as a larger library with gRNAs tiled approximately every 75bp across the locus. We performed proliferation screening in immortalized melanocytes with and without ultraviolet radiation B (UVB) exposure. Inhibition of control genes with known effects on proliferation (*TP53*, *CCND1*) resulted in strong increases and decreases of proliferation, respectively. Consistent with a well-established role as a tumor suppressor in melanocytes, gRNAs inhibiting *CDKN2A*-p16^{INK4A} were significantly enriched, while *CDKN2B*-targeted gRNAs were moderately enriched. In contrast, gRNAs inhibiting *CDKN2A*-p14^{ARF} and *MTAP* were depleted. CRISPRi targeting of multiple CCVs located within intronic regions of p16 had similar effects to p16-targeted gRNAs, while CCVs within the first intron of p14 had effects in both directions, indicating potential roles for CCVs in the regulation of an established risk gene. In contrast, CCVs from a signal intronic to *MTAP* had opposite effects of p16 targeting, suggesting that inhibition of CCV enhancers from this signal may regulate *MTAP* and or potentially lead to upregulation of p16. Targeting of familial variants had a similar effect as targeting p16 and *CDKN2B* and will require further investigation to tease apart target genes. Finally, we identified multiple distant intergenic regulatory elements that have a strong effect on melanocyte proliferation and viability in both non- and UV-exposed cells, allowing for interpretation of additional non-coding variation found in melanoma cases and families. These results paired with region capture C, ATAC-seq, and HiChIP will help disentangle this complex locus.

Cancer Posters - Wednesday

PB1053. Comprehensive study of gene expression outliers and their regulation mechanisms in pan-cancer.

Authors:

J. Han, P. Boutros; UCLA, Los Angeles, CA

Abstract Body:

Cancer is a disease characterized by remarkable heterogeneity, with many molecular features significantly associated with tumour progression in specific subsets of patients. Gene expression varies drastically between tumours and within cells of a single tumour. This variability in gene expression can be caused by gene-regulatory mechanisms such as DNA copy number changes, extrachromosomal DNA, or aberrant methylation, amongst many others. These mechanisms sometimes generate extreme outliers: transcripts that show atypically high or low gene expression in a small percentage of cancers. These outliers increase the molecular and phenotypic diversity between individuals, contributing to tumour heterogeneity. Gene expression in cancer has been well studied with many reports of differential gene expression patterns between specific cancer types or subtypes. Importantly, many of these differences were strongly associated with tumorigenesis and tumour progression. For example, the *BCR-ABL* fusion was discovered in chronic myeloid leukemia with drugs targeting this fusion gene dramatically increasing the survival rate. The *EML-ALK* fusion gene is another example of a gene-expression outlier and drug target. It occurs in about 4% of non-small-cell lung carcinomas and has been routinely screened for. These examples demonstrate the critical role gene expression outliers play in cancer progression and highlight their potential as biomarkers for diagnostics and identifying novel drug targets. Despite their importance, there has not yet been a comprehensive pan-cancer study of gene expression outliers and their general properties. We lack the answers to fundamental questions such as how many outliers exist in a typical tumour, whether this differs across cancer types, what mechanisms generate the most outliers, whether specific clinical or somatic mutational features correlate with the number or type of outliers, and how many recurrent outliers exist. To answer these questions, we performed a comprehensive and systematic analysis of cancer gene-expression outliers using molecular data from multiple cancer genomics projects. We have created a new statistical outlier detection method and applied it to transcriptomics and proteomics data across 33 cancer types. We used this resource to describe the fundamental landscape of gene-expression outliers, including the most likely genetic and epigenetic mechanisms driving them. The resulting outliers will be further studied for their impact on cancer progression and validated hits will serve as clinically relevant biomarkers and targets for future functional and therapeutic investigations.

Cancer Posters - Thursday

PB1054. Computational method for the detection of microsatellite instability in tumor tissue samples

Authors:

J. Budiš^{1,2,3}, M. Sýkora⁴, T. Sládeček^{1,2}, M. Kucharík^{1,2}, W. Krامل^{1,2}, O. Pös^{1,2}, J. Styk^{1,2,5}, T. Adamík², M. Nemeč², I. Lojova^{6,7}, J. Turňa^{2,3,6}, J. Radvánsky^{2,6,7}, S. Tomáš^{1,2,6}; ¹Geneton Ltd., Bratislava, Slovakia, ²Sci. Park, Comenius Univ. in Bratislava, Bratislava, Slovakia, ³Slovak Ctr. of Scientific and Technical Information, Bratislava, Slovakia, ⁴Faculty of Informatics and Information Technologies, Slovak Univ. of technology in Bratislava, Bratislava, Slovakia, ⁵Inst. of Med. Biology, Genetics and Clinical Genetics, Faculty of Med., Comenius Univ. in Bratislava, Bratislava, Slovakia, ⁶Dept. of Molecular Biology, Faculty of Natural Sci., Comenius Univ. in Bratislava, Bratislava, Slovakia, ⁷Inst. of Clinical and Translational Res., BioMed. Res. Ctr., Slovak Academy of Sci., Bratislava, Slovakia

Abstract Body:

Recent advances in sequencing technologies have enabled affordable testing for various genetic diseases. Their application in the field of oncology has great potential for early detection and monitoring of tumor progression. A promising way is the thorough characterization of microsatellites, where the detection of their unstable forms can be used as a cancer biomarker.

We compared sequenced DNA fragments isolated from tumor and control tissues. We first mapped them to the reference human genome. Then, thousands of selected microsatellite loci across the genome were thoroughly characterized to detect typical anomalies of unstable forms. The changes were aggregated across the analyzed loci to obtain compact features for each sample. Finally, the features were analyzed with classification methods to distinguish between tumor and control tissue.

We show that microsatellite instability is readable in our tumor samples. The method has therefore the potential to automate the detection and characterization of ongoing oncology disease from the sequenced genomic data.

The presented work was supported by the Slovak Research and Development Agency (grant ID APVV-18-0319), and by the OP Integrated Infrastructure within projects with ITMS codes 313011W988, 313011F988, and 313011V578, 313011W428, all co-financed by the European Regional Development Fund.

Cancer Posters - Wednesday

PB1055. Concurrent Core-binding factor beta subunit (CBFB) gene rearrangement/inv(16)(p13.1q22) in p210 BCR-ABL1 positive chronic myelogenous leukemia

Authors:

J. Reid¹, K. Semenova¹, K. Dang¹, J. Tso¹, L. Yang¹, A. Fleishman¹, X. Zhao¹, F. Quintero-Rivera²; ¹Univ. of California Irvine Med. Ctr., Orange, CA, ²Univ. of California Irvine (UCI), Irvine, CA

Abstract Body:

The concurrent identification of core-binding factor beta subunit (CBFB) gene rearrangement associated with inv(16)(p13.1q22) in a patient with chronic myelogenous (CML) is an extremely rare phenomenon. Herein, we present a 60-year-old male who was transferred to the emergency department; he was confused and found to have left sided deficits. His CBC showed hyperleukocytosis with WBC of 600,000/MCL and hemoglobin of 7.3g/dL. His peripheral blood smear showed severe normocytic anemia, severe leukocytosis/neutrophilia with basophilia/eosinophilia and increased myeloid blasts (3%) and mild thrombocytopenia. Bone marrow biopsy identified a markedly hypercellular marrow with marked left-shifted myeloid hyperplasia, megakaryocytic hyperplasia with dysplasia and moderate fibrosis, blasts were not increased (1-2%), consistent with CML-chronic phase (CP). Flow cytometric/immunohistochemistry analysis showed 1.5 % myeloid blasts expressing CD34, CD117, HLA-DR, CD38, CD33, and MPO. Cytogenetics analysis identified a 9 and 22 translocation in 20 cells, with half of them also showing an inversion of chromosome 16, consistent with cytogenetics clonal evolution and accelerated phase -AP- (formerly considered blast phase (BP)). FISH detected a CBFB gene rearrangement in 8.5% of nuclei examined. PCR and NGS detected, BCR ABL p210 transcript, and BCR-ABL1 fusion transcript [VAF % 59.2] and CBFB-MYH11 fusion transcript [VAF % 53.4], respectively. Patient underwent angioplasty, and leukapheresis; hydroxyurea was given for the refractory hyperleukocytosis. He also received cytarabine, dasatinib with decadron. His principal diagnosis was acute basal ganglia hemorrhage, and secondary diagnoses: shock, likely hypovolemic, acute hypoxic respiratory failure, right middle cerebral artery stroke, CML-AP, tumor lysis syndrome, acute kidney injury. Patient passed away 6 days after admission. 25 cases of CML with inv(16) and eosinophilia have been reported in which patients are either in CP, BP or AP. The prognosis has been poor consistently displaying aggressive clinical progression of the disease and rapid decline, if treated only with chemotherapy [PMID: 16203287]. Only two patients treated with FLAG-Ida plus tyrosine kinase inhibitors and allo-SCT had a favorable response [PMID:2825353]. Detection of the inv(16) by karyotyping is technically challenging, in addition, FISH for CBFB is not usually run in the context of CML, thus raising the possibility that similar cases have been underdiagnosed. With the routine use of NGS for gene fusion detection, in addition to cytogenetics, more patients like this one should be diagnosed and treated accordingly.

Cancer Posters - Thursday

PB1056. Confirmation of pathogenic germline variants identified by tumour testing in British Columbia

Authors:

P. Cheng, K. Compton, I. Bosdet, T. Tucker, S. Young, S. Yip, K. Schrader, S. Sun; BC Cancer, Vancouver, BC, Canada

Abstract Body:

Background

Multi-gene panel tumour testing (TT) has been used in the evaluation of various solid and hematologic cancers in British Columbia (BC) since mid-2016, including advanced non-small cell lung cancer (NSCLC), colorectal cancer (CRC), melanoma (MEL), low-grade glioma (LGG), and gastro-intestinal stromal tumour (GIST), ovarian cancer (OVA) added in 2018, as well as acute myeloid leukemia (AML), myelodysplastic syndrome (MDS), and myeloproliferative neoplasms (MPN). In addition to somatic driver mutations, TT may reveal potential pathogenic germline variants (pPGVs) associated with inherited cancer predisposition syndromes. Herein, we analyzed rates of pPGV detection by TT and referral to the Hereditary Cancer Program (HCP) for confirmatory germline testing, and clinical implications for patients (pts) with positive results.

Methods

All pts with TT performed from January 1, 2018 to June 30, 2021 were identified. Next-generation tumour sequencing (Oncopanel) was used to screen the following genes: *AKT1, ALK, BRAF, BRCA1, BRCA2, CCND1, CCND3, CIC, EGFR, ERBB2, ERBB3, FUBP1, HRAS, IDH1, IDH2, KIT, KRAS, MAP2K1, MET, NRAS, PDGFRA, PIK3CA, PTEN, ROS1, SDHA, SDHB, SDHC, SDHD*. Additional hereditary genes were added in 2019: *APC, BMPR1A, CDH1, CDK4, CDKN2A, GNA11, GNAQ, MLH1, MSH2, MSH6, MUTYH, NF1, PALB2, PMS2, POLD1, POLE, SMAD4, STK11, TP53*. Targeted genes in the Myeloid panel included: *ASXL1, BCOR, CALR, CBL, CEBPA, CSF3R, DNMT3A, ETV6, EZH2, FLT3, GATA2, HRAS, IDH1, IDH2, JAK2, KDM6A, KIT, KMT2A, KRAS, MPL, NPM1, NRAS, PHF6, PIGA, PRPF40B, PTEN, PTPN11, RAD21, RUNX1, SETD2, SF1, SF3A1, SF3B1, SH2B3, SMC1A, SMC3, SOCS3, SRSF2, STAG2, TET2, TP53, U2AF1, U2AF2, WTI, ZRSR2*.

Results

In total, 8391 TTs were performed, with pPGVs detected in 273 pts. Of 177 pts referred to the HCP, 130 had germline testing, and 57 PGVs were confirmed in 55 pts (2 pts had 2 PGVs), including 33 that would not have met referral criteria based on personal or family history. Confirmation of PGVs did not impact treatment decisions; however, all pts with PGVs received individualized genetic counselling, and 7 family members subsequently self-referred for testing, with PGVs identified in two.

Discussion

TT detected pPGVs in 3.3% of pts with advanced solid organ and myeloid cancers, with 65% referred to the HCP and germline testing confirming at least one PGV in 42% of cases. Increased clinician and pt awareness with targeted follow up of cases with pPGV's detected by TT could increase referral rates and facilitate identification and screening of at risk pts and families.

Cancer Posters - Wednesday

PB1057. Congenital uterine abnormalities and the risk of gynecological cancers

Authors:

N. Abdelmoula, B. Abdelmoula, A. Karra; Med. Univ. of Sfax, Sfax, Tunisia

Abstract Body:

Background: Women with a history of miscarriage and infertility have higher prevalence of congenital uterine anomalies compared with the general population (diagnosed in 2-4% of women with a normal reproductive outcome). These malformations result from developmental disorder during the formation, fusion or resorption of the Müllerian ducts. Generally, development of the female reproductive tract is a complex and dynamic process that involves a complex regulation via multiple signaling pathways. This study assessed the management of the risk of gynecological cancers in female patients harboring congenital uterine abnormalities among patients who presented at our genetic counselling during the genetic exploration of miscarriages and/or infertility. Methods: This observational study retrospectively obtained data on all patients who were admitted to our genetic counselling during the genetic exploration of miscarriages and/or infertility between January 2002 and November 2016. Out of these patients, we selected female patients for whom congenital uterine anomalies were recorded and genetic counselling was delivered. Results: This study recruited ten patients with congenital uterine abnormalities (3% of the population of infertile patients of our genetic counselling). Out of these patients, a bicornuate unicollis uterus was recorded in 50% of patients, a septate uterus with complete vaginal septum in one case (10%) and a segmental Müllerian agenesis with Mayer-Rokitansky-Küster-Hauser syndrome phenotype in four patients (40%). One of them had in association to unicornuate uterus and vaginal diaphragm at the superior third, a unilateral renal agenesis. Complete agenesis of the uterus was recorded in two patients. At the productive level, miscarriages and recurrent pregnancy losses with and without live birth were recorded in two patients, ectopic pregnancies were recorded in one patient and primary amenorrhea was recorded in two patients. The others presented a primary infertility and were candidate to ART techniques. Cytogenetic analysis of the ten patients showed a 46,XX formula. All of them received a genetic counselling and were informed about a minor risk of gynecologic cancers which requires regular monitoring. Conclusion: In fact, there is an evidence that clear cell adenocarcinomas of the ovary, endometrium, cervix, and vagina are associated with Müllerian duct abnormalities with/without renal malformations. These rare histologically gynecologic cancers and tumors are characterized by upregulation of the mTOR and RAS signaling pathways and demonstrate more likely microsatellite instability than other gynecologic cancers.

Cancer Posters - Thursday

PB1058*. Copy number variation at *ATF7IP* is associated with increased risk for testicular germ cell tumors.

Authors:

M. Fan¹, J. Pluta¹, C. Ruan¹, B. Weathers¹, K. P. D'Andrea¹, R. Karlsson², F. Wiklund², P. A. Kanetsky³, K. L. Nathanson^{1,4}, Testicular Cancer Consortium; ¹Div. of Translational Med. and Human Genetics, Dept. of Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ²Dept. of Med. Epidemiology and Biostatistics, Karolinska Inst., Stockholm, Sweden, ³Dept. of Cancer Epidemiology, Moffitt Ctr. and Res. Inst., Tampa, FL, ⁴Abramson Cancer Ctr., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

BACKGROUND Testicular germ cell tumor (TGCT) is one of the most heritable cancer types. Genome-wide association studies (GWAS) have identified 66 susceptibility loci associated with risk of TGCT, which in total account for 44% of disease heritability. The contribution of copy number variation (CNV) to TGCT susceptibility has not been previously evaluated. **METHODS** SNP-array genotyping was conducted using the Infinium HumanCore-24 chip with 6,920 customized and 306,040 pre-designed probes. We used PennCNV (v1.0.5) to generate raw CNV calls in 5,287 men with and 4,967 men without TGCT, matched by age and ancestry. We used HandyCNV (v1.1.7) to call and annotate CNV regions (CNVRs). Associations were modeled using logistic regression and performed using custom scripts in R v4.0.1. After Bonferroni correction, the threshold for significance was $p < 10^{-05}$. rt-PCR was used for *in vitro* validation. To examine the relationship between copy number and gene expression, we utilized copy number, RNA seq, and CpG methylation data from 116 TGCA TGCT cases. We performed a linear regression analysis to assess the association of expression and CNV. **RESULT** We identified 58,960 CNV calls in 8,945 men; 31,448 calls were found in men with TGCT and 27,512 in men without TGCT. After CNVR calling and case-control comparisons, four duplications, and one deletion met the significance threshold. The most significant ($p=1.17 \times 10^{-27}$) CNV was a segmental duplication overlapping the 5' end of *ATF7IP*, which was validated by rtPCR. The other identified CNVs were a duplication of ~69kbp ($p=1.40 \times 10^{-15}$) overlapping *CADPS2*, *TAS2R16*, and *SLC13A1*; a duplication of ~29kbp ($p=5.53 \times 10^{-12}$) overlapping *NIPSNAP3A*, *NIPSNAP3B*, and *ABCA1*; a deletion of ~2kbp ($p=2.42 \times 10^{-09}$) overlapping *ABC5*; and a duplication of ~150kbp ($p=6.09 \times 10^{-08}$) overlapping *CSN1S1*, *CSN2*, *STATH*, *HTN3*, *HTN1*, and *CSN1S2AP*. Three genes had a statistically significant association between CNV and expression: *ABCA1* ($p=0.009$), *NIPSNAP3A*, and *NIPSNAP3B*. *SLC13A1* and *ATF7IP* also showed increased expression with higher copy number, though the results were not statistically significant. **CONCLUSION** This study is the first to identify segmental duplications/deletions significantly associated with an increased risk for TGCT. Variation at *ATF7IP*, a modulator of transcription regulation and chromatin formation, has been associated with TGCT susceptibility before; the identified promoter CNV is independent of the known association. Many genetic regions showing overlap with significant CNVs contain genes involving solute transport (*ABC5*, *CADPS2*, *SLC13A1*), suggesting a novel biologic pathway important to TGCT susceptibility.

Cancer Posters - Wednesday

PB1059. Cost-utility of universal screening for common BRCA1 and BRCA2 variants among Ashkenazi Jewish women: a real-life analysis

Authors:

R. Michaelson-Cohen^{1,2}, A. Ben Shoham^{2,3}, S. Lieberman^{1,2}, M. Cohen^{2,3}, C. Cohen¹, E. Levy-Lahad^{1,2}, A. Lahad^{2,3}; ¹Shaare Zedek Med. Ctr., Jerusalem, Israel, ²Faculty of Med., Hebrew Univ., Jerusalem, Israel, ³Clalit Hlth.Services, Jerusalem, Israel

Abstract Body:

Objectives: Identifying carriers of pathogenic *BRCA1/BRCA2* variants can reduce cancer morbidity and mortality through surveillance and prevention. Universal testing for founder *BRCA1/BRCA2* variants, found in 2.5% of Ashkenazi Jews (AJ), fulfills WHO disease screening criteria. We analyzed the cost-effectiveness of *BRCA1/BRCA2* population screening (PS) in AJ, as an alternative to two existing strategies: cascade testing (CT) in carrier's relatives (>25% carrier probability) and international family history (IFH)-based guidelines (corresponding to >10% carrier probability). **Methods:** A decision analytic model was performed to estimate quality-adjusted life-years (QALY) gained, and incremental cost-effectiveness ratio (ICER) for PS vs. CT and IFH strategies. Analysis was conducted from a payer-perspective using lifetime horizon, based on actual costs. **Results:** Per 1000 women, the PS model vs. CT strategy predicted 21.6 years gained, a lifetime decrease of 3 breast cancer (BC) cases and 4 ovarian cancer (OC) cases, and the PS model vs. IFH strategy predicted 6.3 years gained, a lifetime decrease of 1 case of BC and 1 case of OC. PS was less costly compared with CT (-3097 US\$/QALY). Although PS was more costly than IFH-based testing (+18,968 US\$/QALY), it was still highly cost-effective, from a public health policy perspective. PS was more effective than other strategies in all sensitivity analyses). **Conclusions:** Our findings, based on actual expenditures from health care systems, indicate that from a public health policy perspective, there is an advantage in expanding the use of genetics for identification of women who are *BRCA1/BRCA2* carriers, and thus at high risk to develop cancer. The CT strategy, the standard protocol in Israel till recently, severely restricts the number of carriers identified, and precludes effective surveillance and prevention in many women who are not identified as carriers. Compared to both CT and IFH policies, the PS strategy is cost-effective, and has the greatest effectiveness in gain in lifespan and reductions in BC and OC incidence. Therefore, our study suggests that founder *BRCA* variant testing should be available to all AJ women, irrespective of family history. Implementation of PS for BRCA variants in these women would be the first example of using Precision Medicine for cancer screening. This can also serve as an informative paradigm as genomics is increasingly integrated into large-scale prevention for additional populations.

Cancer Posters - Thursday

PB1060. *CRLF2* Rearrangements in a Patient with B-ALL.

Authors:

Y. Lin^{1,2}, R. Tang^{1,2}, A. Bajpai^{1,2}, W. Yeh^{1,2}, S. Karamooz³, K. Eastwood³, M. T. Guardiola³, C. A. Tirado^{3,4,2}; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²Intl. Circle of Genetics Studies, Los Angeles, CA, ³Baylor Scott & White Hlth., Temple, TX, ⁴Texas A&M Sch. of Med., Temple, TX

Abstract Body:

Leukemia is the most common pediatric malignancy, and B-cell acute lymphoblastic leukemia (B-ALL) is one of the prevalent pediatric leukemias, accounting for 26% of cancers diagnosed in children 0-14 years of age. We present a case of an 11-year-old girl with B-ALL. Conventional cytogenetics showed a karyotype described as

46,XX,del(5)(q31q35),add(6)(q23),del(7)(q32q36),add(11)(q23),ider(21)(q10)add(21)(q22),inc[20].

DNA FISH analysis was performed on a tissue sample from the bone marrow and showed variant rearrangements of *CRLF2*, as well as loss of *ETV6* signals and gain of *RUNX1* signals. The presence of *CRLF2* rearrangements is often associated with *CRLF2* overexpression and poor prognosis. Complex karyotypes also reflect genomic instability and poor prognosis. The patient was in complete remission nine months after diagnosis but passed away a month later from chemotherapy-induced hepatic failure, renal failure, and febrile neutropenia.

Cancer Posters - Wednesday

PB1061. Cross disorder genetic analysis identifies autoimmune disease loci inversely associated with diverse cancer types

Authors:

J. Chen¹, M. Epstein¹, J. Schildkraut², S. Kar³; ¹Emory Univ., Atlanta, GA, ²Emory Univ., Augusta, GA, ³Univ. of Cambridge, Cambridge, United Kingdom

Abstract Body:

The emergence of immune-checkpoint inhibitors has revolutionized the cancer treatment landscape. It is known that increased autoimmune adverse events correlate with better responses to cancer treatment, suggesting a joint mechanism underlying autoimmunity and cancer. We hypothesized that pleiotropic alleles which confer increased risk of autoimmune diseases, but decreased risk of cancer may offer vital insights into this joint mechanism. We searched for such alleles using some of the largest genome wide association study (GWAS) datasets from European-ancestry populations for four common cancer (breast, ovarian, prostate, and endometrial cancers) and their subtypes and four adult autoimmune/autoinflammatory disorders (Rheumatoid arthritis, systemic lupus erythematosus, ulcerative colitis and Crohn's Disease). We performed a meta-analysis for each cancer-autoimmune disease pair using the inverse-variance approach implemented in METAL and clumped the resulting METAL statistics. In order to obtain METAL statistics for variants that specifically have opposite effects on the two outcomes, we multiplied beta coefficients from autoimmune diseases GWAS summary statistics by '-1' before conducting the meta-analysis. Additionally, we performed a Transcriptome-wide association study using sparse-canonical correlation (sCCA-TWAS), which integrates multiple tissues for increased power, to identify genes associated with both cancer and autoimmune disease. We identified 41, 18, 47 and 26 autoimmune disease loci that are inversely associated with Breast, Ovarian, Prostate, and Endometrial Cancers, respectively, where the lead variant was associated with at least one cancer and autoimmune disease at $P < 10^{-3}$ individually and the combined association in METAL had $P < 5 \times 10^{-8}$. For example, allele A of rs2384061, located in *ADCY3*, was negatively associated with breast cancer ($P = 1.02 \times 10^{-11}$) and positively associated with Crohn's disease ($P = 2.85 \times 10^{-6}$), where METAL $P = 6.45 \times 10^{-16}$. Our sCCA-TWAS results showed that gene expression of *ADCY3* was associated with Breast cancer ($P = 4.08 \times 10^{-10}$) and Crohn's disease ($P = 3.10 \times 10^{-7}$). In the analysis of ER-negative breast cancer and autoimmune diseases, multiple genes that have inflammation roles such as *IRF6*, *IFITM2* and *CCL11* were identified. We confirmed autoimmune effects at many known cancer risk loci and identified several cross-disorder loci that were not detected in previous cancer GWAS. These loci may be valuable genomic targets linking autoimmune disorders with cancer development, offering insights into the interplay between immunity, inflammation and cancer risk, prevention, and treatment.

Cancer Posters - Thursday

PB1062. Deciphering a prognostic DNA methylation signature in tubo-ovarian high-grade serous cancer.

Authors:

B. Jorgensen¹, C. Wang¹, J. Cunningham², S. Armasu¹, N. Traficante^{3,4}, The Australian Ovarian Cancer Study Group, D. Garsed^{3,4}, S. Fereday^{3,4}, D. Ariyaratne³, S. Winham¹, D. Bowtell^{3,4,5}, E. Goode¹; ¹Mayo Clinic, Rochester, MN, ²Mayo Clinic & Fndn, Rochester, MN, ³Peter MacCallum Cancer Ctr., Melbourne, Australia, ⁴Sir Peter MacCallum Dept. of Oncology, The Univ. of Melbourne, Parkville, Australia, ⁵Ctr. for Cancer Res., The Westmead Inst. for Med. Res. and Dept. of Gynaecological Oncology, Westmead Hosp., The Univ. of Sydney, Sydney, Australia

Abstract Body:

Background: We have identified a quantitative tumor DNA methylation signature in tubo-ovarian high-grade serous carcinomas (HGSC) that indicates poor prognosis. The signature associates with shorter time to recurrence, independent of clinical factors (N=325 discovery set, HR 2.87, $p=2.2 \times 10^{-13}$; N=715 validation set, HR 1.65, $p=0.015$) and inversely correlates with expression of immune genes on chromosome 6p21.3. Factors underlying this signature remain unknown. **Methods:** We sought to evaluate additional tumor characteristics related to this signature and/or its role in disease outcome in The Cancer Genome Atlas, Mayo Clinic, and Australian Ovarian Cancer Study participants. We examined the following tumor-based data: tumor RNA-based transcriptomic molecular subtype (N=746), expression of the ATP-binding cassette transporter *TAP1* (N=605), and levels of CD8⁺ TILs (N=222). Multivariate testing adjusting for age, stage, study site, and debulking status utilized linear and Cox regression. **Results:** Higher methylation signature scores were found in tumors with no CD8⁺ TILs or with a low/moderate CD8⁺ TIL level, compared to those with high CD8⁺ TIL levels ($p=5.7 \times 10^{-12}$ and 1.1×10^{-4} , respectively). *TAP1* mRNA levels (6p21.3) negatively correlated with methylation signature (p-values < 0.05). Methylation signature associated with molecular subtype: immunoreactive (C2.IMM) subtype tumors had the lowest signatures and proliferative (C5.PRO) subtype, the highest. Considering age, stage, debulking status, study site, *BRCAl/2* germline mutation status, *TAP1* RNA expression, and molecular subtype, only age at diagnosis, *TAP1* RNA expression and, molecular subtype associated with the signature (p-values 0.017, 4.69×10^{-19} ; 1.24×10^{-9} , respectively). Multivariate survival analyses suggest that stage, debulking status, and methylation signature associated with time to recurrence/death at $p < 0.05$ with adjustment for age at diagnosis, study site, *TAP1* RNA expression, and molecular subtype (methylation signature HR 2.22, $p=4.46 \times 10^{-8}$). Similar results were seen when limiting to participants with CD8⁺ TIL and *BRCAl/2* data, wherein methylation signature remained significantly associated with time to recurrence/death ($p=0.009$), while CD8⁺ TIL levels, *TAP1* RNA expression, and molecular subtype did not. **Conclusion:** These results suggest differential methylation in the tumor-immune microenvironment. As multiple yet independent factors appear to be associated with methylation signature (*TAP1* expression, molecular subtype, CD8⁺ TIL levels), we do not expect methylation signature to be a simple surrogate for any of these tumor features.

Cancer Posters - Thursday

PB1063. Deep and error corrected sequencing via the low-cost Ultima Genomics platform enables ultra-sensitive circulating tumor DNA cancer monitoring

Authors:

A. Cheng¹, A. J. Widman², A. Arora², I. Rusinek³, W. F. Hooper², R. Murray², D. Halmos², T. Langanay², G. Inghirami¹, S. Germer⁴, M. Marton², R. Furatero², J. McClintock², L. Winterkorn², Z. Steinsnyder², M. S. Malbari¹, A. Saxena¹, M. K. Callahan⁵, D. T. Frederick⁶, L. Spain⁷, A. Jaimovich³, D. Lipson³, S. Turajlic⁷, M. C. Zody⁴, N. K. Altorki¹, J. D. Wolchok¹, M. A. Postow⁵, N. Robine⁴, G. Boland¹, D-A. Landau¹; ¹Weill Cornell Med., New York City, NY, ²New York Genome Ctr., New York City, NY, ³Ultima Genomics, Newark, CA, ⁴New York Genome Ctr., New York, NY, ⁵Mem. Sloan Kettering Cancer Ctr., New York City, NY, ⁶Mass Gen. Cancer Ctr., Boston, MA, ⁷The Frank Crick Inst., London, United Kingdom

Abstract Body:

In many areas of oncology, we lack sensitive tumor-burden monitoring to guide critical decision making. While circulating tumor DNA (ctDNA) promises to enable disease monitoring, this approach is limited by the sparsity of ctDNA in the plasma. To overcome this challenge, error-corrected deep targeted sequencing has been proposed. Nonetheless, this framework is limited by the low number of genomic equivalents (GEs, typically $\sim 10^3$ /mL of plasma), imposing a ceiling on effective sequencing depth. We have previously shown that genome-wide mutational integration through plasma whole genome sequencing (WGS) can sever the dependency between available GEs and assay sensitivity (Zviran *et al*, Nature Medicine, 2020). In this approach, tumor-informed mutational profiles are applied to plasma WGS, allowing detection of tumor fractions as low as 10^{-5} . However, the higher cost of WGS limits practical depth of coverage (20-30X) and may limit broad adoption.

Lower sequencing costs may thus allow for enhanced ctDNA cancer monitoring via WGS. We therefore applied emerging lower-cost WGS (\$1/Gb, Almogy *et al*, biorxiv, 2022) to plasma from 7 patients with metastatic cancer at 80-100X coverage depth. Read depth profiling and error rates were comparable between matched Ultima and standard platform datasets. Integration of deep learning architectures for signal to noise enrichment (Widman *et al*, biorxiv, 2022) with deeper WGS coverage enabled ctDNA detection at the parts per million range.

We further reasoned that lower sequencing cost can be harnessed for duplex error-corrected WGS. Applying this to PDX samples showed order of magnitude decrease in errors, and error rates $\sim 6 \times 10^{-7}$ in samples from patients with metastatic melanoma. We further used this approach to tackle the more challenging context of cancer monitoring in early-stage melanoma without matched tumor sequencing. While in uncorrected WGS, *de novo* mutation calling yielded limited ability to detect melanoma specific mutations, duplex-corrected WGS allowed us to harness melanoma mutational signatures for disease monitoring without matched tumor profiling.

In summary, these data demonstrate the exciting potential of low cost WGS for ultra-sensitive ctDNA cancer monitoring. In the tumor-informed setting, deeper sequencing increased sensitivity for mutational profile detection. Moreover, the application of duplex error-correction at genome scale allowed for sensitive cancer monitoring without matched tumor profiles. We envision that the era of low-cost sequencing will thus empower ultra-sensitive cancer monitoring via WGS, with transformative impact on cancer care.

Cancer Posters - Wednesday

PB1064. Deep learning identifies gene expression and DNA methylation as highly predictive of lung cancer histologic subtype.

Authors:

M. Betti¹, E. R. Gamazon^{1,2}, M. C. Aldrich¹; ¹Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Clare Hall, Univ. of Cambridge, Cambridge, United Kingdom

Abstract Body:

Genomics-based molecular assays such as RNA-seq and DNA methylation profiling have gained traction in the clinic. This increase in clinical use has been particularly true in the context of cancer diagnosis and treatment, where biomarker-based tumor characterization plays a key role in targeted treatments. While high dimensionality often limits the interpretability of large genomic datasets, deep learning-based approaches have shown enormous promise due of their ability to model non-linear relationships within high-dimensional data types. We aimed to evaluate the applicability of deep learning to histological subtype categorization in lung cancer using datasets from three commonly utilized genomic profiling techniques: RNA-seq, DNA methylation, and miRNA-seq.

From The Cancer Genome Atlas (TCGA), we identified 962 lung cancer cases (477 squamous cell carcinoma and 485 adenocarcinoma) for which RNA-seq, DNA methylation, and miRNA-seq were performed with biopsied tumor tissue. Using each of these three distinct data types, we defined and trained an ensemble of three multilayer perceptron architectures, with hyperparameters optimized independently using grid search, to distinguish between lung cancer histologic subtypes.

Before model training, 20 percent of each training set was randomly partitioned to use as an independent validation set. Classification performance of each model was evaluated using the area under the receiver operating characteristic curve (AUC) statistic. Of the three models, the one trained on CpG methylation showed the highest performance, with an impressive AUC of 0.98 in the validation set. The gene expression-based model was a close second, with an AUC of 0.94. Finally, we found miRNA expression to be the poorest predictor of lung cancer histological subtype, with an AUC of 0.61.

Our results suggest that both gene expression and DNA methylation may have strong potential utility as clinical predictors of lung cancer histologic subtype. Although the same was not observed using miRNA, the use of a more complex model architecture such as a convolutional (CNN) or graph neural network (GNN) might improve the diagnostic accuracy of this data type. The high performance of our trained multi-layer perceptron-based models suggests that gene expression and DNA methylation merit further exploration as potential diagnostic biomarkers of lung histology, capable of enhancing dominant imaging-based approaches.

Cancer Posters - Thursday

PB1065. Deep learning morphology profiling identifies and enriches carcinoma cells from effusion samples in real-time for cytological and molecular analysis

Authors:

A. Mavropoulos¹, J. Cruz¹, J. Nieto¹, K. Saini¹, W. Austin¹, M. Phelan¹, J. Kim¹, J. Mei¹, W. Yu², C. Johnson¹, S. Boutet¹, T. Lee², N. Li¹, J. Rao², M. Salek¹, M. Masaeli³; ¹Deepcell Inc, Menlo Park, CA, ²Univ. of California, Los Angeles, Los Angeles, CA, ³Deepcell, Menlo Park, CA

Abstract Body:

Cell morphology is a fundamental cell feature essential for pathological identifications, but is currently analyzed with low throughput, subjective, and laborious methods. We developed a platform that generates high-dimensional morphology data using high resolution bright-field imaging that can classify and sort cells in real time when coupled with our Deep Neural Network architecture. Enriched cells are label-free, unperturbed, and viable, making them amenable to downstream molecular and functional analysis.

Effusion samples are a valuable source of tumor cells used for diagnostic, precision medicine, and biomarker discovery purposes but detection and molecular characterization of cancer cells in body fluids is generally limited by low tumor cell content and high background of non-malignant cells. To address these challenges, we used the Deepcell platform and trained an AI model using >3 million cell images from malignant effusion samples. In silico analysis was used to validate the model (AUC=0.96). AI model performance was tested on clinical samples, with carcinoma cells detected in 100% (n=25) of the carcinoma-positive cytology clinical samples and percent frequency highly concordant to flow cytometry data.

Cell images were used to generate high-dimensional profiles based on morphological features, and tSNE analysis revealed multiple distinct clusters characteristic of carcinoma cells that were also classified as carcinoma that were present in the malignant and absent in the benign ascites sample. Pap staining showed higher purities in sorted versus pre-sorted samples, verifying malignant cell enrichment capability of the platform. Additionally, sorted samples had increased amplitude of deletion and amplification peaks and increased TP53 mutation frequency compared to pre-sorted samples, as demonstrated by CNV analysis and targeted sequencing, respectively. scRNA-seq analysis showed enrichment of EpCAM(+) tumor cells and presence of EpCAM(-) tumor cells with high levels of Claudin-4 and CD24 in sorted cells.

Combined, these data show the Deepcell platform is consistent with current cytology and biomarker assay results, with added benefits of greater accuracy and enrichment of live malignant cells present in effusion samples. The gentle workflow yields label-free, viable cells with unaltered morphological features and transcriptomic profiles that could be used for further downstream analysis. Additionally, high-dimensional morphology analysis provides a new dimension that can be integrated into single cell multi-omics to elucidate tumor cell heterogeneity within effusion fluids to ultimately improve precision medicine.

Cancer Posters - Wednesday

PB1066. Detecting pleiotropic breast cancer susceptibility variants from genome-wide association studies.

Authors:

X. Li, M. Zhou, M. Henricks, V. Loch; St. Cloud State Univ., St. Cloud, MN

Abstract Body:

Breast cancer is the most common invasive cancer and the second leading cause of cancer deaths in women. There exists compelling evidence that some genetic variants are associated with the risk of multiple cancer sites and other cancer-related phenotypes. (i.e., pleiotropy). For example, cross-cancer Genome-Wide Association Studies (GWAS) for two to five cancers have been conducted to identify pleiotropic variants. Previous work has also estimated the genetic correlation between pairs of cancers using data from GWAS for multiple cancer sites. Currently, the association pathways from a variant to different cancer types and cancer-related phenotypes remain poorly characterized. In this study, we investigated different possible pathways between genetic variants and breast cancer. We began by detecting outliers in a single exposure-outcome mendelian randomization (MR) analysis, hypothesizing they are due to horizontal pleiotropy. In MR analysis, variants that exert horizontal pleiotropy are typically treated as a nuisance. Here, we exploited horizontal pleiotropy as an opportunity to understand the underlying shared biological pathways between ovarian cancer and breast cancer. We performed the MR analysis of ovarian cancer and breast cancer using GWAS summary statistics. We identified five horizontal pleiotropy SNPs, three of them were previously reported to be associated with both breast and ovarian cancer (rs10069690/ 5p15/TERT, rs8170/ 19p13/ BABAM1, rs635634/ 9q34/ABO). Index variants in two additional regions (rs1243180/10p12.31/MLLT10, rs183211/17q21.31/NSF) that were only previously associated with ovarian cancer also showed horizontal pleiotropic effects with breast cancer. We also examined the potential causal association between breast cancer and two risk factors, alcohol consumption and smoking. We found a significant association between genetic liability to alcohol consumption and breast cancer ($P=0.014$). Our findings can enhance the understanding of breast cancer and help to target the investigation of genes with greater clinical evidence.

Cancer Posters - Thursday

PB1067. Detection of androgen receptor splice variants from clinical sequencing

Authors:

S. Hwangbo, S. Lee, S. Kim, H. Yun; Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract Body:

Androgen receptor (*AR*) splice variants (*AR*-Vs) have been widely studied on its important role in prostate cancer (PC) progression. In particular, *AR*-Vs lacking ligand binding domains, such as *AR* splice variant 7 (*AR*-v7) that links exon 3 and cryptic exon 3 or *AR* variant that losses exons 5 to 7 (*AR*^{v567es}), have been demonstrated as one of the resistant mechanisms to androgen deprivation therapy in PC. Despite its clinical importance, the current algorithms to detect fusions such as TopHat-Fusion have limitation to detect intra-chromosomal rearrangements including *AR*-Vs. In this respect, we propose a novel approach to identify *AR*-Vs from targeted RNA sequencing (RNA-seq) dataset. In short, our two-step approach first gathers soft-clipped or divided reads that are adjacent to the splicing sites of *AR*-Vs and then yields the number of splitting reads that support the existence of *AR*-Vs. Next, the algorithm selects the paired-end reads whose one side and the other side are mapped to the preceding and following exons, respectively. The final number of spanning reads are calculated following to the removal of low-quality reads. We validated the proposed approach using two large-scale, independent RNA-seq datasets of PC samples: 54 samples from Seoul National University Hospital (SNUH) and 558 samples from The Cancer Genome Atlas (TCGA) dataset. Overall, our approach successfully identified samples with putative *AR*-Vs, including *AR*-v7 and *AR*^{v567es}. In addition, our approach successfully detected *AR*-v7 with reliable number of supporting reads, from the RNA-seq dataset of *AR*-v7-expressing cell line. Finally, statistical analyses of 54 SNUH PC patients suggested potential detection threshold in clinical sequencing settings (at least 8 split reads). In the SNUH dataset, The *AR*-v7 positive group, which included 12 patients (22.2%) with 8 or more split reads, had higher *AR*-v7 expression levels than the negative group ($p=0.0147$). The similar pattern was also found in the analysis of TCGA dataset. Owing to its flexibility, we also confirmed that the developed approach can be used to detection of other *AR*-Vs, including *AR*-v3 and *AR*^{v567es}. We expect that the further in-depth analyses including larger samples and clinical outcomes can discover clinical applicability of *AR*-Vs.

Cancer Posters - Wednesday

PB1068. Detection of APC and MUTYH germline mutations in two south Indian families with hereditary colorectal cancer.

Authors:

S. Akula, S. Valmiki, K. K. Mandapati, D. R. Vegulada; Genes N Life Hlth.Care, Hyderabad, India

Abstract Body:

Globally, colorectal cancer (CRC) is recognized both as the second most deadly cancer and third most prevalent disease, with significant projected growth in incidence. Addressing this concern requires a thorough understanding of its medical origins. Most colorectal cancers are somatic, but 5-10% of CRC are associated with germline mutations, including within APC and MUTYH genes that cause familial adenomatous polyposis (FAP). This study was carried out with the aim to identify these specific germline mutations in two South Indian families with CRC incidence. The primary step was the collection of formalin-fixed paraffin-embedded tumor tissue sections from the index patient and screening for microsatellite instability by immunohistochemistry to rule out hereditary nonpolyposis CRC, which was negative. To detect germline mutations, adhering to Mayo Clinic and American College of Medical Genetics' guidelines for genetic testing, researchers collect blood samples from the index patient and eleven other family members. All coding regions and intron-exon junctions of APC and MUTYH genes were analyzed using Sanger sequencing. No mutations were detected in the MUTYH gene in the index patient. However, a missense mutation (c.5465A>T p.D1822V) was identified within the APC gene in the index patient. All eleven family members were screened for the targeted c.5465A>T mutation, with 4/11 family members having it. Sigmoidoscopy revealed the presence of the polyps in two of the family members having the mutation. The APC mutation has previously been reported to be the most common allele associated with increased risk of CRC triggered by dietary fat intake, indicating an aspect that affected family members must remain attentive to. Through this study's genetic counseling, these family members were given precise, early diagnoses and guided clinical management for future surveillance, which must be accomplished on a wider scale through further research.

Cancer Posters - Thursday

PB1069*. Detection of circulating tumor DNA in resectable pancreatic ductal adenocarcinoma

Authors:

J-S. Lee¹, S. Cho¹, M-W. Seong¹, J. Jin-Young¹, S. Park²; ¹Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, ²Seoul Natl. Univ Hosp, Seoul, Korea, Republic of

Abstract Body:

Background: Circulating tumor DNA (ctDNA) is a promising biomarker for early tumor detection and minimal residual disease (MRD) assessment in early-stage cancer, but quantifying minute amounts of ctDNA in the early stage of disease is challenging. Here, we adapted an ultrasensitive next-generation sequencing (NGS) technology and performed parallel analysis of pre- and post-operative ctDNA and matched tumor tissues in a prospective cohort of patients with resectable pancreatic ductal adenocarcinoma (PDAC). **Methods:** In total, 70 consecutive patients undergoing curative resection for resectable PDAC were enrolled from August 2020 through October 2021. We performed integrated digital error suppression enhanced cancer personalized profiling by deep sequencing (CAPP-Seq) NGS of triple-matched samples [pre/post-operative plasma cell-free DNA (cfDNA), tumor tissue, and germline DNA] targeting 77 genes. **Results:** Preoperative ctDNA was detected in 37.7% of the evaluable patients, with a median variant allele frequency of 0.09%. Twelve additional oncogenic mutations were detected exclusively in pre-operative ctDNA but not in tissue. The risk of early recurrence tended to be higher in patients with detectable post-operative ctDNA than in patients with negative conversion of ctDNA after surgery. cfDNA variants from 24.5% of patients had features compatible with clonal hematopoiesis. **Conclusion:** An optimized NGS approach might add value beyond tissue analysis through the highly sensitive detection of minute amounts of ctDNA in resectable PDAC. Pre/post-operative ctDNA changes could be a potential tool for MRD assessment. Moreover, parallel analyses of matched tissues and leukocytes might be required to accurately detect clinically relevant ctDNA in the clinical laboratory.

Cancer Posters - Wednesday

PB1070. Detection of *EGFR* exon skipping variants using clinical sequencing and its application

Authors:

S. Kim, S. Hwangbo, S. Lee, H. Yun; Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract Body:

The epidermal growth factor receptor (EGFR) gene is amplified in multiple tumors including brain and lung cancers, and some of those amplifications accompany EGFR exon skipping variants of which resultant transcript lacks a specific range of exons. To date, many EGFR exon skipping variants have been reported, such as vII (skips exon 14 and 15), vIII (skips exon 2 to 7), and vIV (skips exon 25 to 27 or 25 and 26). Since many studies successfully showed their oncogenic roles, those exon skipping variants are considered as a promising anti-cancer drug target. Despite its potential, detecting the exon skipping variants from clinical data is still limited, and only a few studies demonstrated its identification using Next-Generation Sequencing (NGS) technology. In this study, we developed a straightforward approach to identify EGFR exon skipping variants using targeted sequencing which is frequently used for clinical purposes. The developed algorithm enables identification of three EGFR exon skipping variants (vII, vIII, and vIV) from either DNA or RNA targeted sequencing dataset. We applied the variant detecting algorithm to a large-scale clinical sequencing dataset of 2,897 DNA-RNA matched cancer samples and successfully identified 44 putative EGFR exon skipping variants (5 EGFRvII, 30 EGFRvIII, and 9 EGFRvIV) from 35 (1.21%) samples. Most (38/44, 86.4%) of the EGFR variants were identified from glioblastoma multiforme (GBM) and several variants (3/44, 6.8%) came from lung cancer. All the EGFR exon skipping variants except 1 specific case was accompanied by EGFR amplification. In conclusion, our developed approach effectively detected the EGFR exon skipping variants from targeted sequencing data for clinical use. To accurately identify the exon skipping variants, both DNA and RNA sequencing would be necessary.

Cancer Posters - Thursday

PB1071. Detection of somatic copy number alterations using paired DNA and RNA sequencing

Authors:

M. Lastrapes^{1,2}, Z. Ozcan^{1,2}, Y. Jakubek³, J. Wong², P. Scheet^{1,2}; ¹UTHIth.MD Anderson Graduate Sch. of BioMed. Sci., Houston, TX, ²Dept. of Epidemiology, MD Anderson Cancer Ctr., Houston, TX, ³Dept. of Internal Med., Univ. of Kentucky, Lexington, KY

Abstract Body:

Somatic copy number alterations (SCNAs) have long been known to contribute to tumorigenesis and even show prognostic importance in many different types of cancer, though robust detection of SCNAs requires significant coverage across the genome. Detection of SCNAs is classically done using DNA microarray or whole-genome sequencing of DNA, but recent work has shown promising results in detecting SCNAs from whole exome sequencing (WES) data. With the increase in multi-omic cancer datasets, we hypothesize that the combination of paired DNA sequencing and RNA sequencing (RNAseq) data can further improve detection of SCNAs from NGS data by increasing overall genomic coverage. Here, we evaluate SCNA detection using a haplotype-based approach, applied to both WES data alone and in combination with RNAseq from 95 participants of The Cancer Genome Atlas glioblastoma (TCGA-GBM) project. Genotyped single-nucleotide polymorphisms (SNPs) from each data type were aggregated within unique samples, then input into our computational pipeline for detecting SCNAs from SNP data (hapLOH). The resulting detected SCNAs were benchmarked against the Affymetrix SNP array data, and performance of our analysis of combined data was compared to our analysis of WES alone. We report that SCNA detection using combined data resulted in a 10% absolute increase in sensitivity with only a 2% decrease in specificity at the gene-level relative to WES data alone, showing potential for the use of aggregation of DNA and RNA sequencing to better detect SCNAs in cancer tissue. To further account for this decrease in specificity, we plan to extend this work by exploring additional methods of combining genotype information from data sources and controlling for possible confounders such as RNA coverage bias due to highly expressed genes or genotype disagreements between data types. Our initial analysis of SCNA detection from combined DNA and RNA sequencing shows promising improvements compared to using either DNA or RNA alone, highlighting the clinical benefit of fully leveraging multi-omic datasets in genomic characterization of tumors.

Cancer Posters - Wednesday

PB1072. Determining the mechanisms of radioresistance in breast cancer

Authors:

B. McBean¹, A. Michmerhuizen¹, K. Wilder-Romans¹, B. Chandler², L. Lerner¹, M. Liu¹, A. Boyle¹, C. Speers¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Dana-Farber Cancer Inst., Boston, MA

Abstract Body:

Background: Breast cancer (BC) is the most diagnosed cancer globally and is the deadliest for women. Clinical management of BC includes radiation therapy (RT), with upwards of 85% of women receiving RT as part of their treatment regimen after breast conserving surgery. Although effective, over 10% of women will develop a local recurrence despite RT, a rate that is much higher in women with triple-negative or inflammatory BC. Unfortunately, the molecular mechanisms that underly RT response and intrinsic radioresistance or radiosensitivity are poorly understood. We hypothesized that transcriptomic and proteomic changes that occur early and late after ionizing radiation in intrinsically radiosensitive and resistant BC models would offer mechanistic insight into mediators of this differential response. To test this, we conducted time-course RNA-seq and reverse-phase protein array (RPPA) experiments to characterize transcriptomic and proteomic changes following RT across a panel of BC cell lines of varying levels of intrinsic radiation. **Methods:** Clonogenic survival assays were used to measure the surviving fraction (SF) after 2 Gy of RT across 21 BC cell lines where SF was used as a continuous. On 12 of these cell lines, we also treated cells with 4 Gy RT and RNA was collected 3, 12, and 24 hours after treatment for sequencing and differential gene expression analysis with DeSeq2. Protein was collected 1, 12, and 24 hours after radiation for RPPA analysis with SuperCurve. **Results:** Clonogenic survival identifies a range of radiation sensitivity in human BC cell lines (SF 77%-17%) with no significant correlation ($r < 0.3$) to the intrinsic BC subtype. Using the most radiosensitive (ACC-422 and HCC-1937) and radioresistant (MDA-MB-453 and BT-549) cells, we performed pathway enrichment analysis for genes differentially expressed after RT. Among the most highly enriched pathways from the transcriptomic data, we found processes including DNA damage repair, apoptosis, and cell cycle maintenance. From the proteomic data, we found that proteins including p53, Bcl-2 family proteins, and cell cycle proteins exhibit expression changes after 1 hour. **Conclusions:** Ionizing radiation induces transcriptomic and proteomic expression changes that differ between intrinsically sensitive and resistant BC models. These pathways offer potential insight into the mechanisms underlying intrinsic radioresistance and suggest biologic vulnerabilities that may be targeted to more effectively treat women at a high risk of local BC recurrence. Genome wide CRISPR/Cas9 screens are currently underway in these breast cancer models to confirm these vulnerability targets.

Cancer Posters - Thursday

PB1073. Developing a high-quality, sample-to-result Hereditary Breast and Ovarian Cancer Panel Assay Pipeline for using a novel sequencing platform.

Authors:

A. Bhattacharjee¹, G. Liu¹, E. Frise¹, S. Kruglyak², B. Lajoie², K. Blease², J. Zhao², M. Reese¹, S. Levy²; ¹Fabric Genomics, Oakland, CA, ²Element BioSci.s, San Diego, CA

Abstract Body:

Background: Recent testing recommendations for *BRCA1/2* emphasize the changing landscape for hereditary breast and ovarian (HBOC) from medical history-based to genomic screening approach, and the public health benefit due to clear actionability and ascertainment of the latter approach. Currently, 20 genes are recommended for screening for HBOC based on payor reimbursement and evidence-based guidelines. However, prevention remains inadequate; plagued by a) high test cost, and b) challenges in NGS data quality in genes that are difficult to resolve by standard targeted sequence capture. Most sequencing platforms offer cost effective sequencing with substantial multiplexing. Therefore, high-throughput sequencers do not often fit well with smaller scale clinical testing platforms that require fast turn-around times and the ability to process a few samples at a time. **Approach:** We developed a cost-effective comprehensive testing approach using a novel sequencing platform that enables flexibility in scale while matching the cost of the highest throughput platforms. The Element Aviti benchtop sequencer provides a testing solution with flexible scale that enables small batch sizes with low error rate and novel performance for mutation detection of both point mutations and indels. Nine genes (*BRCA1*, *BRCA2*, *EPCAM*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *PTEN*, *STK11*) often include challenging CNV events that are relevant to report in the screening test. We explore software and library preparation methods to enhance the detection of complex variants in these genes. Screening accuracy performance was assessed on all workflows on a blinded series of samples using custom CNV calling tools across the 9 genes. **Results:** In a pilot study of 51 patient samples we have compared current clinical short read sequencing with a new sequencing platform by Element Biosciences. The error profiles look similar and overall concordance for a total of 272 clinical relevant variants from ClinVar show concordance in 270 cases with 2 sites being inconclusive due to a neighboring homopolymer run for both platforms. All 10 pathogenic variants including CNV variants could be detected on both platforms. We will further present a summary and comparison of the utilized workflows for target capture and sequencing with a focus on CNV workflow to detect CNV events in challenging regions. **Conclusion:** A novel NGS algorithm was optimized for detecting SNVs, indels and CNVs including difficult CNV events that met the cost and quality aspects. Its routine use would therefore enable early surveillance. The approach can be extended to other application of screening of public health importance.

Cancer Posters - Wednesday

PB1074. Development and Clinical Validation of a Bioinformatics Pipeline for Cancer Transcriptome Sequencing

Authors:

K. Cao¹, L. Fumin², F. Xu¹, Z. Fan¹, M. Li¹; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Univ. of Los Angeles, Los Angeles, CA

Abstract Body:

Tumor transcriptome profiling has emerged as the method of choice to identify cancer-related genomic alterations, especially for fusion genes and targetable altered pathways. However, the clinical application of transcriptome sequencing (TS) is lagging due to a lack of clinically validated bioinformatics pipeline. We designed, validated, and implemented a clinical bioinformatics pipeline (ConcordR) for the cancer TS test to support the needs of precision cancer care.

The ConcordR aligns paired fastq files of tumor samples to the human reference genome; the alignments are then analyzed for gene fusion using multiple tools to maximize the detection rate. After stepwise-structured multiple-layer filtrations using internal databases, the fusion results from different callers are merged and annotated. In parallel, the TPM of each gene is calculated and normalized based on selected housekeeping genes. The gene expression data of the tumor are then compared to that of pan-cancer and different tumor cohorts to identify differentially expressed (DE) genes for potentially targetable biomarkers, gene expression signature and pathway analyses to inform personalized patient care.

The accuracy and precision of fusion detection using ConcordR were evaluated using 66 tumor samples previously studied. ConcordR correctly identified all fusions in positive cases and no fusion in negative cases in a double-blind manner, resulting in 100% sensitivity and specificity. The reproducibility and repeatability of gene expression profiling were assessed by sequencing five tissue types each in triplicates in different batches. Statistical analyses showed that the identification of DE genes is highly reproducible on all tissue types except FFPE samples, which demonstrated significant variations. Clinical validation showed that ConcordR successfully identified cases with gene overexpression such as *MYCN* overexpression in pediatric neuroblastomas, *CCNE1* and *MYC* overexpression in pediatric osteosarcomas. Clinical application of ConcordR has facilitated the diagnosis of many undiagnosed cancers, such as the identification of a novel fusion *ATXNIL-NUTM2A* in a patient with congenital disseminated tumor of unknown origin leading to the diagnosis of *CIC*-rearranged sarcoma. ConcordR has also detected biomarkers for targeted therapy, such as targeting *EGFR* overexpression in high grade gliomas.

In summary, we have developed a clinical bioinformatics pipeline for pediatric cancer transcriptome sequencing. The validation results demonstrate robust analytic accuracy and precision in gene fusion identification and gene expression profiling in most tissue types.

Cancer Posters - Thursday

PB1075*. Development and Clinical Validation of a Liquid Biopsy Assay for Neurofibromatosis 1.

Authors:

F. Xu¹, C. Kotch¹, M. Koptyra¹, M. Lueder¹, W. Fu¹, K. Cao¹, A. Long¹, Z. Fan¹, S. Brown¹, T. De Raedt¹, J. Foster¹, M. Fisher¹, M. Li^{1,2}; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Univ of Penn, Philadelphia, PA

Abstract Body:

Introduction: Neurofibromatosis type 1 (NF1) is an autosomal dominant tumor predisposition syndrome affecting 1 in 2500 individuals. Common NF1-associated tumors include low and high-grade gliomas, and peripheral nerve sheath tumors. NF1 patients often have a high number of low-grade tumors of which only some will undergo malignant transformation. Detecting this progression early is challenging but extremely important, as the associated mortality is very high. Moreover, NF1 patients often have multiple tumors and standard imaging is unable to distinguish benign from malignant tumors. We developed a liquid biopsy (LB) NGS assay to detect tumor-associated mutations in cell-free DNA (cfDNA) with the goal of early detection of transforming or transformed tumors and defining the molecular makeup of the tumors. **Methods:** The assay interrogates 20 genes commonly mutated in NF1-related malignancies. A total of 45 plasma samples were collected from NF1 or other patients with tumors harboring mutations covered by the assay. Two positive and 9 negative cfDNA controls were used to evaluate the accuracy and precision of the assay. cfDNA was extracted using QIAamp Circulating Nucleic Acid Kit. The yield and integrity of cfDNA were determined by Qubit and qPCR, respectively. The regions of interest were enriched using the anchored multiplex PCR method and sequenced using NovaSeq 6000 or MiSeq. The sequence data were processed using Archer Analytics v6.0. **Results:** All known variants including 19 samples with known NF1 germline mutations and 32 mutations in the positive controls were correctly identified at comparable variant allele fractions (VAF), demonstrating 100% sensitivity. No false-positive variants were identified in the negative controls, indicating 100% specificity. The reproducibility and repeatability of the assay were 100% based on the inter-run and intro-run of two positive controls. The assay reliably detected variants at 0.5% VAF and can detect variants with VAF <0.1% using a modified bioinformatics pipeline. 13 mutations found in the tumor samples were identified in 8 plasma samples with a VAF range from 0.07% to 3.60%. Early data suggest that the presence of tumor-associated mutations in cfDNA appeared to be associated with tumor size; however, more cases are needed to confirm this finding. **Summary:** We developed and validated an NF1 LB panel that can effectively detect mutations in cfDNA with a 100% sensitivity and specificity using validation samples. Hence, this assay could potentially be used as a non-invasive method to monitor patients with NF1 serially to identify early-stage malignancies and potentially tumors under malignant transformation.

Cancer Posters - Wednesday

PB1076. Development of a breast cancer risk prediction model with carrier status, a polygenic risk score, and epidemiologic risk score.

Authors:

S. Kalia¹, N. Boddicker², S. Yadav², C. Hu², F. Couch², **P. Kraft¹**, the CARRIERS Consortium; ¹Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ²Mayo Clinic, Rochester, MN

Abstract Body:

Breast cancer has been associated with monogenic, polygenic, and epidemiologic (clinical, reproductive and lifestyle) risk factors, but direct empirical data on their joint effects is limited. We extended a risk model incorporating pathogenic variants (PV) in six breast cancer predisposition genes and a 105-SNP polygenic risk score (PRS) to include an epidemiologic risk score (ERS). This study was performed in a population-based sample of more than 22,700 cases and a similar number of age-matched controls from the Cancer Risk Estimates Related to Susceptibility (CARRIERS) Consortium. We assessed effect measure modification among the modeled factors, age, and family history of breast cancer: interaction terms were chosen via penalized logistic regression with cross validation. Our final model includes a synergistic interaction between the ERS and PRS, and antagonistic interactions among age, PRS and *BRCA1/2* status. Adding the ERS increases risk discrimination from an AUC of 0.67 (95% CI: 0.66-0.67) for a model with PV and PRS, to an AUC of 0.68 (0.68-0.69) with addition of the ERS. A 40-year-old female at the median PRS with the highest ERS has an odds of breast cancer two times that of an age-matched female at the median PRS but with the lowest ERS, while the same odds ratio among 70-year-old females is 4.5.

Our results illustrate that the ERS, alone and in combination with the PRS, can contribute to clinically meaningful risk stratification across high-risk thresholds for recommending risk-reducing medications and breast MRI screening, especially for carriers of a PV in a moderate penetrance gene such as *ATM* or *CHEK2*. Although all *ATM* or *CHEK2* carriers are classified above the 20% lifetime risk threshold for breast MRI based on their monogenic risk alone, 28.5% of *ATM* carriers and 17.0% of *CHEK2* carriers with a family history of breast cancer and 60.1% of *ATM* carriers and 46.6% of *CHEK2* carriers without a family history are reclassified below the threshold when the ERS and PRS are incorporated into the risk prediction. *CHEK2* carriers at the 10th percentile of the joint distribution of ERS and PRS, without a family history of breast cancer, have a predicted lifetime risk of 12.5%, near the average of 13% for a female in the US. Appropriately integrating monogenic, polygenic, and epidemiologic risk factors to improve breast cancer risk prediction models may inform personalized screening protocols and prevention efforts.

Cancer Posters - Thursday

PB1077. DHGAN, Digital Histology Generative Adversarial Networks Used in Creating Synthetic Genitourinary Tissue Slides

Authors:

H. Arora¹, D. Van Booven², A. Noman³, F. Ismael³, I. Xu¹, V. Sandoval⁴, J. A. Gomez⁵, S. Punnen¹; ¹Univ. of Miami, Miami, FL, ²Univ Miami Miller Sch Med, Miami, FL, ³Dow Univ. of Hlth.Sci., Karachi, Pakistan, ⁴Western Univ., Dept. of Surgery Div. of Urology, London, ON, Canada, ⁵Western Univ., Dept. of Pathology and Lab. Med., London, ON, Canada

Abstract Body:

Introduction: Recent integration of open-source data to machine learning models, especially in the medical field, in the domain of image analysis has opened new doors to studying disease progression and/or regression. However, limitations of the medical data for machine learning approaches is the insufficient quantity and quality (technical variation including bubbles in the slide, discoloration, and tissue folding) of data to a particular medical condition. In this context, synthetic data augmentation by using generative adversarial networks (GAN) could potentially generate high-quality data that preserve clinical variability. **Methods:** Tissue images were downloaded from 11 different genitourinary tissues from the Genotype-Tissue Expression Database (GTEx) Histology Repository. Images were subjected to quality control using HistoQC, and then segmented to patch sizes which ranged from 64, 128, and 256 with PyHIST. A conditional GAN was coded to create new synthetic images based of these training segments. The synthetic segments were then given to board certified pathologists for inspection and quality control. A separate classification machine learning step was used at the final stage that included the RNA sequencing data obtained from GTEx to assign accuracy to classify synthetic images to their proper tissue group. **Results:** Upon inspection 80% of the synthetic tissue images were assigned as “passed QC” by our pathologists. Further, the classification algorithm was able to assign on average 73.2% of the synthetic images to their correct origin. Patch size accuracy was determined to be lower than larger patches (45.5% vs 73.2%). Final accuracy numbers were found to be higher in the RNA sequencing classification network at an average of 91.0%. **Conclusions:** Using the GAN to create synthetic images is approaching a level of accuracy similar to any other successful classification method. With some work on the image analysis, the synthetic images will reach a quality which can assimilate clinical variability of original histology.

Cancer Posters - Wednesday

PB1078. Diagnostic gene fusion detection in cancer by adaptation of short-read CHOP Fusion Panel to Oxford Nanopore long-read sequencing.

Authors:

K. Rybacki^{1,2}, L. Fang², F. Xu², Y. Hu², M. Ahsan², M. Li^{2,3}, K. Wang²; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Children's Hosp. of Philadelphia, Philadelphia, PA, ³Univ. of Pennsylvania Med. Sch., Philadelphia, PA

Abstract Body:

Gene fusion is the process where two distinct genes fuse, typically as the result of structural variants such as chromosomal translocation, interstitial deletion, or chromosomal inversion, or rarely, trans-splicing events. Gene fusions can be used as biomarkers in cancer diagnosis and possible therapeutic targets. However, current short-read next-generation sequencing techniques have limitations in detecting the full spectrum of gene fusions and in resolving repetitive or low-complexity regions. In addition, indexing and pooling multiple samples can take a few days, which is not ideal for time-sensitive scenarios where gene fusions provide decision support for cancer diagnosis, treatment and prognosis. The Division for Genomic Diagnostics (DGD) at the Children's Hospital of Philadelphia (CHOP) previously developed a custom-designed RNA sequencing panel, CHOP Fusion panel, using anchored multiplex PCR technology. The panel interrogates 117 cancer genes known to be involved in gene fusions with diagnostic significance, and it works on bone marrow, blood, fresh/frozen tissue, and formalin-fixed, paraffin-embedded (FFPE) samples. We evaluated the possibility to adapt the CHOP Fusion Panel through long-read sequencing on the Flongle flowcell for fast-turnaround and low-cost sequencing with potential point-of-care applications. We performed pilot sequencing on five positive samples with paired Illumina data and determined gene fusion events by LongGF and other computational tools. Due to the high number of sequencing reads needed to confidently identify gene fusions, PCR validation was carried out for the identified gene fusions from LongGF. We also "Illuminized" the Nanopore long-read data into 150bp paired-end Illumina reads and evaluated short-read based computational pipelines. Our preliminary results suggest that sequencing can be stopped within 8 hours and data analysis can be completed within 1 hour, therefore offering significant advantages over conventional Illumina sequencing. In summary, this study demonstrated technical feasibility to adapt CHOP Fusion Panel and use an automated workflow for rapid detection of known and potential novel gene fusions on the Flongle flowcells via long-read sequencing.

Cancer Posters - Thursday

PB1079. Dietary factors and endometrial cancer risk: A Mendelian randomization study.

Authors:

T. O'Mara, D. Glubb, X. Wang; QIMR Berghofer Med. Res. Inst., Brisbane, Australia

Abstract Body:

Given the strong association between obesity and endometrial cancer risk, dietary factors may play an important role in the development of this cancer. However, observational studies of micro- and macronutrients and their role in endometrial cancer risk have been inconsistent. Clarifying these relationships are important to develop nutritional recommendations for cancer prevention.

Using Mendelian randomization (MR), we investigated the effects of circulating levels of 14 micronutrients (α -carotene, β -carotene, calcium, copper, folate, iron, magnesium, phosphorus, retinol, vitamin B6, vitamin B12, vitamin C, vitamin E and zinc) as well as corrected relative macronutrient intake (protein, carbohydrate, sugar and fat) on endometrial cancer risk.

Two-sample MR for endometrial cancer and its subtypes (endometrioid and non-endometrioid histologies) was performed using genome-wide association study summary statistics from the Endometrial Cancer Association Consortium (up to 12,270 cases and 46,126 controls). Inverse-variance weighted MR analyses were performed, with sensitivity analyses (including MR-Egger and MR-PRESSO analyses) to verify MR assumptions. For macronutrient analyses, multivariable MR was performed to account for genetic correlation between phenotypes.

A relationship was observed between increased genetically predicted circulating vitamin C and endometrial cancer risk (OR per SD increase 1.41 95%CI 1.16-1.72 $P=0.0007$), consistent with a previous MR study. Investigation of macronutrients found increased sugar intake associated with risk of endometrial cancer (0.43 95%CI 0.24-0.79 $P=0.006$), and its subtypes (endometrioid 0.38 95%CI 0.19-0.77 $P=0.007$; non-endometrial cancer 0.09 95%CI 0.02-0.37 $P=0.0008$). Carbohydrate intake associated with reduced risk of endometrioid endometrial cancer risk (0.34 95%CI 0.15-0.74 $P=0.006$). Increased fat intake associated with increased risk of non-endometrioid endometrial cancer risk (10.7 95%CI 1.8-63 $P=0.008$). These results were consistent across sensitivity analyses. There were no significant results observed in multivariable MR analyses including all macronutrients.

In summary, these findings suggest that certain dietary factors influence endometrial cancer risk but further investigation is required.

Cancer Posters - Wednesday

PB1080*. Differentially expressed serum RNAs as potential early markers for primary sclerosing cholangitis-associated cholangiocarcinoma.

Authors:

G. Hemmrich-Stanisak¹, S. Juzenas¹, T. Folseraas², K. Boberg³, C. Schramm⁴, T. Karlsen², A. Franke¹; ¹Univ. of Kiel, Kiel, Germany, ²Res. Inst. for Internal Med., Div. of Surgery, Inflammatory Diseases and Transplantation, Oslo Univ. Hosp. Rikshosp.et and Univ. of Oslo, Oslo, Norway, ³Res. Inst. of Internal Med., Oslo Univ. Hosp. Rikshosp.et, Oslo, Norway, ⁴Univ. Med. Ctr. Hamburg-Eppendorf, Hamburg, Germany

Abstract Body:

Cholangiocarcinoma (CCA) is a common malignancy in patients with primary sclerosing cholangitis (PSC) and carries a high rate of mortality. Although the pathogenesis of CCA in PSC is largely unknown, inflammation-driven carcinogenesis concomitant with various genetic and epigenetic abnormalities are underlying factors. When PSC-associated CCA is diagnosed, most tumors are unresectable, and no effective medications are available. Given the poor therapeutic outcome, the surveillance and management of PSC patients who are at an increased risk of developing CCA are of importance. In this study we aim to identify differentially expressed candidate microRNAs and/or other small RNAs, that could be used as early markers to stratify PSC patients, which will develop PSC-associated CCA. To do so we performed differential expression analysis in a German cohort (n=65) between patients with PSC-associated CCA (n=11), PSC patients who developed CCA during a 3.5y follow-up period (n=21) and PSC patients that did not develop a malignancy (n=33). To identify potential prognostic RNA-markers we furthermore conducted survival and progression analyses on this data set. To validate our findings we plan to use a similar matched Norwegian cohort comprising 21 PSC-associated CCA patients, 19 patients that developed CCA (7y follow-up) and 25 PSC-only patients. Our exploratory results derived from the German data set already indicate several differentially expressed microRNAs that a) represent potential markers for the development and/or presence of PSC-associated CCA and b) are also among the top significant prognostic candidates in the survival analysis, thus might even serve as early cancer markers. All of the identified serum microRNAs await now validation in the above described independent Norwegian cohort.

Cancer Posters - Wednesday

PB1081. Discovery of novel predisposing coding and noncoding variants in familial Hodgkin lymphoma

Authors:

J. Myers¹, J. Flerlage¹, J. Maciaszek¹, N. Oak¹, S. Rashkin¹, Y. Hui¹, Y-D. Wang¹, W. Chen¹, G. Wu¹, T-C. Chang¹, K. Hamilton², S. Tithi¹, L. Goldin³, M. McMaster⁴, M. Rotunno⁵, N. Caporaso⁶, A. Vogt⁶, D. Flamish⁷, K. Wyatt⁷, J. Liu⁷, M. Tucker⁶, C. Mullighan¹, K. nichols¹, M. Metzger¹, J. Yang¹, E. Rampersaud¹; ¹St. Jude Children's Res. Hosp., Memphis, TN, ²Dana-Farber Cancer Inst., Boston, MA, ³DCEG/NCI, Bethesda, MD, ⁴DCEG/NCI/NIH/DHHS, Bethesda, MD, ⁵Natl. Cancer Inst., Bethesda, MD, ⁶NCI, Bethesda, MD, ⁷NIH, Bethesda, MD

Abstract Body:

Familial aggregation of Hodgkin lymphoma (HL) has been demonstrated in large population studies, pointing to a genetic predisposition to this hematological malignancy. To understand the genetic variants associated with the development of HL we performed whole-genome sequencing on 234 individuals with and without HL from 36 pedigrees that had 2 or more first-degree relatives with HL. Our pedigree selection criteria also required one affected individual to be less than 21 years of age and overall, the median age at diagnosis was 21.98 years of age (3 - 55 years). Rigorous family-based segregation analysis was performed for the identification of coding and noncoding variants using linkage and filtering approaches. Our rule-based variant prioritization identified 44 HL risk variants in 28 pedigrees, of which 33 are coding, and 11 are noncoding. Overall, there were 4 highly prioritized recurrent variants: a coding variant in KDR (rs56302315), a 5'UTR variant in KLHDC8B (rs387906223), a noncoding variant in an intron of PAX5 (rs147081110), and another noncoding variant in an intron of GATA3 (rs3824666). The GATA3 variant overlaps a DNase I hypersensitivity locus as well as TCF3 and TCF12 binding motifs and in silico prediction indicates a loss of binding potential for the variant compared to the reference allele. A newly identified splice variant in KDR (c.3849-2A>C) was observed for one pedigree and high confidence stopgain variants affecting IRF7 (p.W238*) and EEF2KMT (p.K116*) were also observed. KDR, KLHDC8B, PAX5, and GATA3 are well known HL candidate genes whereas IRF7 is known to play a role in innate immune response which could have implications with regard to viral infection for that pedigree, and EEF2KMT is a methyltransferase that could affect the epigenetic landscape in the pedigree it was observed in. Gene-level recurrence of different truncating variants was seen for POLR1E in three independent pedigrees as well. POLR1E was not on our candidate gene list; however, the finding is intriguing since interchromosomal arrangements involving POLR1E have been described in other types of B-cell lymphomas. While germline variants in KDR and KLHDC8B have previously been reported, PAX5, GATA3, IRF7, EEF2KMT, and POLR1E represent novel observations. Although there may be environmental factors influencing lymphomagenesis, we observed segregation of candidate germline variants likely to predispose Hodgkin lymphoma in the majority of pedigrees studied.

Cancer Posters - Thursday

PB1082. Discovery of transcriptomic predictors of colorectal cancer mortality by race and ethnicity

Authors:

H. Yin^{1,2}, J. R. Huyghe², D. Redwood³, T. A. Harrison¹, A. D. Dawson², S. S. Thomas², B. W. Seaton², A. K. Koehne², H. Green⁴, K. Nasir⁵, J. J. Tiesinga⁶, C. Whitlow⁴, L. Hsu², J. C. Figueiredo⁵, L. Li⁴, T. K. Thomas³, L. Christopher^{2,1}, U. Peters^{2,1}; ¹Univ. of Washington, Seattle, WA, ²Fred Hutchinson Cancer Ctr., Seattle, WA, ³Alaska Native Tribal Hlth.Consortium, Anchorage, AK, ⁴Ochsner Clinic Fndn., New Orleans, LA, ⁵Cedars-Sinai Med. Ctr., Los Angeles, CA, ⁶Alaska Native Med. Ctr., Anchorage, AK

Abstract Body:

Tribal Health Organizations recognize the high rates of colorectal cancer (CRC) among Alaska Native people and are undertaking initiatives to address it including improved screening and engaging in research. Tumor profiling study in patients from diverse racial and ethnic groups will lead to novel molecular insights about CRC. So, we designed a nested case-control study with including of 840 African American, Alaska Native, Hispanic, and non-Hispanic White CRC patients. We included 70 cases and 140 controls within each group. Cases were patients who died of CRC and controls were CRC patients who survived at least as long as the case and were matched on stage, tumor site, age, sex, race and ethnicity, and year of diagnosis. We are generating RNA-seq data using RNA extracted from formalin-fixed paraffin-embedded (FFPE) tumor tissues. Using the tumor transcriptomic data, our study aims to characterize tumor gene expression and immune microenvironment features, and evaluate their relationships with CRC mortality and risk factors (e.g. family history). Our initial RNA-seq run included 87 patients who were equally distributed across four racial and ethnic groups. Quality control and alignment metrics demonstrated good quality with a mean alignment rate of 96.4%, 80.4% mean specificity for exons, 17,368 mean number of genes detected, and a mean sequencing depth of 14.5 million uniquely mapping paired-end reads. One sample failed due to low exonic rate (17.7%) and 13 samples failed due to low sequencing depth (<5.5). Consensus molecular subtypes (CMS) were called among 74 patients. We observed 11 patients (14.8%) in CMS1, 30 patients (40.5%) in CMS2, 9 patients (12.2%) in CMS3, 7 patients (9.5%) in CMS4, and 17 patients (23.0%) in mixed types. We calculated a T cell-inflamed gene expression profile (GEP) score, a biomarker predicting response to checkpoint inhibitors and CRC-specific mortality. We observed differences in median GEP scores between populations and tumors from African American and Alaska Native patients having the highest GEP scores. Those preliminary results were limited by the small sample size and small proportion of lethal CRC outcomes (16.1%). We expect to have over 300 samples completed by summer 2022 and further analyses will be conducted then. With the inclusion of patients from more diverse populations, our study will help advance understanding of differences and similarities in the molecular and cellular landscape of CRC across racial and ethnic groups. This may also enhance the identification of novel clinically useful predictors of lethal CRC and potential novel therapeutic targets that could meaningfully reduce long-standing CRC disparities.

Cancer Posters - Wednesday

PB1083. Dissecting admixture effects on a 313-variant polygenic risk score model for breast cancer in a census-based cohort of Brazilians.

Authors:

T. Almeida¹, R. Barreiro², A. Farias¹, G. Souza¹, G. Tunes¹, E. Duim¹, C. Gomes³, M. Frederico², M. Farias¹, E. Amaro Jr.¹, B. Domingues Bitarello⁴, H. Brentani², M. Naslavsky⁵; ¹Hosp. Israelita Albert Einstein, São Paulo, Brazil, ²Hosp. Israelita Albert Einstein, Univ.e de São Paulo, São Paulo, Brazil, ³Univ.e de São Paulo, São Paulo, Brazil, ⁴Bryn Mawr Coll., Bryn Mawr, PA, ⁵Hosp. Israelita Albert Einstein, Univ. of Sao Paulo, Sao Paulo, Brazil

Abstract Body:

Complex disorders like breast cancer are described by a heritable polygenic component, environmental factors, and their interactions. Polygenic risk scores (PRS) partially represent the heritability captured by GWASs, but PRS transferability across populations is hampered by a complex interaction of population-specific factors, ultimately leading to differences in variant effects: demographic history, linkage disequilibrium (LD), variant frequency (AF-diff), presence of rare variants, epistasis, and GxE. This study aims to assess the transferability of a 313-SNP PRS (PRS₃₁₃) for breast cancer derived from a European ancestry (Eur) cohort in a census-based tri-hybrid - Eur, African (Afr), Native American - admixed sample of whole-genomes from Brazil (SABE1171, n=1171, abraom.ib.usp.br). We computed PRS₃₁₃ in 750 women (21 cases + 729 controls) in SABE1171 and used the UK Biobank (UKBB) as a gold-standard comparison. We find that the overall PRS distribution for SABE1171 was increased compared to the UKBB. Still, our area under the receiver operator curve (AUC) of 0.62 was equivalent to the UKBB-Eur AUC (0.63). To understand the influence of our small number of cases on the observed AUC, we contrasted the SABE1171—UKBB difference between mean and median PRS₃₁₃ values for cases and controls to 5,000 bootstraps of equally sized cases and control sets from the UKBB. SABE1171 mean PRS₃₁₃ difference adheres to the bootstrap, but the median PRS₃₁₃ difference is inflated, suggesting a few Brazilian cases might disproportionately impact the AUC result. To understand the effects of LD and AF-diff, we compared ancestry-stratified - Eur, Afr, East Asian (Eas) - LD patterns of 400kb windows around each PRS₃₁₃ SNP inferred from the 1000 Genomes Project (1KGP) to the LD pattern observed in SABE1171. For each LD window surrounding a PRS₃₁₃ variant, we calculated principal components (PCs) of LD values and the AF-diff between SABE1171 and each of the 1KGP ancestry groups. Based on the LD PCs and AF-diff, we conclude that SABE1171 cohort is more similar to Eur-1KGP than to either Afr- or Eas-1KGP, in agreement with the equivalent AUC results we obtained. Our preliminary findings indicate that using an LD and AF-diff informed PRS₃₁₃ for the SABE1171 cohort might be beneficial but will probably not suffice to achieve optimal prediction values. Alternatively, using effect sizes derived from admixed cohorts might yield better results. In brief, our study highlights the limitations of currently available PRSs when used outside of their respective discovery cohorts and showcases how deeply genotyped admixed populations can help understand PRS properties in admixed populations.

Cancer Posters - Thursday

PB1084. Does non-coding variation in known breast cancer susceptibility genes contribute to breast cancer risk?

Authors:

M. Akamandisa, B. Wubbenhorst, K. D'Andrea, H. Symecko, R. Gillette, P. Shah, A. Bradbury, K. Maxwell, S. Domchek, K. Nathanson; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Breast cancer (BC) is the most common malignancy and leading cause of cancer mortality in women. Average lifetime BC risk for a woman in the United States is 13%. However, pathogenic variants in BC susceptibility genes such as *BRCA1/2* and *PALB2*, or high polygenic risk scores substantially increase BC risk. Women with a family history of >2 1st or 2nd degree relatives with BC and/or ovarian cancer < age 60, BC < age 40, or BC and second primary tumor are considered high-risk and have greatly increased likelihood of having a pathogenic variant in a BC susceptibility gene. Yet, 70% of women with high-risk BC test negative for coding variants. In these women, possible causal variation could be non-coding in deep intronic or regulatory regions (promoters and enhancers), or cryptic structural variants (SV). To determine whether non-coding variation contributes to BC in 515 high-risk women testing negative for pathogenic coding variants in BC genes, we performed deep intronic splicing analysis using SpliceAI. We identified 7 potentially causal splicing variants in 13 (2.5%) samples in *BRCA1*, *ATM*, *RAD51C*, and *BRIP1*, three within 10bps of the exon border and the rest deeper in the introns, up to 1970bps. One *BRCA1* (c.4986+6T>G) and one *ATM* (c.8418+5_8418+8delGTGA) variant are known pathogenic but had not been previously reported. *In vitro* analysis to assess the impact of the identified variants on splicing and SV evaluation are underway. We performed whole genome sequencing of 24 high risk women (25% EOBC, 79.2% family history under 60, 25% multiple primaries) with no known mutations. We did not identify any splicing variants using SpliceAI or SVs likely to affect protein structure using Manta. To evaluate regulatory variation, we identified rare (<0.1% population frequency), heterozygous promoter and enhancer variants, and bioinformatically selected candidates for functional validation using GREEN-DB. Eighteen candidate non-coding variants in *BRCA1*, *PALB2*, *BARD1*, *ATM*, *CHEK2*, *RAD51C*, and *STK11* were identified in 10 (41%) samples. The variants have predicted scores of less than -1 for promoter/enhancer activity by DeepSea, lie within the 90th percentile of predicted deleteriousness by deep learning noncoding functional predictors JARVIS and/or ncER, and disrupt transcription factor binding sites. The candidate regulatory variants will be functionally validated *in vitro*. Understanding the level of contribution of non-coding variation in known BC susceptibility genes is critically important going forward, as results impact sequencing offered to patients, and if positive, their medical management.

Cancer Posters - Wednesday

PB1085. Dosing-specific modes of resistance to quizartinib in individual acute myeloid leukemia cells exposed by ResolveOME combined genomics and transcriptomics chemistry

Authors:

S. Velivela¹, i. González¹, V. Weigman¹, G. Harton², J. West³, J. Zawistowski⁴; ¹BioSkryb Genomics, Chapel Hill, NC, ²BioSkryb Genomics, Durham, NC, ³BioSkryb, Inc., Durham, NC, ⁴BioSkryb Genomics, Inc., Durham, NC

Abstract Body:

Dose and dose scheduling influence mechanisms of resistance to targeted therapeutics in clinical trials at both genomic and transcriptomic levels. To ascertain interplay between these levels, we created a model of quizartinib resistance in an acute myeloid leukemia (AML) cell line harboring an internal tandem duplication (ITD) mutation in *FLT3* with continual near-IC₅₀ (2 nM) dosing, and a second dose-escalation model where an initial 200 pM dose was increased by 100 pM at weekly intervals until the growth rate of the cells was near that of the treatment-naïve parental cells. We hypothesized that in utilizing single cell analysis between models we would see distinct, single-nucleotide and copy number changes, in conjunction with transcriptional adaptation and therefore employed ResolveOME to concomitantly assess DNA and RNA modulation. Copy number variation (CNV) analysis with BaseJumper showed that treatment-naïve parental single cells harbored Chr.5, 6 & 13 trisomies and pentasomy of Chr.8, consistent with karyotypic analysis. In the continual dosing model, resistance was correlated with gain of Chr. 19q and loss of Chr.5 trisomy to 2n level, whereby these modulations were heterogenous in single cells. By contrast, the dose-escalation model revealed Chr. 1q gain and Chr. 2p loss and ploidy reduction of Chr. 8 to 2n level in a fraction of the individual cells. At the single nucleotide variation (SNV) level, we identified a secondary mutation N841K in the drug target *FLT3*, present in all resistant cells of the continual dosing model and previously found in AML patients. We identified *FLT3* N841K only in a subset of dose-escalation single cells, and, intriguingly, the single cells that harbor this missense mutation lack the CNV specific to that model. This suggests differential mechanisms of resistance among single cells of a given treatment model, whereby N841K may be sufficient to drive resistance in isolation, but in other cells copy number alterations acquired during dosage increases were a primary mode of resistance. In addition to these divergent CNV paradigms we have defined AXL pathway bypass of FLT3 signaling inhibition via *GAS6* upregulation in resistant cells of the continual-dosing model and contrast the magnitude of this adaptation between single dose-escalation single cells. These data highlight distinct and heterogeneous modes of single-cell drug resistance depending on the nature and duration of the dosing and spotlight the necessity of joint genomic and transcriptomic information to comprehensively elucidate drug resistance mechanisms.

Cancer Posters - Thursday

PB1086. Dynamic play between human N a-acetyltransferase D and H4-mutant histones: Molecular dynamic study

Authors:

K. Srivastava, S. Rathod; CSIR-Central Drug Res. Inst., Lucknow, India

Abstract Body:

N-terminal acetyltransferases (NATs) are overexpressed in various cancers. Specifically in lung cancer, human N- α -acetyltransferase D (hNatD) is upregulated and prevents the histone H4 N-terminal serine phosphorylation, leading to the Epithelial-to-mesenchymal transition (EMT) of cancer cells. hNatD facilitates histone H4 N- α -terminal serine acetylation and halts the CK2 α -mediated serine phosphorylation. In the present study, we report the effects of four N-terminal mutant (S1C, R3C, G4D and G4S) histone H4 peptides on their bindings with hNatD by employing a molecular dynamics simulation. We also used graph theory-based analyses to understand residue correlation and communication in hNatD under the influence of WT and MT H4 peptides. Results show that S1C, R3C and G4S mutant peptides have significant stability at the catalytic site of hNatD. However, S1C, G4D and G4S peptides disrupt hNatD structure. Additionally, intramolecular hydrogen bond analysis reveals greater stability of hNatD in complex with R3C peptide. Further, intermolecular hydrogen bond analysis of acetyl-CoA with hNatD and its RMSD analysis in five complexes indicate that cofactor has greater stability in WT and R3C complexes. Our findings support previously reported experimental study on impacts of H4 mutations on its hNatD-mediated acetylation catalytic efficiency. The Betweenness centrality (BC) analysis further gives insight into the hNatD residue communication dynamics that can be exploited to target hNatD using existed or novel drug candidates therapeutically.

Cancer Posters - Wednesday

PB1087. EagleC: A deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps.

Authors:

X. Wang, **Y. Luan**, F. Yue; Northwestern Univ., Chicago, IL

Abstract Body:

Hi-C technique has been shown to be a promising method to detect structural variations (SVs) in human genomes. However, algorithms that can use Hi-C data for a full-range SV detection have been severely lacking. Current methods can only identify inter-chromosomal translocations and long-range intra-chromosomal SVs (>1Mb) at less-than-optimal resolution. Therefore, we develop EagleC, a framework that combines deep-learning and ensemble-learning strategies to predict a full-range of SVs at high-resolution. Importantly, we show that EagleC can uniquely capture a set of fusion genes that are missed by WGS or nanopore. Furthermore, EagleC also effectively captures SVs in other chromatin interaction platforms, such as HiChIP, ChIA-PET, and capture Hi-C. We apply EagleC in over 100 cancer cell lines and primary tumors, and identify a valuable set of high-quality SVs. Finally, we demonstrate that EagleC can be applied to single-cell Hi-C and used to study the SV heterogeneity in primary tumors.

Cancer Posters - Thursday

PB1088. Efficacy of universal testing for hereditary cancer syndromes: A community based cancer center (CBCC) experience.

Authors:

A. Hamblett¹, E. Molle¹, C. Van Der Walt², K. Kunze², M. Golafshar², A. Malon¹, E. Edward³, R. Nussbaum³, J. Samadder², J. Drew¹; ¹Middlesex Hlth., Middletown, CT, ²Mayo Clinic, Scottsdale, AZ, ³Invitae, San Francisco, CA

Abstract Body:

Background: Identification of pathogenic germline variants (PGVs) can inform cancer treatment, identify family members at increased cancer risk, and alter cancer screening/risk reduction recommendations. Historically, genetic testing has been pursued for patients who meet published guidelines based on age at cancer diagnosis, tumor characteristics, or family history. Research has shown that guideline directed germline testing misses 50% of PGVs in cancer patients at academic medical centers, but 85% of cancer patients receive care in community settings. A knowledge gap exists as to the ability of guidelines to predict PGVs in cancer patients seen outside of academic medical centers. The study aimed to evaluate whether using existing guidelines is sufficient for identifying patients with PGVs in a CBCC.

Methods: We completed a prospective study of germline genetic testing among patients with solid tumor malignancies aged 18-75 at a CBCC between December 2019 - May 2022. The patients were unselected for cancer type, stage, family history, and race/ethnicity. PGVs were identified using an 84-gene panel and were stratified by penetrance. National Comprehensive Cancer Networks (NCCN) 2020/2021 guidelines were used to determine incremental PGV rate which describes those patients with PGVs who did not meet NCCN guidelines or had PGVs that were not consistent with their history.

Results: 269 patients were enrolled, majority were female (56.5%), white (90.0%) with a median age of 63.0 years. The cohort was composed primarily of breast (27.9%), prostate (24.2%), and lung (12.3%) cancer patients. A total of 46 PGVs were identified in 44 of 269 (16.4%) patients, with 41.3% of PGVs stratified as high or moderate penetrance. Of those with detected PGVs, 56.8% did not meet NCCN guidelines based on cancer diagnosis or family history. Of those with detected PGVs, those with variants of uncertain significance (VUS), and those with negative results, 43.2%, 51.8%, 52.3% respectively met NCCN guidelines. Although not statistically significant ($p > 0.05$), this shows an 8-9% lower rate of meeting NCCN guidelines for those with PGVs. The incremental PGV rate was 67.4%.

Conclusions: This study found that patients with PGVs were not more likely to meet guidelines than negative or VUS patients, indicating that NCCN guidelines were not more efficient in detecting PGVs. A large percentage of patients had PGVs which would have been missed based on NCCN guidelines or recommended cancer-specific testing panels, suggesting that guideline directed testing was not effective in identifying cancer patients with PGVs in a CBCC.

Cancer Posters - Wednesday

PB1089. EG VEGF in ovarian cancer and preeclampsia as an early diagnostic marker: toward a targeted therapy in the treatment of tumors, especially those resistant to chemotherapy.

Authors:

M. Benfateh, T. Abouessaouira; Univ. hassan ii, Casablanca, Morocco

Abstract Body:

EG VEGF is an angiogenic factor whose biological activity is mediated via two receptors coupled to the G protein, the prokineticin receptor (PROKR1) and (PROKR2). The expression of this factor in the placenta and the female reproductive system opens perspectives of its implication and that of its receptors in many processes other than placentation. This will allow establishing its value as a predictive marker in the occurrence of many pregnancy pathologies, such as preeclampsia, but also in female cancer pathologies, such as ovarian cancer. The aim of this work is to study was first to identify the epidemiological profile of the risk factors for the occurrence of these two pathologies and secondly, to characterize the role of EG VEGF and its receptors PROKR1 / PROKR2 in epithelial ovarian cancer. The thesis project was based on an epidemiological and a molecular study. The study is a retrospective descriptive one, that was undertaken over a period of 2 years, during which we collected the cases of preeclampsia and ovarian cancer, managed in the department of Gynecology-Obstetrics "C" Service and the oncology center of Ibn Rochd University Hospital of Casablanca. The molecular study is a prospective, control cases, spread over four years, during which serum assays and an immunohistochemical studies, were carried out on patients with ovarian cancer. Our results showed that the frequency of preeclampsia was about (7.1%). The epidemiological profile was represented by primiparous woman, with average age 30 years that have not been followed up on their pregnancy. The risk factors frequently found are obesity and chronic hypertension associated with vascular-renal history; abortion and perinatal death associated with history of obstetrics. For ovarian cancer, the prevalence is 2.7%, the average age of patients is 51.47 ± 13 years, and the significant risk factors observed are multiparity, menopause as a personal history and hypertension as a medical history. Cystic-looking tumors, on the right side, cystadeno-carcinoma constituted the most frequent histological type. 74.8% cases of metastasis were reported, including 44.3% of peritoneal carcinosis. The molecular showed a potential involvement of the EG-VEGF / PROKR system in the development and progression of the epithelial ovarian cancer.

The determination of the role of this pro-angiogenic and pro-inflammatory factor could constitute an early diagnostic marker of these pathologies, along with the possibility to constitute a targeted therapy in the treatment of tumors, especially those resistant to chemotherapy.

Cancer Posters - Thursday

PB1090. Eligibility, uptake and response to germline genetic testing in women with DCIS

Authors:

R. Ellsworth¹, L. Turner², L. Lovejoy³, C. Turner⁴, C. Shriver⁵; ¹Murtha Cancer Ctr., Bethesda, MD, ²Anne Arundel Med. Ctr., Annapolis, MD, ³Windber Res. Inst., Windber, PA, ⁴NIH, Bethesda, MD, ⁵Uniformed Services Univ. of the Hlth.Sci. and Walter Reed Natl. Military Med. Ctr., Bethesda, MD

Abstract Body:

Background Ductal carcinoma *in situ* (DCIS) is both a malignant, yet pre-invasive disease of the breast. While the majority of DCIS have a low risk of recurrence, a subset of women with germline pathogenic variants (PV) in cancer predisposition genes are at increased risk for breast cancer recurrence. Uptake of genetic testing and subsequent surgical intervention in women with DCIS has not been well-studied. The aim of this study was to evaluate test eligibility, uptake and impact on surgical decision making in women with DCIS. **Methods** All women diagnosed with unilateral DCIS 2001-2020 enrolled in the Clinical Breast Care Project were identified. Demographic, commercial genetic test results and surgical procedures were extracted from the database. Test-eligibility was assigned using National Comprehensive Cancer Network (NCCN) criteria. Panel genetic testing was performed in the research setting using DNA from 465 women. Statistical analyses were performed using Fisher's exact tests and Chi-square analyses with $p < 0.05$ defining significance. **Results** Forty-two percent of women were test-eligible at diagnosis with an additional 12.8% meeting eligibility criteria >1 year after diagnosis. Within the test-eligible group, 34.5% pursued clinical genetic testing and 17.3% pursued testing only after a second cancer event. Of the 39 women with PV, 20 (51.3%) were detected only in the research setting, with 11 (28.2%) of these women not eligible for genetic testing based on NCCN criteria. In women who did not undergo BM at diagnosis, recurrence was significantly higher ($p=0.001$) in women with PV (33.3%) compared to those without PV (11.8%).

Conclusion Pursuit of genetic testing and subsequent use of risk-reducing surgeries in women with PV was suboptimal in women with a primary diagnosis of DCIS. In conjunction, >50% of PV were detected only in the research setting. Because omission of genetic testing in women with DCIS may represent a lost opportunity for prevention, genetic testing at the time of diagnosis should be standard for all women with DCIS. **Disclaimer** The contents of this publication are the sole responsibility of the author(s) and do not necessarily reflect the views, opinions or policies of Uniformed Services University of the Health Sciences (USUHS), The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., the Department of Defense (DoD), the Departments of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Cancer Posters - Wednesday

PB1091. Elongin C (*ELOC/TCEB1*) gene: a novel cause of von Hippel-Lindau disease

Authors:

A. Andreou¹, B. Yngvadottir¹, L. Bassaganyas¹, G. Clark¹, E. Martin¹, J. Whitworth², A. J. Cornish³, Genomics England Research Consortium, R. Houlston⁴, P. Rich⁵, C. Egan⁶, S. V. Hodgson⁷, A. Y. Warren⁸, K. Snape⁹, E. Maher²; ¹Univ. of Cambridge, Cambridge, United Kingdom, ²Univ Cambridge, Cambridge, United Kingdom, ³Div. of Genetics and Epidemiology, The Inst. of Cancer Res., London, United Kingdom, ⁴The Inst. of Cancer Res., Sutton, United Kingdom, ⁵Dept. of Neuroradiology, St George's Univ. Hosp. NHS Fndn. Trust, London, United Kingdom, ⁶Moorfields Eye Hosp. NHS Fndn. Trust, London, United Kingdom, ⁷St George's Hosp Med. Sch, London, United Kingdom, ⁸Dept. of Histopathology, Cambridge Univ. NHS Fndn. Trust, Cambridge, United Kingdom, ⁹St George's Hosp., London, United Kingdom

Abstract Body:

Background: Around 95% of patients with clinical features diagnostic of Von Hippel-Lindau disease (VHL) have a detectable inactivating germline variant in *VHL*. The VHL protein (pVHL) functions as part of the VCB-CR complex which plays a key role in oxygen sensing and degradation of hypoxia inducible factors. To date, only variants in *VHL* have been shown to cause VHL disease. **Materials and Methods:** We undertook trio analysis by Whole-exome sequencing (WES) in a proband with VHL disease but without a detectable *VHL* mutation. Molecular studies were also performed on paired DNA extracted from the proband's kidney tumour and blood and bioinformatics analysis of sporadic renal cell carcinoma data set was undertaken. **Results:** A *de novo* pathogenic variant in *ELOC* (NM_005648.4:c.236A>G [p.Tyr79Cys]) gene was identified in the proband. *ELOC* encodes elongin C, a key component [C] of the VCB-CR complex. The p.Tyr79Cys substitution is a mutational hotspot in sporadic VHL-competent renal cell carcinoma (RCC) and has previously been shown to mimic the effects of pVHL deficiency on hypoxic signalling. Analysis of a RCC from the proband showed similar findings to that in somatically *ELOC* mutated RCC (expression of hypoxia responsive proteins, no somatic *VHL* variants and chromosome 8 loss). **Conclusions:** These findings are consistent with pathogenic *ELOC* variants being a novel cause for VHL disease and suggest that genetic testing for *ELOC* variants should be performed in individuals with suspected VHL disease with no detectable *VHL* variant.

Cancer Posters - Thursday

PB1092*. Enhanced expression of mutant alleles of cancer driving genes in pan-adenocarcinoma

Authors:

C. Lo^{1,2}, N. Fujito¹, H. Nakaoka³, K. Hosomichi⁴, I. Inoue¹; ¹Human Genetics Lab., Natl. Inst. of Genetics, Mishima City, Shizuoka Prefecture, Japan, ²Dept. of genetics, Sch. of Life Sci., The Graduate Univ. for Advanced Studies (SOKENDAI), Kanagawa Prefecture, Japan, ³Dept. of Cancer Genome Res., Sasaki Inst., Tokyo, Japan, ⁴Dept. of Bioinformatics and Genomics, Kanazawa Univ., Kanazawa Prefecture, Japan

Abstract Body:

Among cancer driver genes, the effects of oncogenesis or the impacts of insufficiency in tumor suppression can be intensified by increasing the expression of mutant alleles through the observed **allelic imbalance in transcriptional efficiency (AITE)** effect. And, this is in contrast to those mutated passenger genes, whose normal alleles are instead increased in the expression for buffering the mutant proteins. We have analyzed large-scale pan-adenocarcinoma samples from The Cancer Genome Atlas (TCGA) program (n=3256). The genotypes and allele specific copy numbers inside each gene were firstly checked by SNP array and whole exome sequencing data (WESeq). Then, based on the paired RNA sequencing data (RNASeq), the AITE effects were screened for those heterozygous positions without existing allele specific somatic copy number alterations (SCNA). Among those somatic mutations, most nonsynonymous SNV mutations of well-known **cancer driver genes biased their expression toward the mutant allele** in statistically significant ways, where the average mutant allele frequency (MAF) in RNASeq was significantly **higher** than that in WESeq (Welch's t-test, $p = 1.63 \times 10^{-5}$ for *TP53*; $p = 1.36 \times 10^{-14}$ for *KRAS*; $p = 9.75 \times 10^{-4}$ for *CTNNB1*; $p = 3.80 \times 10^{-3}$ for *PTEN*). In contrast, the nonsynonymous SNV mutations of **passenger genes biased their expression toward the normal (reference) allele**, where the average MAF in RNASeq was significantly **lower** than that in WESeq (Welch' t-test, $p = 2.33 \times 10^{-20}$ for *COL3A1*; $p = 1.52 \times 10^{-5}$ for *CDH11*; $p = 5.82 \times 10^{-23}$ for *MYH11*; $p = 2.59 \times 10^{-5}$ for *PDGFRB*; $p = 4.41 \times 10^{-3}$ for *TNC*; $p = 1.01 \times 10^{-12}$ for *PTPRC*). Especially, those AITE genes with increased mutant allele expression could mostly be found with a few sites which equipped relatively high mutation frequency, belonged to positive selection signals. Overall, the tumors showed the significant higher AITE occurring rates in their genomes when being compared to the paired-solid-normal tissues (Welch's t-test, $p < 2.2 \times 10^{-16}$, n=152). By further comparing their whole genome sequencing data (WGSeq), more *de novo* transposable elements (TE) insertion sites were also detected inside tumors than that inside normal tissues with significance (Welch's t-test, $p=0.0456$, n=27). Within 1 Mbp around of AITE genes, tumor-specific *de novo* TE insertions were discovered to be associated with AITE occurrence.

Cancer Posters - Thursday

PB1093. Estimating total tumor-specific microRNA content in human tissues using computational deconvolution

Authors:

M. Montierth^{1,2}, G. A. Calin², W. Wang²; ¹Baylor Coll. of Med., Houston, TX, ²Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract Body:

Transcriptomic change is a hallmark of cancer, and thus quantification and characterization of these changes is a key feature of modern clinical oncology. Recently, it has been shown that a deconvolution-based estimate of total mRNA expression per tumor cell (TmS) varies across and within cancer types and is a predictor of patient survival outcomes. It stands to reason that variation in total tumor cell expression of other RNA species may further capture important biological changes. MicroRNAs (miRNAs) are one of the many RNA species presented within a cell and are important post-transcriptional regulators with well-documented diverse roles in cancer, from initiation to progression and metastasis. Methods for studying tumor cell specific miRNAs have lagged behind mRNAs. Single cell miRNA studies are essentially nonexistent, and many existing transcriptomic computational deconvolution methods are inadequate to study miRNAs. To address these needs, we adapted DeMixT, a semi-unsupervised transcriptomic deconvolution method, to miRNA, and we introduce a deconvolution-based estimate of total tumor cell miRNA content (TmiS). Using matched DNaseq/miRNAseq data for 4,167 patient samples across 14 cancer types from The Cancer Genome Atlas (TCGA), we quantified tumor-specific total miRNA expression, using tumor miRNA proportion, which is estimated from deconvolution of bulk miRNAseq data, and further adjusting for tumor purity and ploidy, which are estimable through genomic deconvolution. We found that TmiS varies across cancer types and differs between relevant clinical subtypes within cancer type. We observed that generally across cancers, high TmiS is associated with increased risk of disease progression and death, after adjusting for stage and age. We replicated the associations between TmiS and breast cancer subtypes that we observed in TCGA-BRCA in 1,293 patient samples from the Molecular Taxonomy of Breast Cancer Consortium (METABRIC). Taken together, we find evidence for TmiS as a potential pan-cancer biomarker and prognostic indicator

Cancer Posters - Wednesday

PB1094. Evaluating biomarker potential of germline genomic factors for predicting clinical outcomes in prostate cancer.

Authors:

N. Zeltser¹, K. E. Houlahan¹, S. M. Al-Hiyari¹, S. E. Eng¹, Y. Patel¹, T. N. Yamaguchi¹, S. Tao¹, R-R. Huang¹, R. E. Reiter¹, H. Ye¹, A. S. Kinnaird^{1,2}, P. C. Boutros¹; ¹Univ. of California, Los Angeles, CA, ²Univ. of Alberta, Edmonton, AB

Abstract Body:

Prostate cancer is the second-most diagnosed cancer and the second leading cause of cancer death in American men. Early detection is common, but is followed by the more challenging task of prognosing a highly variable clinical course. Current clinical risk-assessment strategies such as serum abundance of prostate specific antigen (PSA), tumor size and extent, and tumor grade based on biopsy are highly imprecise: over a third of patients are over-treated. An improved method of risk stratification may lie in hereditary factors. Prostate cancer is one of the most strongly inherited ($h^2 = 57\%$), with accumulating evidence associating rare variants, common variants, and genetic ancestry to clinical outcomes. We have performed germline sequencing on blood from thousands of patients diagnosed with localized prostate cancer and with extensive follow-up data. We quantify the interactions of rare and common variants, and demonstrate that germline features provide insights into patient outcomes and optimal management strategies.

Cancer Posters - Thursday

PB1095. Evaluation of miR-21 and miR-938 expression in the progression of gastric cancer in the Magellanic Chilean cohort - MAGIC.

Authors:

D. Zapata-Contreras^{1,2}, A. Altamirano³, S. Karelavic³, C. Urrea³, F. Orellana³, C. Delgado³, M. Iriarte³, M. Puente³, L. Leiva³, L. Godoy³, O. Gallardo¹, P. Zúñiga^{1,2}, Y. Espinosa-Parrilla^{1,2,4}; ¹Med. and Evolutionary Genomics in Magallanes (GEMMa), Ctr. for Ed., Hlth.care and Investigation (CADI-UMAG), Punta Arenas, Chile, ²Sch. of Med. – Univ. of Magallanes, Punta Arenas, Chile, ³Regional Clinical Hosp. of Magallanes, Punta Arenas, Chile, ⁴InterUniv. Ctr. for Hlth.y Aging (CIES), Punta Arenas, Chile

Abstract Body:

Gastric Cancer is a leading cause of global cancer morbidity and mortality, being the third-leading cause of cancer death worldwide. Its clinical outcome is highly dependent on local progression and metastasis development, which is partially controlled by oncogenes and tumour suppressor genes that are further regulated by microRNAs, small endogenous regulatory RNAs involved in the control of multiple biological pathways. Aberrant microRNA expression in cancer has been found to correlate with the diagnosis and stage of many cancer types including gastric cancer and, due to its high stability, microRNAs have emerged as potential effective prognostic and diagnostic biomarkers. In view of the high mortality rates of gastric cancer in Latin America, it is necessary to find specific biomarkers to trace its diagnosis and progression for these specific populations that have not yet been studied and have a different genetic pool, as the Chilean population. For that purpose, we recruited a cohort (MAGIC) in the extreme south of Chile, Magallanes, of more than 500 individuals with suspected gastric disease through endoscopy from 2018 to date. A liquid biopsy, together with tissue from biopsies and gastrectomies (if the procedure was necessary for diagnosis/recovery) were obtained from patients at different stages of gastric cancer progression. As a first approach, a comparative expression analysis of two microRNAs involved in gastric cancer was performed for their validation in the MAGIC cohort. The microRNAs evaluated were miR-21, which has been reported to be upregulated in gastric cancer in other populations, and miR-938 for which we previously reported a genetic association with gastric cancer probably through deregulation of the *CXCL12* axis. The expression analysis by qRT-PCR in tumoral and adjacent normal tissues of patients at different stages of the disease showed an increase of 0,5 in the expression of miR-21 in advanced tumoral tissues compared with normal tissues and of 2 in the expression of miR-938 in early tumoral tissues compared with normal and advanced tumoral tissues (Tukey's Anova test, $p < 0.05$). The expression analysis of these two miRNAs and their target genes is now under further investigation in the MAGIC cohort. These results are consistent with the recognized value of miR-21 as a biomarker of gastric cancer in other populations and point towards a potential significance of miR-938 as a diagnostic biomarker of early gastric cancer.

Cancer Posters - Wednesday

PB1096. Evaluation of non-coding regulatory variants in susceptibility of Chordoma

Authors:

S. Yepes Torres¹, J. Bai², C. Li², S. Gui², W. Luo¹, J. Liu¹, H. Koka¹, B. Hicks¹, A. Goldstein¹, R. Yang¹; ¹Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Bethesda, MD, ²Beijing Tiantan Hosp., China, China

Abstract Body:

Chordoma is a rare bone cancer with etiologic factors largely unknown. Brachyury, which is encoded by *TBXT*, a transcription factor critical for notochord develop, is essential for chordoma pathogenesis. We previously identified *TBXT* as a major chordoma susceptibility gene. However, genetic causes in the majority of familial and sporadic chordoma patients remain largely unknown. The role of non-coding variants in chordoma susceptibility has not been explored. Computationally predicting non-coding regulatory variants is important because of the high volume of causal candidate variants in regulatory regions and the challenges of functional validation. The goal of this study is to identify potentially functional non-coding regulatory variants in a previously described 1.5-Mb *TBXT* super-enhancer associated region and the 5p15.33 *TERT* promoter region, which is a well-characterized regulatory region that contains genetic variants associated with multiple cancer types, in a Whole-Genome Sequencing (WGS) dataset of 80 Chinese patients with skull-base chordoma. Prediction scores based on different assumptions of causality and complementary statistical models from aggregated scores and ensemble prediction models from the comprehensive resource regBase, which integrates non-coding variant prediction scores from 23 tools (Zhang S, Nuc Acid Res, 2019), were applied for variant prioritization. We also used the GWAS catalog to incorporate significant disease-associated SNPs (P-value < 5E-8) at the *TERT* region from 10 GWAS fine-mapping studies that confer risk of multiple cancers to investigate those variants in the chordoma dataset. The analyses identified several score peaks, with approximately 19 score peaks colocalized with significant disease-associated variants identified from the *TERT* promoter region, e.g., rs13172201, rs10069690, and rs2853669. Among prioritized variants identified by regBase models, rs2853669 showed one of the most significant scores. This SNP was previously validated and shown to disrupt the *TERT* promoter and confer cancer risk by extensive functional experiments. Potential variants upstream of *TBXT*, such as g.166478324 (G>A), had the largest cancer driver potential prediction score and warrant further investigation. Analyzing the score spectrum of the ensemble prediction models applied across the studied regions has facilitated the identification of non-coding regulatory variants with functional, pathogenic, and driver potential that may contribute to chordoma susceptibility. Future investigations include expanding candidate gene regions and replicating top variants in independent datasets.

Cancer Posters - Thursday

PB1097. Evidences for polygenic inheritance in familial cancer

Authors:

M. Lista^{1,2}, M. Baldassarri^{1,2,3}, M. Mencarelli³, D. Maffeo^{1,2}, E. Pasquinelli^{1,2}, M. Carullo^{1,2}, L. Adamo^{1,2}, E. Antolini^{1,2}, F. Fava^{1,2,3}, K. Zguro², S. Furini², G. Doddato^{1,2}, R. Tita³, M. Bruttini^{1,2,3}, A. Fabbiani^{1,2,3}, C. Lo Rizzo³, A. Pinto³, A. Carrer^{1,2,3}, L. Loberti^{1,2,3}, P. Ilardi^{1,2,3}, M. Manara^{1,2,3}, F. Mari^{1,2,3}, F. Ariani^{1,2,3}, A. Renieri^{1,2,3}; ¹Med. Genetics, Univ. of Siena, Siena, Italy, ²Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Italy, ³Genetica Medica, Azienda Ospedaliera Univ.ria Senese, Siena, Italy

Abstract Body:

The majority of cancer driver genes have moderate to low penetrance. However, the molecular bases of incomplete penetrance have still not been fully understood yet. We wanted to explore the model of transmission by performing Exome Analysis (EA) in 40 families in which the DNA of at least two affected members was available. In a relevant percentage of cases (30%) the pathogenic variant (P) identified in one cancer gene did not segregate with the disease in the family. We found another variant of unknown significance (VUS) in another cancer driver gene segregating with the disease. Relevant cases will be presented, including one familial melanoma with not segregating *P/RAD50* and segregating *VUS/PMS2*, one familial breast cancer with not segregating *P/ERCC2* and segregating *VUS/ERCC5*, one familial colon cancer with not segregating *P/MSH6* and segregating *VUS/MSH3*. We then explored additional cases (80) in which only one affected member was available. In a significant fraction of patients (10%) we identified P variants in 2 different cancer driver genes. Combinations among others were: *P/ATM/P/MSH3* in pancreatic cancer, *P/SMAD9/P/SEC23B* in breast cancer, *P/RAD50/P/PRKN* in melanoma, *P/SBDS/P/BRCA2* in prostate cancer, *P/RBI/P/ERCC3* in melanoma, *P/BRCA1/P/SLC26A3* in breast cancer, *PMS2-deletion/P/ERCC4* in colon cancer. Finally, we explored if the number of rare P-VUS (MAF<0,01) in cancer driver genes of cancer patients (30 patients) differs from that identified in non-cancer subjects (50 subjects). We identified a mean of 2,8 P-VUS in patients and a mean of 1,6 P-VUS in controls with a p-value between them of 0,0008 obtained with Mann-Whitney U Test. In summary, these data can give us the evidence that cancer may have a polygenic inheritance and that also familial cancer genetic susceptibility, could be the result of more than one variant in cancer driver genes.

Cancer Posters - Wednesday

PB1098. Exome-wide discovery of the contribution of coding variants identifies novel genes associated with breast cancer risk.

Authors:

N. Wilcox¹, M. Dumont², A. González-Neira³, S. Carvalho¹, C. Beauparlant², M. Crotti⁴, C. Luccarini¹, S. Dubois², G. Pita³, J. Castela⁵, J. Chang-Claude^{6,7}, J. Dennis¹, T. Dork⁸, C. Engel^{9,10}, M. Gago-Dominguez^{11,12}, P. Guénel¹³, E. Hahnen^{14,15}, M. Hartman^{16,17,18}, J. Li¹⁹, K. Michailidou^{1,20}, S. Teo^{21,22}, M. Schmidt^{23,24}, R. Schmutzler^{14,15,25}, P. Pharoah^{1,4}, A. Droit^{2,26}, A. Dunning⁴, A. Kvist²⁷, P. Devilee^{28,29}, D. Easton^{1,4}, J. Simard²; ¹Ctr. for Cancer Genetic Epidemiology, Dept. of Publ. Hlth. and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, ²Genomics Ctr., Ctr. Hosp.ier Univ.ire de Québec – Université Laval Res. Ctr., Québec, QC, Canada, ³Human Genotyping Unit-CeGen, Human Cancer Genetics Programme, Spanish Natl. Cancer Res. Ctr. (CNIO), Madrid, Spain, ⁴Ctr. for Cancer Genetic Epidemiology, Dept. of Oncology, Univ. of Cambridge, Cambridge, United Kingdom, ⁵Oncology and Genetics Unit, Inst. de Investigación Sanitaria Galicia Sur (IISGS), Xerencia de Xestión Integrada de Vigo-SERGAS, Vigo, Spain, ⁶Div. of Cancer Epidemiology, German Cancer Res. Ctr. (DKFZ), Heidelberg, Germany, ⁷Cancer Epidemiology Group, Univ. Cancer Ctr. Hamburg (UCCH), Univ. Med. Ctr. Hamburg-Eppendorf, Hamburg, Germany, ⁸Gynaecology Res. Unit, Hannover Med. Sch., Hannover, Germany, ⁹Inst. for Med. Informatics, Statistics and Epidemiology, Univ. of Leipzig, Leipzig, Germany, ¹⁰LIFE - Leipzig Res. Ctr. for Civilization Diseases, Univ. of Leipzig, Leipzig, Germany, ¹¹Genomic Med. Group, Intl. Cancer Genetics and Epidemiology Group, Fundación Pública Galega de Med. Xenómica, Inst. de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hosp.ario Univ.rio de Santiago, SERGAS, SERGAS, Spain, ¹²Moore's Cancer Ctr., Univ. of California San Diego, La Jolla, La Jolla, CA, ¹³Team 'Exposome and Heredity', CESP, Gustave Roussy, INSERM, Univ. Paris-Saclay, UVSQ, Villejuif, France, ¹⁴Ctr. for Familial Breast and Ovarian Cancer, Faculty of Med. and Univ. Hosp. Cologne, Univ. of Cologne, Cologne, Germany, ¹⁵Ctr. for Integrated Oncology (CIO), Faculty of Med. and Univ. Hosp. Cologne, Univ. of Cologne, Cologne, Germany, ¹⁶Saw Swee Hock Sch. of Publ. Hlth., Natl. Univ. of Singapore and Natl. Univ. Hlth. System, Singapore, Singapore, ¹⁷Dept. of Surgery, Natl. Univ. Hlth. System, Singapore, Singapore, ¹⁸Dept. of Pathology, Yong Loo Lin Sch. of Med., Natl. Univ. of Singapore, Singapore, Singapore, ¹⁹Human Genetics Div., Genome Inst. of Singapore, Singapore, Singapore, ²⁰Biostatistics Unit, The Cyprus Inst. of Neurology & Genetics, Nicosia, Cyprus, ²¹Breast Cancer Res. Programme, Cancer Res. Malaysia, Subang Jaya, Selangor, Malaysia, ²²Dept. of Surgery, Faculty of Med., Univ. of Malaya, UM Cancer Res. Inst., Kuala Lumpur, Malaysia, ²³Div. of Molecular Pathology, The Netherlands Cancer Inst., Amsterdam, Netherlands, ²⁴Div. of Psychosocial Res. and Epidemiology, The Netherlands Cancer Inst. - Antoni van Leeuwenhoek Hosp., Amsterdam, Netherlands, ²⁵Ctr. for Molecular Med. Cologne (CMMC), Faculty of Med. and Univ. Hosp. Cologne, Univ. of Cologne, Cologne, Germany, ²⁶Département de Médecine Moléculaire, Faculté de Médecine, Ctr. Hosp.ier Univ.ire de Québec Res. Ctr., Laval Univ., Québec, QC, Canada, ²⁷Div. of Oncology and Pathology, Dept. of Clinical Sci. Lund, Lund Univ., Lund, Sweden, ²⁸Dept. of Pathology, Leiden Univ. Med. Ctr., Leiden, Netherlands, ²⁹Dept. of Human Genetics, Leiden Univ. Med. Ctr., Leiden, Netherlands

Abstract Body:

Genetic susceptibility to breast cancer is known to be conferred by a combination of common variants identified through genome-wide association studies together with rarer coding variants with higher disease risks. The latter include protein-truncating variants (PTVs) and/or rare missense variants in *ATM*, *BARD1*, *BRCA1*, *BRCA2*, *CHEK2*, *RAD51C*, *RAD51D*, *PALB2* and *TP53*. However, these variants explain ~20% of the familial relative risk of breast cancer and the overall contribution of coding variation

to breast cancer is unclear. To evaluate the role of rare coding variants more comprehensively, we performed a meta-analysis across three large whole-exome sequencing datasets: 2 datasets from the Breast Cancer Association Consortium together with UK Biobank. These datasets comprised 16,498 cases and 182,142 controls primarily of European ancestry. Burden tests were performed for protein-truncating and rare missense variants in 16,562 and 18,681 genes respectively. To improve power, we incorporated data on family history of breast cancer. Associations between protein-truncating variants and breast cancer were identified for 7 genes at exome-wide significance ($P < 2.5 \times 10^{-6}$): the five known susceptibility genes *BRCA1*, *BRCA2*, *CHEK2*, *PALB2* and *ATM*, together with novel associations for *ATRIP* and *MAP3K1*. Predicted deleterious rare missense or protein-truncating variants were additionally associated at $P < 2.5 \times 10^{-6}$ for *SAMHD1*. All 3 novel genes have prior evidence of involvement in tumorigenesis or DNA repair. *MAP3K1* is a well-established breast-cancer GWAS locus; however, the observed PTV burden was independent of the GWAS associations. We used an empirical Bayes approach to estimate the overall exome-wide contribution of protein-truncating variants to the familial relative risk (FRR) of breast cancer: under the best fitting model, the contribution of rare PTVs beyond previously known genes explains ~2% of the FRR, with ~130 genes being associated. We demonstrate that large exome sequencing studies, combined with efficient burden analyses, can identify additional breast cancer susceptibility genes, and that further large-scale analyses are likely to identify further genes, but that most of the “missing” heritability is likely to be found in the non-coding genome.

Cancer Posters - Thursday

PB1099. Expanding the clinical and molecular spectrum of hereditary *TINF2* cancer predisposition syndrome reporting a family with Hodgkin Lymphoma

Authors:

E. Agolini¹, F. Locatelli¹, F. Berardinelli², A. Alesi¹, S. Russo¹, S. Genovese¹, A. Sgura², R. Amato², M. Matraxia¹, S. Petrocchi¹, D. Cocciadiferro¹, M. Lodi¹, K. Girardi¹, R. De Vito¹, A. Novelli¹, L. Vinti¹; ¹Ospedale Pediatrico Bambino Gesù, Rome, Italy, ²Università Roma Tre, Rome, Italy

Abstract Body:

TINF2 (TERF1 Interacting Nuclear Factor 2) gene encodes TIN2 protein, a critical subunit of the shelterin complex, which protects telomere ends from being recognized as damaged DNA sites and also serve to regulate telomere length maintaining chromosome stability. Pathogenic missense and truncating *TINF2* variants are causative for dyskeratosis congenita (DC), a rare bone marrow failure syndrome, inherited as autosomal dominant disorder, characterized by mucocutaneous abnormalities associated with very high risks of developing aplastic anemia, myelodysplastic syndrome, leukemia, and solid tumors. TIN2-DC variants localize to a 30 amino acid coding stretch in exon 6 called the 'DC cluster'. Recent reports show that specific *TINF2* truncating variants act as high penetrance cancer predisposition alleles outside DC context, with a relevant risk to develop different tumors including melanoma, thyroid and breast cancer. Based on these reports, *TINF2* acts as a haploinsufficient tumor suppressor resulting in excessive telomere elongation. Here, we describe a novel heterozygous loss-of-function germline variant c.248_249del (p.Leu83GlnfsTer53) in *TINF2* (NM_001099274.3) in a family with two young sisters affected by nodular sclerosing Hodgkin lymphoma without any of the DC symptoms. This variant was identified by family-based clinical exome sequencing and was inherited from the apparently healthy mother. Further studies on patient's lymphocytes to evaluate telomere length and additional cytogenomics analysis on the tumor DNA are currently in progress. In conclusion, this report expands the clinical and molecular spectrum of hereditary *TINF2* cancer predisposition syndrome, reporting for the first time an association between a novel deleterious variant of this gene and Hodgkin lymphoma risk.

Cancer Posters - Wednesday

PB1100. Exploring the non-invasive potential of miRNA 145 and miRNA 363 in Prostate Cancer

Authors:

A. Manoj, A. MAHDI, M. AHMAD, M. Kumar; King George's Med. Univ., Lucknow, India

Abstract Body:

Owing to the misdiagnosis and over-treatment of prostate cancer, due to its indolent growth and false-positive errors produced by diagnostic ways like PSA level, DRE, and biopsy. Consequently, there is an urgent need for reliable diagnostic and detection technologies to diagnose aggressive prostate cancer non-invasively by exploring the potential of circulating microRNAs. Efficient and reliable markers within the body fluids can help in personalized treatment decisions for monitoring disease and survival. MicroRNA (miRNA) are highly conserved small non-coding RNA that modulate genes involved in numerous biological processes and form part of complex networks that play a significant role in prostate cancer initiation and progression. In keeping this view, we have designed our study to investigate biomarker potential miRNA 145 and miRNA 363 in prostate cancer. We determined the comparative expression of miRNA 145 in benign prostatic hyperplasia (BPH) and prostate cancer blood samples by Real Time-PCR and performed the Receiver Operating Characteristics (ROC) curve analysis. Further, we explored target genes of miRNA using In-Silico tools like TargetScan, miRDB, and miRWalk3.0 and performed gene and pathway enrichment analysis using DAVID 6.8. Our observation suggests a significant tumor-suppressive role of miRNA 145 and the oncogenic role of miRNA 363 in prostate cancer compared to BPH. Computational analysis reveals that miRNA 145 and miRNA 363, both target genes like E2F3, NRAS, PTEN, FOXO1, etc., play a significant role in prostate cancer survival and progression. Finally, ROC curve analysis predicts that miRNA 145 (AUC 0.822***) and 363 (AUC 0.852***) have good biomarker potential for distinguishing BPH from prostate cancer. Collectively these results reveal that miRNA 145 and miRNA 363 expression may act as a non-invasive biomarker for effective diagnosis of Prostate Cancer. For future suggestions, the target genes evaluation may provide therapeutic targets for better treatment and management of prostate cancer.

Cancer Posters - Thursday

PB1101. Family History-Based Risk of BRCA Mutations in the Malian Populations

Authors:

B. Baba; USTTB, Bamako, Mali

Abstract Body:

In 2021, the US Preventive Services Task Force recommended that women at a high risk of deleterious BRCA mutations, based on family history, be referred by primary care providers for genetic counseling and evaluation for testing. It is estimated that 2% of U.S. women have a family history necessitating referral for genetic counseling. The estimated combined frequency of BRCA mutations in the U.S. is approximately 0.25%. We assessed the proportion of women in Mali at high risk of carrying a BRCA mutation based on family history. We analyzed data from the 2021 Mali Health Interview Survey (CHIS), a population-based telephone survey that collects state and local data on a range of health topics. The 2021 survey included a detailed cancer family history module. We estimated the probability the respondent carried a BRCA mutation using BRCAPRO. We imputed the current age of family members and their age at cancer diagnosis. We estimated the current age as a function of the respondent's age. We estimated the age of cancer diagnosis of family members who had cancer before age 50 years as a random variable from a uniform distribution with a lower limit of the estimated current age and an upper limit of 50 years. The minimum ages for breast cancer and ovarian cancer onset were set at 25 and 35 years, respectively. The minimum estimated probability was 0.0000014% and the maximum was 94%. The 98.5 percentile for the population was 0.63%. Therefore, <1.5% of the population had an estimated probability of BRCA mutations > 1.0%; <0.5% had an estimated probability of BRCA mutations > 10.0%. The CHIS data had some other limitations. The respondent's history of ovarian cancer was not available. No information was available on BRCA mutation testing, so the predictive value of the risk estimate could not be determined. CHIS asks respondents about breast cancer twice -during the regular questionnaire and in the family history module, but only asks for age of diagnosis in the family history module. Therefore, we had to estimate the age at diagnosis for women who reported breast cancer only in the regular questionnaire. The methods we used to estimate the age of cancer diagnosis are biased towards underestimating the age of cancer diagnosis, since cancer incidence increases with age. Since younger age at diagnosis is associated with an increased risk of having a BRCA mutation, we may have overestimated the probability of carrying a BRCA mutation.

Cancer Posters - Wednesday

PB1102. Filling gaps in whole genome analysis in hematology: A chance for optical mapping and long-read NGS.

Authors:

J. Savara^{1,2}, J. Manakova¹, R. Nesnadna¹, A. Petrackova¹, J. Minarik³, T. Papajik³, P. Gajdos², E. Kriegova¹; ¹Dept. of Immunology, Faculty of Med. and Dentistry, Palacky Univ. and Univ. Hosp., Olomouc, Czech Republic, ²Dept. of Computer Sci., Faculty of Electrical Engineering and Computer Sci., VSB-Technical Univ. of Ostrava, Ostrava, Czech Republic, ³Dept. of Hemato-Oncology, Faculty of Med. and Dentistry, Palacky Univ. and Univ. Hosp. Olomouc, Olomouc, Czech Republic

Abstract Body:

Background: Genomic profiling is widely recognised as a key approach for patient management and the development of treatment strategies in hematology. Yet traditional methods have limitations and novel techniques such as whole genome mapping, long-read/single-cell sequencing and epigenomics may further impact our knowledge about the disease-associated aberrations.

Methods: To demonstrate the potential of novel whole genome approaches, we applied optical mapping (OM; Bionano Genomics; ~350x coverage) and long-read next-generation sequencing (lr-NGS) (TELL-Seq; ~100x coverage) to analyze myeloma cells (CD38+CD138+, >86%) enriched from paired samples of peripheral blood (PB; infiltration 27%) and bone marrow aspirate (BM; infiltration 62%) obtained from a patient with extramedullary multiple myeloma (MM; female, 39 years, IgG lambda, stage IIIA, ISS II). Comparison of OM and lr-NGS data was performed using om-annotsv-svc tool (<http://olgen.cz/en/resources>).

Results: OM in BM-derived MM cells detected five translocations (VAF 12-63%), two large deletions on chromosomes (chr)1 (30.6 Mbp) and chr6 (29.3 Mbp), and 1249 deletions (size 1.8 kbp-4.5 Mbp), most of them hardly detected by short-read NGS. Out of these structural variants (SVs) in BM, PB-derived MM cells carried t(1;8), two different t(4;7), large chr6 deletion, but not t(2;15), t(14;19) and large chr1 deletion. Comparison with lr-NGS confirmed 57% (4/7) translocations/large deletions detected in BM. Regarding other SVs, 45% deletions (561/1249), 3% inversions (5/167) and 8% duplications (6/74) detected by OM were confirmed using available lr-NGS variant callers. However, our preliminary results showed that many of missed SVs detected by OM are also present in lr-NGS data. Detected translocations encompassed numerous MM-associated genes (e.g. t(1;8) *TENT5C* and *POU5F1B* in close proximity to *MYC* and t(4;7) *NSD2* and *CREB3L2* genes) as well as genes associated with other cancers (e.g. t(14;19) *NACCI* and t(12;15) *SLC12A6* genes). All aberrations detected by diagnostic FISH/array-CGH (4p16 *NSD2/FGFR3*, 8q24 *MYC*, 13q14 *RBI*, trisomy chr3, chr15 and monosomy chr13, chr23) were confirmed by OM.

Conclusions: Our data revealed that OM may detect additional disease-associated genetic aberrations in hematologic cancers, not detected by NGS. Our pilot analysis suggests that these aberrations are missed in lr-NGS due to the limitations of currently available variant callers. Both OM and lr-NGS may help to improve our understanding of the genomic landscape, not only in hematology.

Grant support: IGA_LF_2022_011, CZ.02.2.69/0.0/0.0/19_073/0016713 and MH CZ-DRO (FNOL, 00098892).

Cancer Posters - Thursday

PB1103. Finding Therapeutic Targets in Cancer Cells Using Essential Genes

Authors:

T-V. Tran, K. Pull, K. Bussey; Midwestern Univ., Glendale, AZ

Abstract Body:

Essential genes are those with functions required for an organism to live and reproduce. Essential genes in one species may not be essential in other species. This suggests essentiality is dependent on context. The atavism theory of cancer proposes cancer represents a reversion to ancestral unicellular pathways and phenotypes. We hypothesize that mode of life, e.g., unicellular or multicellular, is a context for essentiality. Furthermore, if cancer is a reversion to unicellular behavior, genes that are essential in the unicellular context but non-essential in the multicellular context represent targets for therapeutic development.

To test our hypothesis, we focused on nine model species (*E. coli*, *S. cerevisiae*, *S. pombe*, *A. thaliana*, *D. melanogaster*, *C. elegans*, *D. rerio*, *M. musculus*, and *H. sapiens*) that are well-characterized both genetically and phenotypically. We mined Ensembl PanCompara Genomes to obtain a list of orthologous genes that included all the species of interest. This ensured that a lack of essentiality was not due to a lack of an orthologous gene. We used the Online Gene Essentiality (OGEE) Database v2 to define whether a gene was essential for a species. We then classified orthologous gene groups by how many species in which the gene was essential. Genes that were only found to be essential in *E. coli*, *S. cerevisiae*, and *S. pombe* were defined as essential in unicellular life (UCL). Orthologs that were essential in all examples of multicellular life, but not essential in UCL, were defined as essential in multicellular life (MCL).

We did not find any genes that were essential in all species. We discovered 43 genes that were essential in only yeast and mice. Functional enrichment analysis using DAVID revealed these genes are enriched in functions related to glucose transport and metabolism. We identified 12 genes that are essential in a UCL context and 11 genes that are essential in a MCL context. Using Ensembl VEP, we examined the differences in the types and functional outcome of observed variation between UCL and MCL essential genes in multicellular organisms. Chi-square analysis of mouse data demonstrated that the distribution of variant consequences was different between those genes classified as essential in UCL versus those classified as essential in MCL. UCL essential genes had three times as many observed variants as did MCL essential genes. UCL essential genes also had all classes of variants observed, while MCL essential genes had no instances of variations leading to alterations in stop codons or frameshifts and proportionally more synonymous variants.

The findings of the study demonstrate that mode of life can serve as a context for gene essentiality.

Cancer Posters - Wednesday

PB1104. Frequently mutated TOP2B binding sites reveal candidate structural and regulatory cancer drivers of hepatocellular carcinoma

Authors:

L. Uuskula-Reimand¹, C. Lee², D. Abd-Rabbo², S. Ahktar Alvi¹, M. Bhat³, M. Wilson⁴, J. Reimand²; ¹The Hosp. for Sick Children, Toronto, ON, Canada, ²Ontario Inst. for Cancer Res., Toronto, ON, Canada, ³Univ. Hlth.Network, Toronto, ON, Canada, ⁴SickKids Res. Inst./ Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Whole genome sequencing (WGS) of many cancer types has revealed an enrichment of somatic mutations at the binding sites of CTCF and the cohesin complex at boundaries of topologically associating domains (TADs). This is an unexplained phenomenon that is especially evident in hepatocellular carcinoma (HCC). We previously demonstrated that topoisomerase II beta (TOP2B), an enzyme that catalyses transient DNA double strand breaks (DSBs), interacts and extensively co-localizes with CTCF and cohesin in mouse liver, leading to the hypothesis that TOP2B activities may contribute to the mutational processes in HCC.

To understand the role of TOP2B in somatic mutagenesis in HCC, we characterized the genome-wide binding of TOP2B, CTCF and the cohesin complex subunit RAD21 in the genomes of primary human HCCs. Our analysis of somatic mutations in the WGS data of 300 HCC tumors demonstrated that the enrichment of somatic mutations is focal to the co-binding of TOP2B, CTCF and RAD21, including TAD boundary regions shared by a large number of human tissues. Interestingly, TOP2B-RAD21 binding sites associated with liver enhancers are also strongly enriched in mutations. Genome-wide cancer driver discovery revealed 49 frequently mutated candidate drivers in HCC, 36 of which occurred at binding sites of TOP2B. Furthermore, we observed significantly frequent non-coding driver mutations in additional cancer types that are exposed to exogenous carcinogens, such as esophagus, colorectal, lung and skin cancers. Since the majority of candidate non-coding driver mutations are located in TOP2B-associated structural and regulatory elements of the cancer genome, we propose that TOP2B-mediated DSBs represent a mutational mechanism at CTCF and cohesin binding sites that is active in liver and other cancer types.

Cancer Posters - Thursday

PB1105. Functional characterization of the 1p36.33 pancreatic cancer GWAS locus

Authors:

K. Connelly¹, E. Abdolalizadeh¹, J. Zhong¹, D. Eiser¹, J. Hoskins², A. O'Brien¹, A. Jermusyk³, I. Collins⁴, Pancreatic Cancer Cohort Consortium, Pancreatic Cancer Case-Control Consortium, R. Stolzenberg-Solomon¹, A. Klein⁵, J. Shi⁶, B. Wolpin⁷, J. Smith⁸, G. Petersen⁹, L. Amundadottir¹⁰; ¹Natl. Cancer Inst., Bethesda, MD, ²Natl. Cancer Inst., Rockville, MD, ³Natl. Cancer Inst., NIH, Bethesda, MD, ⁴NIH, Bethesda, MD, ⁵Johns Hopkins Med Sch, Baltimore, MD, ⁶Natl. Cancer Institute, Bethesda, MD, ⁷Dana Farber Cancer Inst., Boston, MA, ⁸Georgetown Univ., Washington, D.C., DC, ⁹Mayo Clinic, Rochester, MN, ¹⁰NIH, Rockville, MD

Abstract Body:

Pancreatic ductal adenocarcinoma (PDAC), the third leading cause of cancer-related deaths in the US, is often difficult to diagnose. Genome-wide association studies (GWAS) have identified twenty common susceptibility loci for PDAC. One locus identified, at chr1p36.33 (tagged by rs13303010, OR=1.26, $P=8.36 \times 10^{-14}$), maps to a gene-dense topologically associated domain (TAD) containing four genes: *NOC2L*, *KLHL17*, *PLEKHNI1*, and *SMAD11*. Bayesian and other fine-mapping strategies identified 13 candidate causal variants within this genetic signal to carry forward for functional studies. Of these 13 variants, four demonstrated allele-specific binding in electrophoresis mobility shift assays (EMSA). These four variants were then examined for allele-specific regulatory effects using luciferase assays. Three variants (rs13303327, rs13303010, rs13303160) demonstrated allele-specific regulatory effects with the alternate (protective) alleles having higher activity. To identify transcription factors that exhibit allele-specific binding and mediate allele-specific gene regulation, we performed SNP-proteomics and *in silico* transcription factor binding prediction on these variants. We identified several proteins exhibiting allele-specific binding. For rs13303327, we focused on ELF2, an ETS family transcription factor, identified in both the proteomics and *in silico* analysis. We further confirmed that ELF2 is involved in allele-specific binding of rs13303327 by EMSAs. For rs13303160, *in silico* analysis revealed AP1 transcription factors as likely allele-specific binding proteins. Ongoing *in vitro* experiments aim to validate the *in silico* prediction. The top candidate functional variants in 1p36.33 have significant expression quantitative trait loci (eQTL) associations with *NOC2L* (rs13303160, $P=8.2 \times 10^{-6}$, normalized effect size = -0.27) and *KLHL17* (rs13303160, $P=2.3 \times 10^{-10}$, normalized effect size = 0.37, GTEx v8, pancreas). Co-localization analyses suggest the GWAS, *KLHL17* and *NOC2L* eQTLs represent a single, common genetic signal (posterior probability= 0.99 and 0.75, respectively). *KLHL17* and *NOC2L* knockdown decrease cell proliferation in pancreatic cancer cell lines. Ongoing work aims to validate the allele-specific binding of the functional variants and better understand their downstream biological consequences influencing PDAC pathogenesis.

Cancer Posters - Wednesday

PB1106*. Functional screens implicate candidate breast cancer risk genes in T cell-mediated immune-surveillance

Authors:

J. Beesley¹, W. Shi¹, J. M. Burrows¹, A. Civitaresse¹, J. Rosenbluh², C. Smith¹, G. Chenevix-Trench¹; ¹QIMR Berghofer, Brisbane, Australia, ²Monash Univ., Melbourne, Australia

Abstract Body:

Most loci identified by genome-wide association studies (GWAS) contain many candidate causal variants (CCVs), with molecular pleiotropy implicating multiple target genes, not all of which may be causal. CRISPR (clustered regularly interspaced short palindromic repeats) screens using trait-relevant phenotypes can provide an efficient way of identifying causal genes. We previously performed such screens to identify breast cancer (BC) risk genes involved in proliferation, tumorigenicity and DNA damage repair. To extend to other cancer hallmarks, we hypothesized that some risk genes are involved in immune system processes to detect and destroy transformed cells. Our previous analysis of the 191 target genes we computationally predicted at BC risk loci found immune system processes amongst the most enriched pathways. In order to identify GWAS target genes expressed by a BC cell line (MCF7) which might render it more sensitive or resistant to T cell killing (the dominant effector of immune-surveillance), we established a MCF7 and HLA-matched HER2-restricted T cell co-culture cytotoxicity assay. Using this assay, we have performed two CRISPR knockout (ko) screens: one genome-wide screen, and a custom library containing guides targeting predicted BC risk genes. We detected positive control genes (e.g. *IFNGR1*, *IFNGR2*, *JAK1* and *JAK2*) in both screens, and observed strong overall correlation of gene-level statistics ($r=0.45$). The genome-wide screen identified 13 genes and the custom screen identified 85 genes ($FDR<0.1$), including *CASP8*, *GATA3*, and *ESR1* (knockout rendered the cells resistant to T cell killing) and *CFLAR*, *SRM* and *EWSR1* (sensitivity genes) common to both. In order to determine whether *CASP8* expression might be regulated by the single fine-mapped CCV, we cloned a putative enhancer that contains rs3769821. The protective allele (associated with increased expression in breast tissue in GTEx) was significantly more active in a luciferase assay than the risk allele. Our results therefore suggest that one way in which *CASP8* acts as a BC risk gene is by conferring resistance to T cell killing when expression levels are reduced. We are now carrying out CRISPR inhibition screens, validation, bioinformatic analyses and luciferase assays to identify the causal variants regulating these risk genes. In conclusion, CRISPR screens are an efficient method to follow up GWAS findings.

Cancer Posters - Thursday

PB1107. Functionalizing methylation-derived gene panels in the UK Biobank, TCGA, and HCA datasets

Authors:

U. Mudgal, M. Pietras, Z. Pitluk, S. Moore, S. Sarangi; Paradigm4, Waltham, MA

Abstract Body:

Epigenetic perturbations are inheritable gene expression alterations without modification of the DNA sequence, and accumulation of such alterations are implicated in different types of cancer pathogenesis. One of the major epigenetic mechanisms involving direct chemical modification to the DNA is called DNA methylation, which has been shown to regulate gene expression by recruiting proteins involved in gene repression or by inhibiting the binding of transcription factors to DNA. Studies that identify methylation-derived gene biomarker panels are crucial for better understanding the etiology of cancers and to discover possible therapeutic targets. However, much work needs to be done to validate and functionalize these gene signatures for prospective classification of risk. Conducting population-scale hypotheses-driven validation studies pose challenges such as efficient storage of multi-omics data (e.g., TCGA, UK Biobank, Human Cell Atlas), running complex queries and computations at scale and across datasets, and creation of cohorts based on different metadata fields. The REVEAL FLASH software stack, comprising of a suite of data-specific apps, elastic scaling infrastructure, and a POSIX compliant networked file system, is ideal for performing these kinds of studies. In this poster, we present a use-case with a methylation-derived eight gene panel (*TCTEX1D4*, *MALE*, *LIME1*, *KLHL38*, *HPDL*, *ESR1*, *UCP2*, and *COMMD7*) from a published study on breast cancer (*Frontiers in Genetics*, 2020). First, we ran GWAS analysis using the UK Biobank for associations between variants in the gene panel members from the 200K Whole Exome Sequence dataset and phecodes derived from ICD10 diagnoses of multiple cancer types and other phenotypes of interest. We used SAIGE & REGENIE for the GWAS analysis and a custom linkage disequilibrium (LD) plus burden test algorithm for testing pairs of variants. Next, we looked for evidence of synthetic interactions in the TCGA dataset by querying for co-occurrence in the gene panel of - loss-of-function (LOF) mutations (present in the UK Biobank dataset), LOF mutations and copy number variants, and LOF mutations and non-synonymous mutations. To assign cell type specificity, we queried the Human Cell Atlas and Tabula Sapiens datasets for expression of members of the gene panel. Finally, we perform a Kaplan-Meier survival analysis using TCGA and UK Biobank data for the top associations and interactions.

Cancer Posters - Wednesday

PB1108. Gene fusion detection and characterization in long-read cancer transcriptomes with FusionSeeker

Authors:

Y. Chen¹, Y. Wang², W. Chen¹, Y. Song¹, H. Chen¹, Z. Chong²; ¹Univ. of Alabama at Birmingham, Birmingham, AL, ²UAB, Birmingham, AL

Abstract Body:

Gene fusions are prevalent in a wide array of cancer types and often play critical roles in tumorigenesis and progression. Long-read RNA sequencing technologies, such as PacBio Iso-Seq and Nanopore direct RNA sequencing, can generate full-length transcript sequencing reads and therefore show great potential in gene fusion detection. However, only two long-read gene fusion detection tools are currently available, and their performance is limited in terms of sensitivity and precision. Here, we present FusionSeeker to comprehensively characterize gene fusions and reconstruct accurate fused transcripts using long-read cancer transcriptome data.

FusionSeeker consists of three major steps: raw signal detection, candidate event clustering and filtering, and transcript sequence reconstruction. It first scans the read alignments for split-read patterns and records gene fusion raw signals if two alignments from the same read reside in distinct genes. All raw signals are then clustered with a density-based clustering algorithm into candidate fusions. FusionSeeker discards candidates supported by few reads as noise and only reports confident gene fusions. For each event, FusionSeeker collects fusion-containing reads to generate an accurate consensus sequence with partial order alignment algorithm. The final output of FusionSeeker includes a list of confident gene fusion events and their transcript sequences.

We benchmarked the performance of FusionSeeker, JAFFAL, and LongGF on both *in silico* and real cancer transcriptome datasets. In three replicate simulated datasets, FusionSeeker outperformed the other two fusion callers, with accuracy over 93% for both Iso-Seq and Nanopore-like reads. FusionSeeker also successfully reconstructed full-length transcript sequences for more than 99.5% of simulated gene fusion events, with an average sequence identity over 99%. When applied on SKBR-3 cell line, FusionSeeker identified 15 previously validated gene fusion events, while JAFFAL and LongGF detected 13 and 10 events, respectively. When comparing gene fusion calls from the three tools, 19 events were reported only by FusionSeeker. 17 out of these 19 FusionSeeker-unique events were cross-validated in DNA sequencing data with a validation rate of 89.47%, which was higher than JAFFAL-unique (27.27%) and LongGF-unique events (60.00%). In MCF-7 cell line, we designed PCR primers and validated 7 novel gene fusion events that have not been previously reported. In both SKBR-3 and MCF-7 cell lines, FusionSeeker transcript sequences showed significantly higher base accuracy than raw reads, indicating that most sequencing errors have been corrected.

Cancer Posters - Thursday

PB1109. Gene-environment interaction of antioxidant genes and a healthy lifestyle index (HeLiX) on breast cancer risk reduction

Authors:

L. Gómez Flores Ramos¹, L. Sánchez-Zamorano², A. Ángeles-Llerenas², R. Rodríguez-Valentín², G. Torres-Mejía²; ¹CONACYT-Natl. Inst. of Publ. Hlth., Cuernavaca, Mexico, ²Natl. Inst. of Publ. Hlth., Cuernavaca, Mexico

Abstract Body:

Oxidative stress (OS) is a well-known risk factor for tumorigenesis. Lifestyle factors are associated with breast cancer (BC) through different pathways, including OS. Antioxidant enzymes are the endogenous defense against OS damage, and this response might change by the variation in these enzyme-coding genes. We aimed to analyze the synergistic effect of genetic variants in antioxidant family genes and a Healthy Lifestyle Index (HeLiX, composed by principal components of Western dietary pattern, alcohol consumption, smoking, and physical activity) on BC risk. We analyzed data from a population-based case-control study with 636 BC cases and 678 controls. We included 176 SNPs from *SOD*, *GPX*, and *CAT* family genes. Only SNPs in Hardy-Weinberg equilibrium in controls were included in the final models. Twelve *CAT* SNPs were associated with BC after adjustment for Native American Ancestry (*rs475043*, *rs494024*, *rs511895*, *rs533425*, *rs554576*, *rs560807*, *rs2073058*, *rs2179625*, *rs2300181*, *rs7104301*, *rs7933285*, *rs12270780*). Only *CAT rs554576* (g.28423A>T) remained significant after correction for multiple comparisons. In postmenopausal women, compared to the reference category (lowest HeLiX tertile T1 and *AA* genotype), women in the highest HeLiX tertile (T3) with *TT* or *AT* genotypes showed 0.15 times the odds of BC (OR= 0.15, 95% CI 0.07-0.32). Women in the medium HeLiX tertile (T2) with *TT* or *AT* genotypes also showed a protective effect (OR= 0.26 95% CI 0.13-0.54; OR= 0.24, 95% CI 0.12-0.45, respectively). For premenopausal women, we observed an odds reduction in women at the highest tertile of HeLiX (T3) with the *TT* (OR= 0.21, 95% CI 0.08-0.51) and *AT* genotypes (OR= 0.29, 95% CI 0.13-0.62). We observed an inverse association for the medium tertile of HeLiX (T2) for women with the *TT* genotype (OR= 0.39 95% CI 0.17-0.87). *CAT rs554576* is a polymorphism located at intron 9; it is a modifier variant and an expression of quantitative trait loci (eQTL) with a cis-regulatory effect on *CAT* and the upstream gene *ABTB2*. *CAT* has higher expression in adipose and breast tissues and is responsive to lifestyle exposures. Our study shows a significant synergistic gene-environment interaction, contributing to understanding BC etiology and prevention pathways.

Cancer Posters - Wednesday

PB1110. Genetic ancestry correlates with somatic mutations in endometrial cancer.

Authors:

R. Bremseth-Vining¹, V. Borda², M. Cote^{3,4}, N. Bateman⁵, G. Dyson^{3,4}, D. Craig^{3,4}, J. Ruterbusch^{3,4}, J. Boerner^{3,4}, T. Conrads⁶, G. Maxwell⁷, K. Darcy^{8,9}, S. Makhon-Moore⁹, T. O'Connor²; ¹Univ. of Maryland, Baltimore, Baltimore, MD, ²Univ. of Maryland Sch. of Med., Baltimore, MD, ³Dept. of Oncology, Sch. of Med., Wayne State Univ., Detroit, MI, ⁴Karmanos Cancer Inst., Detroit, MI, ⁵Women's Hlth.Integrated Res. Ctr., Gynecologic Cancer Ctr. of Excellence, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, ⁶Women's Hlth.Integrated Res. Ctr., Women's Service Line, Inova Hlth.System, Annandale, VA, ⁷Women's Hlth.Integrated Res. Ctr., Inova Hlth.System, Falls Church, VA, ⁸Gynecology Cancer Ctr. of Excellence, Dept. of Gynecologic Surgery and Obstetrics, Uniformed Services Univ. of Hlth.Sci., Walter Reed Natl. Med. Ctr., Bethesda, MD, ⁹The Henry M Jackson Fndn. for the Advancement of Military Med., Inc., Bethesda, MD

Abstract Body:

Endometrial cancer incidence has risen over the last 20 years. Self-described black individuals in the United States show higher incidence and worse outcomes for endometrial cancer than white individuals. Socioeconomic factors and comorbidities that disproportionately affect black individuals play a major role in creating these disparities but do not explain them entirely. Molecular, histopathologic, and genetic elements also contribute to disparities in patient outcome and treatment response. Here we examine how genetic ancestry correlates with somatic mutation rates in endometrial cancer. Germline genetic ancestry and somatic mutation status were examined using normal adjacent tissue and tumor-derived whole exome sequencing data from 297 patients (including 171 self-described black individuals). This data was obtained from Karmanos Cancer Institute and Henry Ford Hospital, and supplied by collaborators at Wayne State University. Local ancestry analysis was performed using RFMix ver. 2. Correlations between African local ancestry and the number of somatic mutations were examined in each local ancestry block along the genome using a negative binomial regression model. A single region in chromosome 8p11.21 reached genome-wide significance. This region contains the gene *SFRP1*, a tumor suppressor; wherein mutations are associated with worse prognostic outcomes in endometrial cancer. This finding necessitates further exploration into the relationship between germline African ancestry and mutations in *SFRP1*. An intra-individual comparison of somatic mutation rates under ancestral diplotypes (a combination of ancestral backgrounds, eg. [European(Eur)-Eur] or [African(Afr)-Eur]) in self-described black individuals showed that somatic mutation rates were significantly higher in Afr-Eur ancestral regions than in Eur-Eur regions. This trend does not persist when comparing Afr-Afr regions to Eur-Eur regions, suggesting that heterozygosity may influence somatic mutation rates in these individuals. An interaction between heterozygosity and somatic mutation burden has yet to be investigated in endometrial cancer. To explore this hypothesis, an F-statistic (inbreeding coefficient) and somatic mutation abundance were computed for target windows of 1Mb for each of the 171 self-described black individuals in this cohort. A meta-analysis of this distribution, using Fisher's combined probability test, showed that heterozygosity correlated with a higher rate of somatic mutations in these individuals. These findings help to shed light on the relationship between germline genetic ancestry and somatic mutations in endometrial cancer.

Cancer Posters - Thursday

PB1111*. Genetic ancestry differences in tumor mutation between early and average onset colorectal cancer

Authors:

B. Rhead¹, D. Hein², Y. Pouliot¹, J. Guinney¹, F. M. De La Vega¹, N. Sanford²; ¹Tempus, Chicago, IL, ²UT Southwestern Med. Ctr., Dallas, TX

Abstract Body:

The incidence and mortality of early onset (EO) colorectal cancer (CRC), defined as CRC diagnosed prior to age 50, are rising when compared to the declining rates of average onset (AO) CRC. Epidemiologic trends for CRC differ by race/ethnicity, which could be connected to underlying differences in tumor mutation and gene expression. Previous research has used self-reported race/ethnicity categories, which are often missing or uninformative for understanding biological underpinnings of such differences, particularly in highly admixed groups such as Black and Hispanic. Hence, we compared tumor mutation profiles of EO and AO CRC - within a de-identified dataset of 1,643 and 3,175 patients, respectively - in the context of genetic ancestry groups. Ancestry informative markers were utilized to estimate ancestry proportion likelihoods for the five continental groups defined in the 1000 Genomes Project: Africa, Americas, East Asia, Europe, and South Asia. Recognizing the complexity of ancestry and race relationships, we imputed several race/ethnicity categories using admixture thresholds based on literature and comparisons with available metadata. Similar numbers of Black and Asian patients were observed in both EO and AO groups (13% and 5-6%), however, more Hispanic/Latino patients were found in the EO subgroup (16%) compared to AO (10%). We assessed the association of genetic ancestry proportions (per 20% increase in each ancestry proportion) with the presence of protein altering somatic mutations or copy number changes in 22 CRC driver genes. Across all ages of onset, African ancestry was associated with higher odds of somatic mutations in *APC* (OR=1.11) and *KRAS* (OR=1.14) and lower odds of *BRAF* mutations (OR=0.84). East Asian ancestry was associated with higher odds of mutations in *TP53* (OR=1.15). Furthermore, we found a novel association between Native American ancestry with higher odds of somatic mutations in the *MLH1* gene (OR=1.61), unexplained by misclassified germline variants in this gene (*i.e.* Lynch syndrome), as we limited our analysis to cases with matched tumor/normal sequencing. When testing for interactions between the effects of onset age (EO vs. AO) and genetic ancestry on gene mutations, we observed a significant interaction ($p=0.031$) between African ancestry and onset age on *APC*: African ancestry was associated with mutations in this gene in AO (OR=1.16) but not EO (OR=0.98). When compared to race/ethnicity classes, genetic ancestry provides a more quantitative and precise profile of shared genetic background that can underlie biological race differences in cancer etiology and outcomes in EO and AO CRC.

Cancer Posters - Wednesday

PB1112. Genetic Modifiers of *KRAS*-Mutant Colorectal Cancer

Authors:

N. Pollock¹, J. Ramroop², T. Gandhi¹, P. T. Campbell³, A. T. Chan^{4,5}, S. Gallinger⁶, G. G. Giles^{7,8}, M. J. Gunter⁹, T. A. Harrison¹⁰, M. Hoffmeister¹¹, P. A. Newcomb¹², S. Ogino¹³, A. I. Phipps¹², F. QU¹², R. E. Schoen¹⁴, A. Pellatt¹⁵, M. O. Woods¹⁶, B. Van Guelpen¹⁷, P. Stevens¹, H. Hampel¹⁸, U. Peters¹², J. P. McElroy¹, A. E. Toland¹; ¹The Ohio State Univ., Columbus, OH, ²The City Coll. of The City Univ. of New York, New York, NY, ³Albert Einstein Coll. of Med., Bronx, NY, ⁴Harvard Med. Sch., Boston, MA, ⁵Massachusetts Gen. Hosp., Boston, MA, ⁶Univ. of Toronto, Toronto, ON, Canada, ⁷The Univ. of Melbourne, Melbourne, Australia, ⁸Cancer Council Victoria, Melbourne, Australia, ⁹Intl. Agency for Res. on Cancer, World Hlth. Organization, Lyon, France, ¹⁰Univ. of Washington, Seattle, WA, ¹¹German Cancer Res. Ctr., Heidelberg, Germany, ¹²Fred Hutchinson Cancer Ctr., Seattle, WA, ¹³Brigham & Women's Hosp, Boston, MA, ¹⁴Univ. of Pittsburgh Med. Ctr., Pittsburgh, PA, ¹⁵Univ. of Utah, Salt Lake City, UT, ¹⁶Mem. Univ. of Newfoundland, St John's, NL, Canada, ¹⁷Umeå Univ., Umeå, Sweden, ¹⁸City of Hope Natl. Med. Ctr., Duarte, CA

Abstract Body:

Somatic mutations in *KRAS* are a common driver of colorectal cancer (CRC) tumorigenesis in 30-40% of patients, yet *KRAS*-mutant tumors remain difficult to treat. Germline variants may influence what somatic mutations occur or provide a selective advantage such that a mutated cell is more likely to progress to a cancer. This “GxM” effect, or genetic variants impacting somatic mutation status, is observed in other cancers and supported by our preliminary studies in CRC. African Americans with CRC are more likely to have tumors with *KRAS* mutations than non-Hispanic Whites, even after considering differences in socioeconomic status and other risk factors. Relatedly, tumors of the proximal colon are more frequent in African Americans and are more likely to have *KRAS* mutations than those of the distal colon. Transcriptional differences between the proximal and distal colon may influence GxM variant interactions in tumorigenesis. We hypothesized that germline variants will alter molecular pathways that support *KRAS* tumorigenesis and differ by proximal-distal colon location. To test this hypothesis, we performed an association study using existing genome-wide germline genotypes and mutation data from 6,386 non-Hispanic White colorectal cancer patients from the Cancer Genome Atlas and the GECCO consortium. We compared germline single nucleotide variants (SNVs) in patients with *KRAS*-mutant tumors to patients with *KRAS*-wild type tumors. No variants met strict thresholds for genome-wide significance. We identified 3 SNVs that associated with *KRAS* mutation status with p -values $<1 \times 10^{-6}$ and 50 with p -values $<1 \times 10^{-5}$. Of 53 SNVs with p -values $<1 \times 10^{-5}$ or $<1 \times 10^{-6}$, 35 had a minor allele frequency $>10\%$ in individuals of African or European ancestry. In a stratified analysis, we identified 29 common SNVs specifically associated with *KRAS* mutations in the proximal colon, 2 in the distal colon, and 10 in the rectum (p -value $<1 \times 10^{-4}$). By patient sex, 32 SNVs specifically had an effect in males and 34 in females (p -value $<1 \times 10^{-4}$). From excluding SNVs in linkage disequilibrium and reviewing relevant genomic features, we identified 90 loci which will be validated in our future studies using a more diverse dataset of ~8,000 patients from multiple sources. In summary, we identified SNVs with suggestive evidence for association with *KRAS* somatic mutations in CRC which may inform the cellular context that supports development of colorectal tumors from cells with *KRAS* mutations.

Cancer Posters - Thursday

PB1113. Genetically Predicted Telomere Length and Risk of Multiple Primary Cancers in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

Authors:

S. He¹, W-Y. Huang¹, M. J. Machiela¹, N. Freedman¹, S. I. Berndt²; ¹Natl. Cancer Inst., Bethesda, MD, ²Natl. Cancer Inst., Rockville, MD

Abstract Body:

Objectives: Telomeres are important in cell division and senescence and critical for chromosomal stability. Leukocyte telomere length is highly heritable and genome-wide association studies have identified over 190 variants associated with telomere length, but little is known about its link with the risk of developing multiple primary cancers. **Methods:** To determine if genetically predicted telomere length (TL) is associated with multiple primary cancers, we conducted a case-control study including 6,399 multiple primary cancer cases and 75,012 controls (92.5% European, 4.1% African, 3.4% Asian) from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. Multiple primary cases were individuals with two or more confirmed primary cancers at different sites. Using up to 160 previously identified variants, we estimated polygenic risk scores (PRS) for telomere length. Logistic regression was used to estimate odds ratios (ORs) for the association of PRS with multiple primary cancers, adjusting for age, sex, genotype platform and ancestry. The PRS was analyzed both as a continuous measure and as quintiles. We also evaluated the association between the PRS of telomere length and the risk of subgroups of multiple primaries, including subgroups composed of multiple primaries with at least one malignancy being adenocarcinoma (N=3,454), squamous cell cancer (N=470), lymphoma (N=895), or sarcoma (N=956). **Results:** Genetically predicted telomere length was associated with an increased risk developing multiple primary cancers (OR=1.78 for one unit increase in TL PRS, 95% confidence interval (CI): 1.53-2.08, P=4.7e-13). The increased risk was strongest for multiple primary cases with at least one cancer being a lymphoid malignancy (OR=2.26, 95%CI: 1.60-3.19, P=8.2e-6). Significant associations were also observed between genetically predicted telomere length and higher risk of multiple primaries with at least one cancer being adenocarcinoma (OR=1.83, 95%CI: 1.54-2.17, P=7.3e-12) or sarcoma (OR=2.19, 95%CI: 1.56-3.07, P=1.1e-5). No significant association was observed for multiple primaries with at least squamous cell cancer (OR=1.17, 95%CI: 0.77-1.76, P=0.47). **Conclusions:** Our study demonstrates that higher genetically predicted telomere length is associated with an increased risk of developing multiple primary cancers, particularly where at least one cancer is a lymphoid malignancy. Further studies are needed to understand the biological mechanisms between telomere length and the risk of multiple primary cancers.

Cancer Posters - Wednesday

PB1114. Genome and transcriptome landscape of familial pancreatic cancer and implications for altered glucose metabolism in moderate-penetrance cancer susceptibility.

Authors:

K. Dixon¹, Y. Shen², C. Hoeschen³, H. Wong³, C. Cremin³, J. Senz⁴, A. Lum⁴, K. O'Neill², C. Ch'ng², Q. Hong³, J. Karasinska⁵, J. Topham², E. D. Pleasance², M. Jones², E. Zhao², S. Young⁶, K. L. Mungall², A. J. Mungall², R. Moore², Y. Ma², A. Fok², J. Nelson², M. K. C. Lee⁷, S. Yip⁶, H. Lim⁷, A. Karsan², S. J. M. Jones², J. Laskin⁷, M. M. Marra², D. F. Schaeffer⁶, D. J. Renouf⁷, D. G. Huntsman⁶, K. A. Schrader¹; ¹Dept. of Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, ²Canada's Michael Smith Genome Sci. Ctr., BC Cancer, Vancouver, BC, Canada, ³Hereditary Cancer Program, BC Cancer, Vancouver, BC, Canada, ⁴Dept. of Molecular Oncology, BC Cancer, Vancouver, BC, Canada, ⁵Pancreas Ctr. BC, Vancouver, BC, Canada, ⁶Dept. of Pathology and Lab. Med., Univ. of British Columbia, Vancouver, BC, Canada, ⁷Dept. of Med. Oncology, BC Cancer, Vancouver, BC, Canada

Abstract Body:

Pancreatic cancer is one of the leading causes of cancer-related deaths worldwide. Although genetic factors contribute to a significant proportion of pancreatic cancer risk, the causal genes are largely unknown. Around 15-20% of non-syndromic familial pancreatic cancer (FPC), defined as a family with at least two affected first-degree relatives, is caused by pathogenic germline variants in *ATM*, *BRCA2*, and *PALB2*. Genes involved in cancer or tissue biology, including mediators of metabolic, immune and microenvironmental response, may also play a role in pancreatic cancer susceptibility. To investigate the molecular pathogenesis of FPC, we performed tumour-normal genome sequencing and tumour RNA sequencing in two affected siblings from a genetically undiagnosed kindred with a multigenerational history of pancreatic cancer. The FPC tumours were characteristic of the stable genomic and classical mRNA subtypes, showing common somatic alterations in RAS, p16-mediated and TGF-beta signalling pathways. Notably, aberrant expression of the genes encoding PDX1, insulin, and other regulators of glucose-mediated hormone secretion in the familial tumours suggested a putative role for altered glucose metabolism in tumourigenesis. These findings could not be explained by an enrichment of pancreatic endocrine cells estimated by transcriptome-based cell type deconvolution. A germline loss-of-function variant was identified in *PIK3C2G*, a gene encoding a liver-specific class II phosphoinositide 3 kinase that has been associated with metabolic phenotypes in humans and mice. History of dyslipidemia and type II diabetes mellitus indicated the presence of other possible related phenotypes in these siblings. Our findings suggest a putative role for altered insulin signalling in pancreatic cancer susceptibility associated with a candidate moderate-penetrance gene. Further studies will be required to evaluate the role of *PIK3C2G* in aberrant glucose metabolism, insulin resistance, and pancreatic cancer.

Cancer Posters - Thursday

PB1115*. Genome-wide Analysis of Rare Haplotypes Associated with Breast Cancer Risk: Discovery, Replication, and Generalizability Evaluation

Authors:

F. Wang¹, W. Moon¹, W. Letsou¹, Y. Sapkota¹, Z. Wang¹, C. Lm², J. Baedke¹, L. Robison¹, Y. Yasui¹; ¹St. Jude Children's Res. Hosp., Memphis, TN, ²Univ. of Alberta, Edmonton, AB, Canada

Abstract Body:

Studies have shown that breast cancer (BCa) risk is influenced by common low-risk variants in numerous loci and rare pathogenic variants in high-risk genes (e.g., *BRCA1/2*). However, the full genetic architecture of BCa risk remains elusive. An inference from the Nordic female twin data provides evidence that rare high-risk variants may be the major determinants of BCa risk. Testing rare-variant associations with disease risk requires large-scale whole-genome sequencing (WGS) data, which are currently being collected. Here, we use haplotypes, sets of variants on a single chromosome, rather than single rare variants, to identify rare loci that explain inherited BCa risk. With 646,446 computationally phased, genotyped variants from 181,034 “white British” women provided by the UK Biobank (UKBB), we conducted a genome-wide haplotype analysis using sliding windows of 5-500 consecutive variants. In the discovery stage, haplotype associations with BCa risk were evaluated retrospectively in the pre-study-enrollment portion of the UKBB data, which included 5,487 BCa cases. BCa hazard ratios (HRs) for additive haplotypic effects were estimated using Cox regression with age as the time axis, adjusting for the top ten genotype principal components. The replication analysis included women free of BCa at enrollment, of whom 3,524 later developed BCa. This two-stage analysis detected 193 rare haplotypes (frequency less than 1%) clustered into 13 distinct loci; each locus spanned a region of 0.1-1.4 Mb and was associated with an appreciable BCa risk increase (discovery: HRs=2.84-6.10, P-value<5x10⁻⁸; replication: HRs=2.08-5.61, P-value<0.01). In contrast, variants that formed these rare haplotypes individually exhibited no or small risk effects. In a separate validation analysis using phased, imputed genotypes from 30,064 female cases and 25,282 controls of European ancestry in the DRIVE OncoArray case-control study, six of the 13 rare loci showed a significant risk increase (logistic-regression-based odds ratio estimates, controlling for the same covariates and imputation uncertainty: 1.48-7.67, P-value<0.05). Functional annotation revealed extensive regulatory DNA elements in BCa-related cells underlying the validated haplotypes, suggesting the influence of intrachromosomal interactions on BCa risk. This study demonstrates both the advantages of using rare haplotypes to find novel risk loci and the possibility of explaining additional BCa heritability. Because our analysis was limited to variants on the UKBB's genotyping array, it implies the discovery of many more risk haplotypes once large germline WGS datasets become available.

Cancer Posters - Wednesday

PB1116. Genome-wide association study identifies 4 novel risk loci including a missense variant in *LGR5* for small intestinal neuroendocrine tumors

Authors:

A. Kumar^{1,2}, M. Aavikko¹, J. Karjalainen^{1,3}, J. Mehtonen¹, K. Palin¹, N. Valimak¹, A. Karhu¹, S. Benjamin⁴, H. Runz⁴, H. Joensuu⁵, T. P. Mäkelä¹, S. Ollila¹, I. E. Kostiaainen¹, C. S. Jantti⁵, FinnGen, A. Palotie¹, L. A. Aaltonen¹, M. J. Daly¹; ¹Univ. of Helsinki, Helsinki, Finland, ²Finnish Cancer Inst., Helsinki, Finland, ³Broad Inst., Boston, MA, ⁴Biogen Inc., Cambridge, MA, ⁵Univ. of Helsinki and Helsinki Univ. Hosp., Helsinki, Finland

Abstract Body:

Small intestine neuroendocrine tumor (SI-NET) is a rare disease with variable clinical symptoms potentially leading to delayed diagnosis and treatment. Understanding of the genetic risk factors underlying SI-NETs might help in the early detection and identification of the individuals at risk. To our knowledge, we report here the largest genome-wide association study (GWAS) of SI-NET with 472 cases and 679,833 controls. We used samples from 307 cases and 287,137 controls in the FinnGen study database to discover SI-NET-associated genetic variants. Next, we performed a meta-analysis with the summary statistics from the United Kingdom Biobank cohort (92 SI-NET cases and 392,696 controls; Fisher test $p = 4.93 \times 10^{-18}$). We found novel associations of SI-NETs with 4 loci near *LGR5*, *FERMT2*, *SEMA6A*, and *CDKAL1* at a genome-wide significance level of (p less than 5×10^{-8}) and replicated the earlier suggested loci near *LTA4H-ELK3* and *KIF16B*. Interestingly, the top variant at the *LGR5* locus (rs200138614, $p = 1.80 \times 10^{-19}$) is a missense variant (p.Cys712Phe) highly enriched in the Finnish population (approximately 25-fold). We further validated the association of *LGR5* p.Cys712Phe with SI-NET in an independent collection of 73 SI-NET patients from Finland and in the United Kingdom Biobank cohort. Functional analyses suggest that *LGR5* p.Cys712Phe is likely a loss of function mutation. Our results demonstrate 4 new associated GWAS loci to SI-NET, among them a novel missense mutation (rs200138614) at *LGR5*.

Cancer Posters - Thursday

PB1117*. Genome-wide association study of intracranial germ cell tumors: a common deletion at *BAK1* attenuates the enhancer activity and confers risk for the rare disease.

Authors:

K. Sonehara¹, Y. Kimura^{2,3}, Y. Nakano^{3,4}, T. Ozawa³, M. Takahashi⁵, K. Suzuki¹, T. Fujii^{6,7,8}, Y. Matsushita^{7,8}, A. Tomiyama^{6,7,8}, T. Kishikawa^{1,9}, K. Yamamoto¹, T. Naito^{1,10}, T. Suzuki¹¹, S. Yamaguchi¹², T. Miwa¹³, H. Sasaki¹³, M. Kitagawa¹⁴, N. Ohe¹⁵, J. Fukai¹⁶, H. Ogiwara², A. Kawamura¹⁷, S. Miyawaki¹⁸, F. Matsuda⁵, N. Kiyokawa², K. Ichimura^{3,8}, R. Nishikawa¹¹, Y. Okada^{1,19,20,10}, K. Terashima²; ¹Osaka Univ. Graduate Sch. of Med., Suita, Japan, ²Natl. Ctr. for Child Hlth. and Dev., Tokyo, Japan, ³Natl. Cancer Ctr. Res. Inst., Tokyo, Japan, ⁴The Univ. of Tokyo Hosp., Tokyo, Japan, ⁵Kyoto Univ. Graduate Sch. of Med., Kyoto, Japan, ⁶Natl. Defense Med. Coll., Tokorozawa, Japan, ⁷Juntendo Univ. Sch. of Med., Tokyo, Japan, ⁸Graduate Sch. of Med., Juntendo Univ., Tokyo, Japan, ⁹Aichi Cancer Ctr. Hosp., Nagoya, Japan, ¹⁰Graduate Sch. of Med., the Univ. of Tokyo, Tokyo, Japan, ¹¹Saitama Med. Univ. Intl. Med. Ctr., Hidaka, Japan, ¹²Faculty of Med., Hokkaido Univ., Sapporo, Japan, ¹³Keio Univ. Sch. of Med., Tokyo, Japan, ¹⁴Shizuoka Children's Hosp., Shizuoka, Japan, ¹⁵Graduate Sch. of Med., Gifu Univ., Gifu, Japan, ¹⁶Wakayama Med. Univ. Sch. of Med., Kimiidera, Japan, ¹⁷Hyogo Prefectural Kobe Children's Hosp., Kobe, Japan, ¹⁸Faculty of Med., the Univ. of Tokyo, Tokyo, Japan, ¹⁹Immunology Frontier Res. Ctr. (WPI-IFReC), Osaka Univ., Suita, Japan, ²⁰RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan

Abstract Body:

Intracranial germ cell tumors (IGCTs) are a heterogeneous group of rare brain tumors mainly diagnosed in children and adolescents, with a substantially higher incidence in East Asian countries than in Western countries (e.g., an incidence of 2.7/million/year in Japan but 0.6/million/year in the United States). Due to the low disease incidence, only a limited amount of basic research has been conducted on this rare disease. Here, through nationwide efforts in patient recruitment, we perform an initial genome-wide association study (GWAS) of 133 IGCTs cases and 762 controls, both of Japanese ancestry. The case sample size is the largest ever reported on IGCTs genetics, enabling us to identify a genome-wide significant locus at 6p21 ($P = 2.4 \times 10^{-9}$, odds ratio = 2.46 [95% CI 1.83-3.31]), which is replicated in an independent dataset of 99 cases and 1026 controls ($P = 1.7 \times 10^{-7}$, odds ratio = 2.22 [95% CI 1.63-3.03]). The lead variant rs3831846 (risk allele frequency = 0.43) is a common 4-bp deletion polymorphism residing in a candidate *cis*-regulatory element E38E2460759, a promoter-proximal enhancer-like element by ENCODE, adjacent to the promoter of *BAK1*. Expression quantitative trait locus (eQTL) analysis using the GTEx dataset reveals that the risk allele of rs3831846 has a down-regulating effect on *BAK1* expression in a wide range of tissues ($P < 0.05$ in the 42/49 tissues, minimum $P = 5.9 \times 10^{-14}$). We further perform *in-vitro* reporter assays to test the allelic difference of rs3831846 in regulatory effect, validating that the risk allele attenuates the enhancer activity of the E38E2460759 sequence, consistent with the eQTL analysis. Finally, given the histological resemblance between IGCTs and testicular GCTs (TGCTs), a subtype of the more common gonadal GCTs, we refer to the previous European TGCTs GWAS and examine the statistics in our IGCTs GWAS for 57 TGCTs risk loci. The effect sizes show a positive overall correlation between Japanese IGCTs and European TGCTs ($P = 1.3 \times 10^{-4}$, Spearman's $\rho = 0.48$). Of the 57 loci, 11 exhibit significant association with IGCTs ($P < 0.05$ with a consistent direction of effect): *CLPTMIL*, *PITX1*, *SPRY4*, *TNXB*, two loci of *BAK1*, *KATNA1*, *DEPTOR*, *GAB2-NARS2*, *HNF1B*, and *TKTL2*. These loci were implicated in a broad range of biological pathways, including KIT/KITLG signaling, apoptosis regulation, and telomerase activity. Our study indicates the shared genetic background of GCTs beyond ethnicity and primary sites, not limited to a specific biological pathway.

Cancer Posters - Wednesday

PB1118. Genome-wide profiling of DNA N6methylation from a breast cancer and a matched normal cell lines

Authors:

C. Xiao, V. Schneider; NIH, Bethesda, MD

Abstract Body:

DNA N6-methyladenine (m6A) modification has been found widely presented in the human genome, and genome-wide DNA methylation profiling in cancer may reveal epigenetic signatures with significant clinical outcomes. Whole genome sequencing data from PacBio single-molecule real-time (SMRT) system provides signals for identifying the presence of m6A in human genomic DNA. We identified 343,199 m6A modification sites with an average density of 122 per Mb in the HCC1395 breast cancer cell line, whereas 722,303 m6A modification sites with average density of 257 per Mb were observed in a matched normal HCC1395BL cell line, meaning that the total number and the average density of m6A methylation sites in the cancer cell line were reduced more than 50% than that in normal cell line. Only small fraction of the methylation sites was found to be shared between the two cell lines, indicating that significant de-methylation (loss) and new methylation (gain) events occurred in HCC1395 cancer cells. A broad distribution of m6A methylation sites across autosomal chromosomes was observed, but the density of m6A methylation sites on chromosome X was extremely low for both HCC1395 and HCC1395BL cell lines. In contrast, the m6A densities on chromosome 7 and chromosome 22, particularly on their q-arms, from HCC1395 were substantially higher (hypermethylation) than other autosomal chromosomes, suggesting that copy number variations may be associated with these chromosomal regions. Most of the m6A methylation sites were located in intergenic and intronic regions, whereas only about 2~3% of the m6A methylation sites were situated in exonic regions, and 12% on ncRNAs. We investigated enriched motifs from the flanking sequences of the m6A sites and found that the m6A modifications occurred the most at the sites with GAG motif in both HCC1395BL (57.3%) and HCC1395 (58.26%). The next most frequent motifs (7 ~ 9% each) for m6A sites is GAA/AAG/GAC for HCC1395BL, and GAC/AAG/GAA for HCC1395. Understanding the genome-wide distinction of DNA methylation between cancer and matched normal cell lines makes it possible to ask more targeted questions and further investigate the role of methylation in cancer.

Cancer Posters - Thursday

PB1119*. Genomic landscape of malignant peripheral nerve sheath tumor (MPNST) reveals novel pathways of tumor evolution that correlate with tumor behavior.

Authors:

D. Miller¹, C. D. Steele², K. Piculell¹, A. Al-Ibraheemi³, V. Eulo⁴, M. M. Bui⁵, A. Chatzipli⁶, B. C. Dickson⁷, D. C. Borchering⁸, A. Feber⁹, A. Galor¹⁰, J. Hart¹¹, K. B. Jones¹², J. T. Jordan¹³, R. H. Kim¹⁴, D. Lindsay¹⁵, C. Miller¹⁶, Y. Nishida¹⁷, P. Proszek¹⁰, J. Serrano¹⁸, R. T. Sundby¹⁹, J. J. Szymanski²⁰, N. J. Ulrich²¹, D. Viskochil²², X. Wang²³, M. Snuderl¹⁸, P. J. Park⁶, A. M. Flanagan², A. C. Hirbe⁸, N. Pillay², I. Cortes-Ciriano¹⁶; ¹Div. of Genetics and Genomics, Boston Children's Hosp., Boston, MA, ²Res. Dept. of Pathology, Univ. Coll. London Cancer Inst., London, United Kingdom, ³Dept. of Pathology, Boston Children's Hosp., Boston, MA, ⁴Div. of Oncology, Dept. of Internal Med., Univ. of Alabama at Birmingham, Birmingham, AL, ⁵Dept. of Pathology, Moffitt Cancer Ctr. & Res. Inst., Tampa, FL, ⁶Dept. of BioMed. Informatics, Harvard Med. Sch., Boston, MA, ⁷Dept. of Lab. Med. and Pathobiology, Univ. of Toronto; Dept. of Pathology and Lab. Med., Mount Sinai Hosp., Toronto, ON, Canada, ⁸Div. of Oncology, Dept.s of Internal Med. and Pediatrics, Siteman Cancer Ctr., Washington Univ. Sch. of Med. in St. Louis, St. Louis, MO, ⁹Clinical Genomics Translational Res., Inst. of Cancer Res., Royal Marsden NHS Fndn. Trust, London, United Kingdom, ¹⁰Univ. of Oxford, Wellcome Ctr. for Human Genetics, Oxford, United Kingdom, ¹¹Dept. of Pathology, Lifespan Lab., Rhode Island Hosp., Providence, RI, ¹²Dept.s of Orthopaedics and Oncological Sci.; Huntsman Cancer Inst., Salt Lake City, UT, ¹³Pappas Ctr. for Neuro-oncology, Massachusetts Gen. Hosp., Boston, MA, ¹⁴Div. of Med. Oncology and Hematology, Princess Margaret Cancer Ctr., Sinai Hlth.System, Hosp. for Sick Children, Toronto, ON, Canada, ¹⁵Dept. of Histopathology, Royal Natl. Orthopaedic Hosp., NHS Trust, Middlesex, United Kingdom, ¹⁶European Molecular Biology Lab., European Bioinformatics Inst., Cambridge, United Kingdom, ¹⁷Dept. of Rehabilitation Med., Nagoya Univ. Hosp., Nagoya, Japan, ¹⁸Dept. of Pathology, New York Univ. Langone Hlth., Perlmutter Cancer Ctr., New York City, NY, ¹⁹Pediatric Oncology Branch, Ctr. for Cancer Res., Natl. Cancer Inst., NIH, Bethesda, MD, ²⁰Div. of Cancer Biology, Dept. of Radiation Oncology, Washington Univ. Sch. of Med., St. Louis, MO, ²¹Dept. of Neurology, Boston Children's Hosp., Boston, MA, ²²Div. of Med. Genetics, Huntsman Cancer Inst., Univ. of Utah, Salt Lake City, UT, ²³GeneHome, Dept. of Individualized Cancer Management, H. Lee Moffitt Cancer Ctr. and Res. Inst., Tampa, FL

Abstract Body:

Malignant peripheral nerve sheath tumors (MPNSTs), a type of soft tissue sarcoma, are the most common cause of early mortality among people with neurofibromatosis type 1 (NF1). Currently available treatments offer less than a 50% 5-year survival. With the goal of identifying either tumor vulnerabilities or biomarkers for early tumor detection, we formed an international consortium to facilitate a comprehensive genomic analysis of 90 MPNSTs (61 NF1-related; 29 sporadic). All samples were assessed by whole genome sequencing (WGS), RNAseq, and methylation array profiling. Multi-regional exome sequencing was performed on a subset. Comparison of our methylation profiles to pathology and clinical data confirmed prior reports that H3K27 trimethylation (Me3) status correlates with prognosis. Our approach yielded new genomic information, including: (1.) that complex chromosomal rearrangements involving *CDKN2A*, and subsequently the PRC2 complex, are early events in a majority of the tumors, but do not involve recurrent breakpoints; (2.) recurrent gains or losses of particular chromosome arms are indicative of H3K27me3 status, but are actually more predictive of prognosis than H3K27me3 status, pathology characterization, or clinical history; (3.) the combination of WGS, multi-regional exome sequencing, and ploidy analysis facilitated identification of a previously unrecognized complex pattern of divergent tumor evolution along two main pathways that correlate with H3K27me3

status. MPNSTs with loss of H3K27Me3 showed amplification of chromosome 8 and extensive LOH, whereas those with H3K27Me3 retention instead evolve through extensive chromosome instability and chromothripsis, and display more heterogeneous karyotypes; (4.) Transcriptomic data revealed that tumors with loss of H3K27me3 show downregulation of markers of immune infiltration and activation of the adaptive immune system compared to tumors with retention of H3K27me3, suggesting that tumors retaining H3K27Me3 may be more responsive to immunotherapy; (5.) The genomic profiles of different subsets of MPNSTs can be detected through analysis of cell-free DNA (cfDNA). Overall, our results provide the foundation for an MPNST clinical care model that incorporates the genomic architecture of the tumor to refine the diagnosis, prognosis, and treatment approach.

Cancer Posters - Wednesday

PB1120. Germline functional non-coding variants associated with ovarian cancer predisposition

Authors:

S. Ezquina¹, E. Dicks¹, R. Corona², K. Lawrenson², S. Gayther³, M. Jones⁴, M. Freedman⁵, R. Drapkin⁶, P. Pharoah⁷; ¹Univ. of Cambridge, Cambridge, United Kingdom, ²Cedars-Sinai Med. Ctr., Los Angeles, CA, ³Cedars Sinai Med. Ctr., Los Angeles, CA, ⁴Cedars Sinai Med Ctr, Los Angeles, CA, ⁵Broad Inst., Boston, MA, ⁶Univ. of Pennsylvania, Philadelphia, PA, ⁷Univ Cambridge, Cambridge, United Kingdom

Abstract Body:

Epithelial ovarian (EOC) cancer has a substantial hereditary component - rare coding variants in multiple genes that confer moderate to high penetrance have been identified and GWAS have identified many low-penetrance common risk alleles. However, the known risk alleles explain <50% of the inheritable component of risk. Some of the missing heritability is likely to be due to rare variation in the non-coding genome. The aim of this study was to assess whether rare variation in active PAX8 transcription factor binding sites is associated with risk of EOC. PAX8 is highly expressed in high grade serous ovarian cancer (HGSOC) and fallopian tube surface epithelial cells (FTSEC), the putative precursor cells of HGSOC. If rare variation in these regions is disease-associated, the number of variants in these regions would be expected to differ between cases and controls. We used H3K27ac ChIP-seq data from two FTSEC and five primary HGSOCs to identify active chromatin, and PAX8 ChIP-seq from three FTSEC and five ovarian cancer cell lines (JHSO4, Kuramochi, OVSAHO, IGROV, Heya8) to identify PAX8 binding sites. We defined 1,687 active regions covering 309Kbases of genomic DNA as those with overlap between H3K27ac marks in at least three of seven tumour or normal fallopian tube cell lines and PAX8 marks in at least three of eight tumour or normal fallopian tube cell lines. Variants were called in germline genomes of 581 ovarian cancer patients and 2750 non-cancer controls sequenced by Genomics England for the UK 100k Genomes Project. Variants with a frequency higher than 0.01 in the gnomAD dataset and in our dataset were excluded. A Kruskal Wallis test was used as a simple burden test to compare the number of variants carried by cases and controls. The median number of variants in active PAX8 binding sites carried was 5 (IQR 3 - 6) in both controls and cases (P = 0.47) suggesting that rare variation in these regions are not associated with a moderate risk of EOC. The power to detect a weak association is very limited. It is also possible that the functional genomic data used to define active PAX8 binding sites is insufficiently sensitive or specific - some or many of the regions identified may not be functionally important. Furthermore, some of the variants identified might not be functionally active, that is they may not disrupt the PAX8 binding site. Further work is required to identify better functional regions and to determine which variants in such regions are functionally disruptive. This study has shown an approach to detecting disease-associated rare genetic variation in the non-coding genome. Rare variation in other transcription factors may be important in the aetiology of EOC and other cancers.

Cancer Posters - Thursday

PB1121. Germline genetics of venous thromboembolism (VTE): Non-cancer polygenic risk scores improve risk stratification in cancer patients

Authors:

S. Groha¹, A. H. Nassar², E. El-Am³, A. Bejjani², V. Naranbhai⁴, A. Gusev⁵; ¹Dana Farber Cancer Inst., Boston, MA, ²Brigham and Women's Hosp., Boston, MA, ³Indiana Univ. Sch. of Med., Indianapolis, IN, ⁴Massachusetts Gen. Hosp., Cambridge, MA, ⁵Dana-Farber Cancer Inst., Boston, MA

Abstract Body:

Background

VTE is a significant secondary cause of mortality in cancer patients, with up to 20% experiencing a VTE during their cancer treatment, constituting a 5-7 times increased risk as compared to the non-cancer population (Razak et al, Cancers 2018). The leading clinical predictor, 'Khorana score' is based on few clinical factors and has been challenging to translate into a clinically effective biomarker. Germline variants are known to be significant factors for VTEs in patients without cancer. In large GWAS studies, the predictive value of polygenic risk scores (PRS) was comparable to well-known genetic risk factors such as factor V Leiden and prothrombin G20210A mutations (Klarin et al, Nature Genetics, 2019). Here, we investigate the utility of a non-cancer VTE PRS to clinically stratify VTE risk in cancer patients and compare to existing clinical risk models.

Methods

The study was conducted in a discovery cohort of 5,760 patients with solid tumors sequenced at Dana-Farber Cancer Institute (DFCI) and treated with anti-neoplastic therapy. The VTE phenotype for the 5,760 patients was abstracted using ICD-10 codes. For validation, we performed manual chart review on a subpopulation of 806 patients with pancreatic cancer sequenced at DFCI. Targeted tumor sequencing was used to impute a VTE PRS from external GWAS data of ~420,000 individuals. Predictive accuracy was quantified in a cause-specific way in a time-to-event analysis and benchmarked against the clinically established Khorana score (Khorana, Blood 2008) for cancer-associated thromboembolism. Associations were then validated in an independent cohort of 927 patients sequenced after 2020 to evaluate predictive shift.

Results

In a univariate analysis, the VTE PRS alone yielded comparable concordance ($c=0.55$ [0.54-0.56]) to the clinical Khorana score ($c=0.56$ [0.54-0.58]). Combining these predictors increased the cause specific concordance ($c=0.57$ [0.56-0.58]), indicating independent information from the germline predictor. This is furthermore corroborated on the level of effect sizes. Importantly, we find that the PRS predictor remains highly significant (cause-specific HR=1.2, $p=1.3 \times 10^{-7}$) after controlling for known clinical factors (cancer type, cancer stage, ECOG performance status, chemotherapy medication, stationary hospital stay, metastasis status, gender, age at start of treatment and genetic ancestry), whereas the Khorana score was no longer significant (HR=0.99 [0.91-1.1], $p=0.88$). A germline VTE PRS could improve stratification of VTE risk in patients with cancer.

Cancer Posters - Wednesday

PB1122. Germline sequencing in Brazilian pancreatic carcinoma patients.

Authors:

L. Munhoz Rodrigues¹, S. Maistro², L. Antônio Senna Leite², M. del Pilar Estevez Diz², U. Ribeiro Jr.¹, R. Guindalini³, M. Folgueira¹; ¹Faculdade de Med. da Univ.e de Sao Paulo, Sao Paulo, Brazil, ²Inst. do Cancer do Estado de Sao Paulo, Sao Paulo, Brazil, ³Oncologia D'Or, Bahia, Brazil

Abstract Body:

Background: Hereditary susceptibility is scarcely known in pancreatic cancer patients in Latin America, which presents a multiethnic population. We evaluated an unselected cohort of Brazilian pancreatic carcinoma (PC) patients to detect the spectrum and frequency of germline mutations in cancer predisposing genes. **Methods:** Patients from Instituto do Cancer do Estado de Sao Paulo (Sao Paulo, located in the southeastern region of Brazil), with histopathological diagnosis of non-endocrine PC were included between 2018-2021. Genomic DNA was obtained from peripheral blood for Next Generation Sequencing using a panel of 113 cancer predisposing genes. Copy number variation (CNV) was assessed by CNV-Atlas and confirmed by Multiplex Ligation-dependent Probe Amplification. **Results:** A total of 180 participants were evaluated, with mean age 61 years (27-87), among whom, 17% with young age (≤ 50 years) and 108 women (60%). Regarding risk factors, 106 (59%) participants reported cases of any cancer in first-degree relatives, 101 (56%) reported a history of smoking and 39 (22%) of alcohol abuse, 125 (70%) reported previous overweight or obesity, 38 (21%) reported diabetes for at least one year before cancer diagnosis and 26 (14%) had a history of occupational exposure. No participants reported Ashkenazi Jewish ancestry and 69 (38%) participants self-declared black or brown - “pardo” skin color. Sixteen participants (9%) had more than one diagnosis of primary cancer and 110 (61%) presented with advanced disease (clinical stage III or IV). We observed that 28 (16%) participants carried pathogenic/likely pathogenic variants, in the following genes: *CHEK2* (n=5), *ATM* (n=4), *BRCA1*, *BRIP1*, *FANCM*, *MTIF*, *MUTYH*, *PALB2* (each gene in two patients) and *BRCA2*, *CDKN2A*, *FANCE*, *MRE11*, *MSH2*, *RAD51C*, *RECQL4*, *SDHA* (each gene in one patient), as well as deletion of 2 exons of *ATM* and duplication of 13 exons of *CHEK2* (each alteration in one participant). Among these, two patients carried two gene variants and one patient carried one variant and one CNV. **Conclusions:** The use of multigene panel in this unselected and multiethnic group of Brazilian PC patients revealed that 16% were carriers of pathogenic germline variants or CNVs, mainly in *CHEK2* and *ATM*.

Cancer Posters - Thursday

PB1123. Germline sequencing of DNA damage repair genes in two hereditary prostate cancer cohorts reveals rare risk-associated variants

Authors:

G. Foley¹, J. Marthick¹, S. Lucas¹, K. Raspin¹, A. Banks¹, J. Stanford², E. Ostrander³, L. FitzGerald¹, J. Dickinson¹; ¹Univ. of Tasmania, Hobart, Australia, ²Fred Hutchinson Cancer Res Ctr, Seattle, WA, ³NHGRI (NIH), Bethesda, MD

Abstract Body:

Rare, inherited variants in DNA damage repair (DDR) genes play an important role in prostate cancer (PrCa) susceptibility. Here, two independent high-risk familial PrCa datasets were interrogated to identify novel, rare DDR variants that contribute to disease risk. Massively parallel sequencing data from Australian and North American familial PrCa datasets were examined for rare, likely deleterious variants in 35 DDR genes. Putative high-risk variants were prioritised based on frequency (minor allele frequency <1%), mutation type (nonsense, missense, or splice), segregation with disease, and *in silico* predicted deleteriousness. Six prioritised variants were genotyped in a total of 1,963 individuals (700 familial and 459 sporadic PrCa cases, 482 unaffected relatives, and 322 screened controls) and M_{QLS} association analysis was performed. Statistically significant associations between PrCa risk and rare variants in *ERCC3* (rs145201970, $p=2.57 \times 10^{-4}$) and *BRIP1* (rs4988345, $p=0.025$) were identified in the combined Australian and North American datasets. A variant in *PARP2* (rs200603922, $p=0.028$) was significantly associated with risk in the Australian dataset alone, while a variant in *MUTYH* (rs36053993, $p=0.031$) was significantly associated with risk in the North American dataset. Our study implicates several rare germline DDR variants in familial PrCa risk, indicating that additional genes in DDR pathways should be considered when evaluating inherited genetic risk. Understanding the functional and/or biological effects of these variants may have wide-reaching implications on the burden of PrCa, especially in their roles as potential gene-based therapeutic targets. Further, we provide evidence that a proportion of rare DDR variants will elude screening confined to early-onset or high-grade familial disease, with implications for the selection of gene-based therapies targeting DDR pathways in PrCa patients.

Cancer Posters - Wednesday

PB1124. Germline variants in the Ah Receptor in Colombian smoker patients with squamous cell carcinoma of the head and neck.

Authors:

N. Trujillo¹, C. I. Vargas C.¹, Á. A. Herrera H.¹, S. Guauque-Olarte², C. Fong³, L. Cifuentes C.⁴; ¹Univ. Industrial de Santander, Bucaramanga, Colombia, ²Univ. Cooperativa de Colombia, Envigado, Colombia, ³Univ. Cooperativa de Colombia, Santa Marta, Colombia, ⁴Univ. Cooperativa de Colombia, Pasto, Colombia

Abstract Body:

Head and neck squamous cell carcinoma (HNSCC) evolves from the mucosal epithelium in the oral cavity, pharynx and larynx, and collectively represents the 7th most common cancer by incidence worldwide. Colombia by 2020 reported 3.148 new cases and at least 1343 deaths associated to this neoplasm. The strongest risk factor for developing HNSCC is Tobacco consumption. The Aryl hydrocarbon Receptor (AhR) encoded by the AHR gene, induces the activation of xenobiotic-metabolizing enzymes. However, interindividual variation in the activity of these enzymes involved in the detoxification of tobacco carcinogens can contribute to cancer susceptibility. We aimed to identify genetic variants in the AHR gene in Colombian smoker patients with HNSCC, and to establish their current effect on the structure - function of the protein product. **Materials and Methods:** This study included the analysis of 23 blood samples from HNSCC smoker patients. The AHR gene coding sequence and its exon-intron boundaries were sequenced through Sanger's method. We determined the presence and localization of variants using BLAST and SW ChromasPro. Also, in silico structural, and functional predictions of the protein were performed by using bioinformatics tools like VEP, SIFT, POLYPHEN, FATHMM, Mutation Taster, PROVEAN and Human Splicing Finder. **Results:** We reported for the first time genetic variants in the AHR gene in HNSCC colombian patients. Also, we identified variants not previously reported for the Latin American population. Some of the variants have been classified as having unknown clinical significance, while others have not been previously reported. Through in silico analysis some variants were found to possibly affect the protein structure, and therefore the function of the Ah Receptor. Allele frequencies were establish for each variant according to the 1000 Genomes Project, GnomAD, and ExAC databases. **Conclusion:** The findings represent a first approach to identify SNPs in the AHR gene in Colombian population. This study contributes to the understanding of the genetic susceptibility to HNC in Colombia. **Financial Support:** CONADI-UCC (INV2085).

Cancer Posters - Thursday

PB1125. GLUT10 is a biomarker for poor prognosis and a predictor of vitamin C combination therapy in breast cancer

Authors:

Y-C. Lee, H-L. Chou, T-T. Nguyen; Academia Sinica, Taipei, Taiwan

Abstract Body:

Breast cancer is the most common malignancy and the leading cause of cancer mortality among women worldwide. Treatment failure and breast cancer mortality are primarily caused by cancer heterogeneity and metastasis. Metastasis cancer cells adapt to the environmental changes during the metastatic cascade through selection and metabolism reprogramming. Thus, the discovery of new markers and development of treatments targeting metabolism reprogramming will benefit breast cancer patients at high risk of metastasis. Here we report the identification of the increased expression of glucose transporter 10 (GLUT10) is associated with poor survival in all subtypes of breast cancer patients. Furthermore, we show that GLUT10 expression enhances cancer progression through promoting mitochondrial function and HIF-1-mediated pathways in a glucose concentration-dependent manner. GLUT10 transports the dehydroascorbic acid, the oxidize form of ascorbic acid (vitamin C) more efficiently than glucose. Notably, this GLUT10-induced breast cancer progression can be disrupted by vitamin C supplementation to inhibit mitochondrial function and enhance HIF-1 degradation. Our study reveals that GLUT10 is a crucial regulator of metabolism reprogramming and breast cancer metastasis, serving as a biomarker and an actionable therapeutic target for breast cancer.

Cancer Posters - Wednesday

PB1126. Going beyond the atlas: Mapping the tumor landscape with complete genomic signatures

Authors:

K. Kennedy¹, J. Dahl¹, M. Graybill¹, T. Morozova¹, J. Zawistowski², J. Blackinton³, J. West⁴; ¹BioSkryb Genomics Inc, Durham, NC, ²BioSkryb Genomics, Inc., Durham, NC, ³BioSkryb Genomics, Durham, NC, ⁴BioSkryb, Inc., Durham, NC

Abstract Body:

Glioblastoma multiforme (GBM) represents an aggressive malignancy with dismal outcome. Despite recent advances in tumor biology, there are limited data capable of linking tumor landscape with genomic signatures to improve patient outcomes. Current tissue atlas methods capably identify cellular location and phenotype in tumor sections, however, they woefully underrepresent the totality of the genomic and transcriptomic landscape, frequently focusing on a pre-defined panel of genes or transcripts. One powerful combination capable of revealing such data is the CellCelector platform from ALS paired with ResolveOME to reveal the complete genome and transcriptome of a single cell. Here, we present an unbiased mapping approach that simultaneously allows for the complete assessment of the tumor genome and transcriptome at the single cell level in a highly heterogenous and deadly tumor type. In this study, GBM tissues were digested *in-situ* and single cells selected for analysis using the CellCelector platform from ALS. For this pilot project, a total of 88 cells were evaluated in an unbiased survey across the landscape of the tumor. Single cells were selected distanced 10uM apart and importantly, no pre-defined stains were utilized to ensure random cell sampling. Single cells were processed using ResolveOME, enabling simultaneous evaluation of the complete genome and transcriptome. Whole genome and whole transcriptome sequencing was performed targeting 70M PE reads and 2M PE reads respectively. After sequencing, genomic and transcriptomic data were processed using the BaseJumper platform to identify cell type, cellular function, and differential gene expression in relation to genomic heterogeneity. We observe substantial heterogeneity between cells that can be organized into clear lineage tracing relationships between both spatial arrangement and mutational signature. This enables an improved understanding of the development of the tumor, and importantly, key insights into potential treatments.

Cancer Posters - Thursday

PB1127*. HDM2 differentially modulates the molecular phenotype of lung tumor cells of varying p53 genotypes

Authors:

Y. Wang, C. Li, K. Ramos; Texas A&M Inst. of BioSci.s and Technology, Houston, TX

Abstract Body:

Retrotransposons such as the Long Interspersed Nuclear Element-1 (LINE-1) represent the most powerful remodeling force of the human genome. LINE-1 is the only autonomous retrotransposon remaining in the human genome and is strongly linked to the poor prognosis of non-small cell lung cancers. Previous studies by our group identified a 4kb intronic LINE-1 insertion into the NACC2 locus of nickel-transformed human bronchial epithelial cells. This insertion led to dramatic reductions in NACC2 at both the mRNA and protein levels. The mRNA levels of HDM2, a protein repressed through NuRD and a negative regulator of p53 through polyubiquitination, were markedly increased in transformed cells. To evaluate the functionality of the HDM2/p53 genetic network, HDM2 was overexpressed in three independent clones of NCI-460 (p53^{wt}), NCI-520 (p53^{mut}), and NCI-1299 (p53^{mut}) cells, and the protein expression of critical molecular targets examined by Western blotting. HDM2 overexpression increased LINE-1 ORF-1p levels in all three clones of wild-type NCI-460 cells but decreased protein expression in all NCI-520 and NCI-1299 clones. TP53 increased in NCI 460 cells and as expected was undetectable in NCI-520 and NCI 1299 clones. The expression level of EGFR did not change irrespective of the p53 genotype, while RB decreased in NCI-H460 clones, but remained unchanged in NCI-520 and NCI-1299 clones. These findings establish the functional integrity of the HDM2/p53 axis and the strong influence of TP53 interactions with LINE-1 in defining lung cancer phenotypes. Together, these data implicate LINE-1 as a targetable pathway for precision therapeutics in lung cancer.

Cancer Posters - Wednesday

PB1128. Hemophagocytosis related to acute promyelocytic leukemia is also positive for the *PML::RARA* gene fusion

Authors:

A. García Romero¹, M. Magaña-Torres², R. González-Arreola¹, C. Borjas-Gutiérrez², J. González-García²; ¹Univ. de Guadalajara, Guadalajara, Mexico, ²Inst. Mexicano del Seguro Social, Guadalajara, Mexico

Abstract Body:

Introduction: Acute promyelocytic leukemia (APL) is characterized by the chromosomal translocation t(15;17)(q24.1;q21.2), which fuses the *PML* and *RARA* genes, raising the hybrid gene *PML::RARA* considered the direct cause of the pathology. In several types of leukemias with different chromosomal arrangements, an hemophagocytic phenomenon has been observed, where the macrophages show a great phagocytic activity of other blood cells; this phenomenon has previously been reported in association with APL on four occasions. The objective of this work was to determine the presence of the *PML::RARA* fusion in phagocytic and phagocytosed cells observed in APL. The study was carried out in a total of 22 patients with APL. Samples were taken before antineoplastic treatment. Patient's data of their clinical characteristics and blood studies were evaluated. Hemophagocytic binomials (phagocytic cell/phagocytosed cell) were identified in bone marrow smears stained with Wright; then, we performed a fluorescence in situ hybridization (FISH) study with the commercial *PML/RARA* dual color/dual fusion in order to determine the presence or absence of the *PML::RARA* gene fusion in these cells. Results: The average age of the patients was 25.6 years, 13 adults and 9 children; 14 men and 8 women. The most common clinical symptoms were cytopenia and coagulopathy. Hemophagocytic binomials were observed in all patients. A total of 548 binomials were evaluated. It was found that the most frequent combination of binomials was the one where both the phagocytic and phagocytosed cells were positive for the *PML::RARA* fusion (81.57%) (p<0.0001; Yates corrected χ^2 test). An unexpected finding was the observation of cells showing a netotic-like morphology, which sometimes appeared trapping erythrocyte cells. These cells were observed in all patients and all without exception were positive for the *PML::RARA* fusion. Conclusions: The hemophagocytic phenomenon was observed in all APL patients. The combination where both cells (phagocytic cell/phagocytosed cell) were positive for the *PML::RARA* gene fusion, was the most frequent of the four possible combinations. Almost all phagocytic cells (544/548) were positive for the *PML::RARA* gene fusion. All netotic-like cells were positive for the *PML::RARA* gene fusion. It is possible that the mutational event that gave rise to the *PML::RARA* gene fusion had occurred in a common progenitor cell of the monocyte/macrophage and myelocyte lineage.

Cancer Posters - Thursday

PB1129. Hereditary Cancer Variants & Homologous Recombination Repair Deficiency in Biliary Tract Cancer

Authors:

H. Nakagawa¹, Y. Okawa¹, T. Nakamura², S. Hirano², Y. Momozawa¹; ¹RIKEN IMS, Yokohama, Japan, ²Hokkaido Univ. Faculty of Med., Sapporo, Japan

Abstract Body:

Background: Biliary tract cancer (BTC) is a rare malignant tumor, but the incidence of BTC is high in Japan, East Asia and South America and increasing worldwide. The heritability and actionability of molecular therapies for BTC are uncertain. Although associations between BTC and *BRCA* germline variants have been reported, homologous recombination deficiency (HRD) has not been investigated in BTCs. **Methods:** We here analyzed germline variants in 27 cancer-predisposing genes (JAMA Oncology 2022) in 1,292 Japanese BTC patients, which were enrolled in Biobank Japan (BBJ) and Hokkaido University Hospital, and 37,583 controls without a personal and family cancer history. We compared the pathogenic germline variant frequency between cases and controls and documented the demographic and clinical characteristics of carrier patients. In addition, whole-genome sequencing analysis of 45 BTC tissues was performed to evaluate somatic mutations and homologous recombination deficiency status. **Results:** Targeted sequencing identified 5,018 germline variants, which were classified into 317 pathogenic, 3,611 variants of uncertain significance, and 1,090 benign variants. Seventy-one BTC cases (5.5%) had at least one pathogenic variant of the 27 cancer-predisposing genes, indicating hereditary cancer. Pathogenic germline variants in BTCs were significantly ($P < 0.00185$) enriched in *BRCA1* (OR=13.6), *BRCA2* (OR=6.5), *APC* (OR=18.2), and *MSH6* (OR =5.2). *PALB2* variants were marginally associated with BTC ($P=0.01$, OR=5.3). The mean BTC diagnosis age of the carriers was significantly younger than the non-carriers and the carriers had more personal history of breast cancer ($P=0.026$, OR=5.8) and family history of breast cancer ($P=0.047$, OR=2.5). Whole-genome sequencing analysis demonstrated that BTCs with pathogenic germline variants in *BRCA2* and *PALB2*, accompanied with loss of heterozygosity (LOH), displayed homologous recombination DNA repair deficiency phenotypes. Conversely, pathogenic germline variants without LOH in homologous recombination-related genes showed homologous recombination proficient phenotypes. **Conclusions:** This study described the heritability and the actionability of homologous recombination deficiency targeted treatments and provides possibilities for expanding therapeutic strategies such as PARP inhibitor and platinum-based drugs, and screening for BTC.

Cancer Posters - Wednesday

PB1130*. Heterozygous loss of *AURKB* in the germline associates with protection from metastasis and death from cancer

Authors:

L. Ward, W. Broom, C. Willis, A. Deaton, A. Holleman, L. Krohn, R. Hoffing, P. LoGerfo, M. Plekan, P. Nioi; Alnylam Pharmaceuticals, Cambridge, MA

Abstract Body:

Both germline and somatic genetic variants contribute to cancer formation and progression, and identifying these variants has led to the discovery of tumor suppressors and oncogenes, leading to the rational development of chemotherapies. While tumor sequencing and expression studies have been useful at understanding somatic cancer genetics, germline genetic associations from GWAS and family studies are more difficult to interpret: well-powered associations with common variants are not always easily assigned to a gene, and associations with rare variants are best powered to detect risk rather than protection. We took advantage of the unprecedented scale of the UK Biobank exome sequencing data to search for rare coding variants protective against cancer among 397,212 people. For phenotypes, we aggregated cancer at the ICD10 “block” level and used data from the cancer and death registries, inpatient hospitalizations, self-reports, and primary care diagnoses, and included BMI, smoking, and drinking behavior as covariates; for genotypes we aggregated rare (MAF < 1%) variants predicted to be loss of function (pLOF) or predicted damaging missense (CADD > 25). The strongest protective association was between *AURKB* variants (N = 2847 carriers / 397,212 participants) and ICD10 block C76-C80, which includes secondary malignant neoplasms. Looking more closely, we detect protection from having a primary tumor (C00-C97 excluding C77-C79; OR = 0.82; $p = 1.8 \times 10^{-5}$), metastasis (C77-C79; OR=0.58; $p = 9.5 \times 10^{-8}$), and death from cancer (OR = 0.64; $p = 6.4 \times 10^{-5}$). Limiting the analysis to only 97,500 individuals with a primary malignancy diagnosed, we find that conditioned on having cancer, *AURKB* damaging variant carriers (N = 617) have a better prognosis, being less likely to experience metastasis (OR = 0.64; $p = 7.9 \times 10^{-5}$) or death (OR = 0.71; $p = 5.4 \times 10^{-3}$). PheWAS indicates no other phenome-wide significant associations, suggesting *AURKB* as a safe target for cancer. Furthermore, complementary TCGA data show *AURKB* overexpression and duplication associating with poorer prognosis. While small molecules inhibiting *AURKB* and other aurora kinases have showed disappointing efficacy and safety in the clinic, inhibition using other modalities such as RNAi may have therapeutic application in oncology, although additional research is needed.

Cancer Posters - Thursday

PB1131. High tumor mutational burden by whole-genome sequencing in resected non-small cell lung cancer in never smokers is associated with worse prognosis.

Authors:

L-J. Ruel¹, Z. Li¹, N. Gaudreault¹, C. Henry¹, V. Saavedra Armero¹, D. K. Boudreau¹, T. Zhang², M. T. Landi², C. Labbé¹, C. Couture¹, P. Desmeules¹, P. Joubert¹, Y. Bossé^{1,3}; ¹Inst. universitaire de cardiologie et de pneumologie de Québec – Université Laval, Québec, QC, Canada, ²Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, ³Dept. of Molecular Med., Université Laval, Québec, QC, Canada

Abstract Body:

Introduction: Tumor mutational burden (TMB) is a measure of the number of somatic mutations in a tumor. In non-small-cell lung cancer (NSCLC), TMB is a surrogate biomarker of neoantigen load and immunotherapy efficacy. Data are scarce about TMB as a biomarker in never smokers with NSCLC. Using whole-genome sequencing (WGS), we aimed to 1) test whether TMB is a prognostic biomarker of survival after surgery in never smokers with NSCLC, 2) define a prognostic TMB cutoff, and 3) compare TMB derived from WGS with exome sequencing (WES) and commercial NGS gene panels.

Methods: TMB was assessed by WGS and *in silico*-reduced WES and targeted NGS panels in 93 paired tumor-normal samples from never smokers who underwent lung cancer resection with curative intent. Kaplan-Meier and multivariate Cox proportional hazards regression analyses were performed to test for association with survival after surgery and to identify the optimal prognostic TMB cutoff. TMB derived from WGS was compared to *in silico*-reduced sequencing size of WES and targeted NGS panels.

Results: Tumors of never smokers with NSCLC had low TMB scores (median of 1.10 mutations/Mb, range of 0.01-14.71). A TMB cutoff of 1.40 mutations/megabase was associated with a 5-year overall survival of 54% in the high-TMB (31% of cases) compared to 82% in low-TMB patients (log-rank $p = 0.0017$). TMB scores from WGS and WES were highly correlated (Spearman $\rho = 0.91$, $p < 2.2e-16$). TMB scores from NGS panels demonstrated high intra-individual fluctuations and identified high-TMB patients with 54% concordance in average compared to WGS.

Conclusions: In resected NSCLC of never smokers, high TMB was associated with worse prognosis. WES provided a good estimate of TMB while targeted NGS panels seem to lack adequate depth and resolution in the setting of low mutation burden. Further studies are needed to evaluate TMB as a predictive biomarker for immunotherapy in never smokers.

Cancer Posters - Wednesday

PB1132*. High-throughput functional screening of candidate causal variants from lung cancer susceptibility loci

Authors:

H. Patel¹, E. Long¹, A. Golden¹, J. Yin¹, K. Funderburk¹, J. Feng¹, R. Hung², C. Amos³, L. Song⁴, J. Shi⁵, N. Rothman⁶, Q. Lan⁶, International Lung Cancer Consortium, J. Choi¹; ¹Lab. of Translational Genomics, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, ²Lunenfeld-Tanenbaum Res. Inst., Sinai Hlth. System, Univ. of Toronto, Toronto, ON, Canada, ³Inst. for Clinical and Translational Res., Baylor Coll. of Med., Houston, TX, ⁴Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, ⁵Biostatistics Branch, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, ⁶Occupational and Environmental Epidemiology Branch, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD

Abstract Body:

Lung cancer (LC) is the leading cause of cancer deaths globally. While smoking is the principal cause, genetic factors also contribute to LC as shown by 12-21% heritability. GWAS have identified multiple LC risk loci including those that are distinct in the major histological types. However, functional characterization of GWAS loci is challenging because most GWAS variants are non-protein-coding and linkage disequilibrium (LD) limits fine-mapping resolution. Hence, we employed massively parallel reporter assays (MPRA) to test transcriptional activity of risk-associated variants *en masse*. We selected 2,245 variants based on the association *P* values (log likelihood ratio <1000) or LD ($R^2 \geq 0.8$) from 41 risk loci from 3 recent GWAS of Asian and/or European populations (each including >20,000 cases). MPRA library was constructed using 145-bp genomic sequences flanking the variants with each allele in both strands and tagged with 25 unique 12-bp tags. Final library containing 239,800 oligonucleotides was cloned into a luciferase plasmid (input DNA), which was transfected into lung cancer cells representing two histological subtypes (A549: adenocarcinoma; NCI-H520: squamous cell carcinoma). We also incorporated benzo[a]pyrene (BaP) exposure (major tobacco-smoke carcinogen) or DMSO to analyze the impact of smoking on variant function. BaP dosage and exposure duration were optimized based on the induction of known target genes and cell viability. Transfection efficiency was optimized to achieve 25 million transfected cells in each of 5 transfections per treatment for each cell line. Regulatory potential of variants was evaluated by tag sequencing and counting in DNA input and mRNA output. Our QC results showed high input DNA tag recovery (>98%) in MPRA mRNA and high inter-transfection correlation ($r > 0.8$) indicating reproducibility among the replicates. At FDR <0.01, we identified 509 and 525 variants (DMSO and BaP groups combined) with significant allelic transcriptional activities in A549 and H520 cells, respectively (~23% of tested variants). Moreover, we found 101 and 93 variants only significant in BaP exposed A549 and H520 cells, respectively, highlighting potential effect of tobacco smoke exposure on variant function. While 190 of the MPRA significant variants were common in both cell types, 319 and 335 variants were unique to A549 and H520, respectively, indicating distinct variant function based on the histological context. Thus, MPRA with relevant cell type and exposure prioritized functional variants, which can be further analyzed by lung-eQTL and chromatin interaction data to identify new target genes and elucidate genetic basis of LC.

Cancer Posters - Thursday

PB1133. Identification of key molecular mechanisms in IDH mutant brain tumors to enable precise risk stratification

Authors:

S. Bhattacharya¹, A. Dawood¹, H. Barseghyan², E. Vilain³, M. Bornhorst⁴; ¹Children's Natl. Hosp., Washington DC, DC, ²Childrens Natl. Med. Ctr., Washington, DC, ³Children s Natl. Med. Ctr., Washington, DC, ⁴Children s Natl. Hosp., Washington, DC

Abstract Body:

Isocitrate dehydrogenase 1 and 2 (IDH) mutations are the most prevalent genetic alterations in adult low- to intermediate grade diffuse astrocytomas, anaplastic astrocytomas, and oligodendrogliomas. Previous studies of adult IDH-mutant tumors have revealed unique tumor-specific signatures, allowing for stratification based on molecular phenotype and clinical presentation. However, very little is known about pediatric IDH-mutant tumors, resulting in fragmented clinical care. To perform precise molecular profiling of IDH-mutant brain tumors to predict patient outcomes and guide potential future therapeutic developments First, we harmonized data from nine pediatric and adult brain tumor datasets, along with our pediatric IDH-mutant tumor samples, to determine the genetic and epigenetic profiles of the brain tumors. We then performed a comprehensive and integrated analysis of genetic, epigenetic, transcriptomic and demographic data from IDH-mutant tumors to identify co-occurring somatic/germline and epigenetic changes within the tumors. Finally, we determined the impact of genomic, epigenomic and demographic parameters on patient outcomes. Using exome sequencing we identified mutations of tumor suppressors like *TP53*, *ATRX* and *NF1*, co-occurring with *IDH* mutations. Additionally, we also identified genes like *MUC3A* and *MUC16*, which has not been previously found to co-occur with *IDH* mutations. Combining the genomic, with the transcriptomic and epigenetic data, we would enable precise risk stratification that can be used to optimize treatments for all patients with IDH-mutant tumors.

Cancer Posters - Wednesday

PB1134. Identification of target genes and pathways involved in triple-negative breast cancer proliferation and stemness using genome-wide CRISPR screens.

Authors:

T. Li¹, G. Lu², W. Wang³, W-Y. Chan⁴, Z. Zhao¹; ¹The Chinese Univ. of Hong Kong, Hong Kong, China, ²Sch. of BioMed. Sci., Hong Kong, China, ³The Chinese Univ. of Hong Kong, Hong Kong, Hong Kong, ⁴The Chinese Univ. of Hong Kong, Shatin, China

Abstract Body:

Identification of target genes and pathways involved in triple-negative breast cancer proliferation and stemness using genome-wide CRISPR screens. Breast cancer is the leading cause of cancer death among women. Unlike other subtypes of breast cancer, triple-negative breast cancer (TNBC) is acknowledged as one of the most aggressive and refractory subtypes of breast cancer that accounts for most of breast cancer-related mortality. For decades, the therapeutic strategies of TNBC were limited to several traditional treatments, which can only eliminate the bulk tumor population. However, the remaining core population - cancer stem cells (CSCs) can develop new tumor tissues with a few cells to initiate disease relapse and invade other organs that contribute to cancer metastasis. Advanced targeted strategies - immunotherapies provide exciting prospects in TNBC treatment. However, it is paramount to identify cancer-specific antigens in advance. Genome-wide CRISPR-Cas9 screens have proven to be an exquisitely powerful method for discovering and functional annotation of genetic or epigenetic elements in biological processes. This study used two CRISPR-Cas9 screening libraries to perform high-throughput pooled screens in TNBC cell lines. One is the human membrane protein activation library, which aims to identify candidate genes encoding antigens that can serve as tumor-specific targets in immunotherapy. The other is the human epigenetic knockout library, which we hope can reveal the epigenetic “drivers” or driver genes regulated by epigenetic factors in TNBC. After weeks of consecutive culture, gene-edited cells were collected, and skewed gRNA distribution in the population will indicate essential genes in cancer cell proliferation. Non-cancer stem cells and cancer stem cells (mesenchymal-like, CD44+/CD24-; epithelial-like, ALDH1+) were separated by FACS, genes critical to stemness transformation will also be screened out. Many known TNBC-specific or promoting genes, such as *SUZ12*, *RAD51* and *PELPI*, were among the top ranks in our screening results. These coincidences confirm the biological reliability of our findings. KEGG pathway analysis showed most candidate genes were involved in cell cycle, NF- κ B, and thyroid hormone signaling pathways. The GEPIA database and Kaplan-Meier method also show that many top hits are tumor-specific and critical for patient survival. In the end, tissue microarray will be used to determine the clinical significance of candidates. Overall, our results provide essential genetic and epigenetic targets for developing TNBC treatments.

Cancer Posters - Thursday

PB1135. Identification of *USP9X* as a leukemia susceptibility gene.

Authors:

S. Sisoudiya^{1,2}, H. Li³, J. Schraw², M. Scheurer², A. Sabo³, P. J. Lupo², S. E. Plon^{1,2,3}; ¹Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, ²Section of Hematology-Oncology, Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, ³Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX

Abstract Body:

Introduction: Data from the Genetic Overlap Between Anomalies and Cancer in Kids (GOBACK) Study (Lupo et al., 2019) indicated that children with multiple non-chromosomal anomalies (MCAs) were 5.9-times (95% CI 5.3-6.4) more likely to develop childhood cancer compared to children without these conditions. Based on this, we hypothesize that molecular characterization of children with MCAs and cancer can reveal novel cancer predisposition syndrome genes.

Method: Whole genome sequencing (WGS) was performed on DNA extracted from saliva samples of GOBACK families. Structural variants (SVs) were called using five callers. A consensus call was defined by greater than 50% reciprocal overlap and made by at least 2 callers. We also analyzed RNA-seq and overall survival data of 142 B-cell acute lymphoblastic leukemia (B-ALL) tumor samples in the TARGET program.

Results: SV analysis of WGS data (n = 20 trios) identified a novel 5kb *de novo* heterozygous inframe deletion in *USP9X* in a female proband with MCAs and B-ALL. Her phenotype was consistent with female-restricted X-linked syndromic intellectual developmental disorder-99 (MRXS99F). We extracted genotype and phenotype data from 42 female probands in the literature with *USP9X* germline variants, two of whom had B-ALL. The patient with the inframe deletion had anomalies similar to the pattern in subjects with loss-of-function variants. Notably, the cumulative incidence of B-ALL among female patients with germline variants in *USP9X* below the age of 8 years (14%) is significantly higher than an age- and sex-matched cohort (0.002%) from the Surveillance, Epidemiology, and End Results (SEER) database (P<0.0001, log-rank test). In sporadic B-ALL, somatic loss of function *USP9X* mutations occur, consistent with a tumor suppressor mechanism. *USP9X* encodes a highly conserved deubiquitinating enzyme. Expression analysis of high-risk sporadic B-ALL patients demonstrates that pathways such as cell cycle, TGF-beta signaling are associated with changes in *USP9X* expression. Overall survival data shows that high-risk ALL patients with lowest quartile *USP9X* expression had significantly poorer overall survival compared to patients with top quartile expression (HR = 2.00; 95% CI: 1.09 - 3.67; P = 0.02), which was validated in an independent B-ALL cohort.

Conclusion: Overall, we describe *USP9X* as a novel leukemia susceptibility gene associated with multiple congenital anomalies and B-ALL. *USP9X* also serves as a tumor suppressor in sporadic B-ALL where its low expression could be used to inform clinical management of high-risk B-ALL patients.

Cancer Posters - Wednesday

PB1136*. Identifying molecular markers of early stage ovarian cancer pathogenesis through whole genome CRISPR screening of mutant *BRCA1* fallopian tube cells.

Authors:

S. Dhungana¹, K. Ayaluri¹, F. Dezem¹, S. Chen^{1,2}, B. Davis^{1,2}, J. Plummer^{1,2}, S. Gayther^{1,2}; ¹Ctr. for Bioinformatics and Functional Genomics, Dept. of BioMed. Sci., Cedars-Sinai Med. Ctr., Los Angeles, CA, ²Applied Genomics, Computation and Translational Core, Cedars-Sinai Cancer, Cedars-Sinai Med. Ctr., Los Angeles, CA

Abstract Body:

An accumulation of genetic alterations transforms normal fallopian tube cells into high-grade serous ovarian cancers (HGSOC). A lack of understanding of the earliest stage molecular mutations in disease development continues to be an impediment for the prevention and care of the disease. Somatic *TP53* mutations are almost ubiquitous HGSOCs, acting as a driving factor for the dysfunction of critical biological processes involved cell division and apoptosis. *BRCA1* mutations predispose to HGSOC and usually occur in conjunction with *TP53* alterations. *BRCA1* functions in the DNA damage repair pathway and it is the breakdown of this process that leads to characteristic accumulation of gross interstitial alterations throughout HGSOC development. In the current study, we created early stage models of fallopian tube epithelial cells carrying both *TP53* and *BRCA1* pathogenic mutations and then used whole genome CRISPR screening of 19,500 genes targets in these models to identify genes associated with early stage neoplasia. We performed this negative selection CRISPR screen in FTE cells that were *BRCA1* wild-type (FT282^{*BRCA1:WT*}) *BRCA1* heterozygous (FT282^{*BRCA1:+/-*}) and *BRCA1* homozygous (FT282^{*BRCA1:-/-*}). Using the ‘Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout’ (MAGeCK) method we identified genes that were enriched and depleted in FTE cells engineered to express different *BRCA1* mutations including several known cancer genes including *PAWR*, *AMOTL2*, *MDM2*, *MYH9*. We next integrated transcriptomic and proteomic profiling in these cell lines to further substantiate the genes found from MAGeCK analysis including: *PAWR*, a regulator of *WT1* (master transcription factor of HGSOC) that downregulates the anti-apoptotic protein *BCL2*; *MYH9*, which plays role in cytokinesis, cell shape and regulates *TP53* stability and localization; *AMOTL2*, which inhibits the *Wnt*/beta-catenin signaling pathway and plays role in polarity, proliferation, and migration; *C16ORF72*, a identified binding partner for *HUWE1*, which is a DNA damage response gene with a role in ubiquitin mediated degradation of *TP53*; and the *Myc* oncogene which is frequently amplified in ovarian and other cancer type. In conclusion, this highly novel study to screen the entire repertoire of coding genes in a single assay has identified multiple target genes that represent new candidates for the early stage detection and diagnosis of HGSOC, chemoprevention and targeted treatments and provides unique opportunities to identify common vulnerabilities for the three most frequent oncogenic alterations in fallopian tube epithelium.

Cancer Posters - Thursday

PB1137. Identifying strong instrumental variables for pleiotropic assessment in Mendelian randomization: CALGB/SWOG 80405 (Alliance).

Authors:

A. Yazdani¹, H-J. Lenz², G. Pillonetto³, R. Mendez-Giraldez⁴, M. Bertagnolli⁵, A. P. Venook⁶, M. J. Ratain⁷, N. Rashid⁸, B. G. Vincent⁹, X. Qu¹⁰, A. Yazdani¹¹, Y. Wen¹², W. F. Symmans¹³, A. B. Nixon¹⁴, M. Kosorok⁸, C. M. Perou^{15,16}, F. Innocenti¹; ¹Div. of Pharmacotherapy and Experimental Therapeutics, Eshelman Sch. of Pharmacy, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ²USC Norris Comprehensive Cancer Ctr., Los Angeles, CA, ³Dept. of Information Engineering, Univ. of Padova, Padova, Italy, ⁴Biostatistics and Computational Biology Branch, Natl. Inst. of Environmental Hlth.Sci., Durham, NC, ⁵Dana-Farber/ Partners Cancer Care, Harvard Med. Sch., Boston, MA, ⁶Univ. of California at San Francisco, San Francisco, CA, ⁷Div. of the Biological Sci., Univ. of Chicago, Chicago, IL, ⁸Dept. of Biostatistics, Gillings Sch. of Global Publ. Hlth., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁹Dept. of Microbiol. and Immunology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ¹⁰Genentech, South San Francisco, CA, ¹¹Ctr. of Perioperative Genetics and Genomics, Dept. of Anesthesiology, Perioperative and Pain Med., Brigham & Women's Hosp., Harvard Med. Sch., Boston, MA, ¹²Alliance for Clinical Trials in Oncology, Chicago, IN, ¹³Dept. of Pathology, Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, ¹⁴Duke Ctr. for Cancer Immunotherapy, Duke Univ., Durham, NC, ¹⁵Lineberger Comprehensive Cancer Ctr., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ¹⁶Dept. of Genetics, Sch. of Med., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract Body:

Integrating genotype, gene expression, and clinical outcome data by Mendelian randomization (MR) technique provides potential opportunities for uncovering causal mechanisms underlying disease etiology. In this context, fulfilling MR assumptions on genetic variants as instrumental variables (IVs) are essential steps to yield unbiased results. Therefore, here we proposed an approach for selecting valid IVs to perform MR analysis. This approach allows for multiple IVs selected from expression quantitative trait loci (eQTL) analysis. We used a predictive model to assess if the accumulated effects of selected *cis*-eQTLs on gene expression can control unobserved sources of variability in investigating gene expression and outcome relationships. The predictive model accounts for linkage disequilibrium and interconnectivity among genes since it is a multivariable polygenic model. If the predicted gene expression significantly affects the outcome, the accumulated effect of IVs is strong. We considered no pleiotropy for the IVs with a *p*-value > 0.01 in GWAS analysis and potential pleiotropic effect for the IVs with a *p*-value < 0.01. To test the latter's pleiotropy, we simultaneously estimated the effects of IVs and gene expressions on the outcome. If only the effect of IVs were significant in the model, it was an indicator of the pleiotropic action of the IVs.

We used tumor RNA-seq and germline genotype data from 1,284 patients with metastatic colorectal cancer treated either with cetuximab or bevacizumab monoclonal antibodies through a randomized phase III trial (CALGB/SWOG 80405). Cetuximab targets epidermal growth factor receptor (EGFR) and bevacizumab vascular endothelial growth factor (VEGF). We found genes causally associated with the overall survival (OS) and IVs that have both direct and indirect effects on OS, specific to each treatment. For instance, we found *WDR62* (*p*-value: 0.639 vs. 0.019) and *SCD5* (*p*-value: 0.056 vs. 0.47) genes with unfavorable putative causal effects on OS in patients treated with cetuximab and bevacizumab, respectively. *SCD5*, as a key regulator of energy metabolism, highlighted the possible risk of dyslipidemia in the use of vascular endothelial growth factor (VEGF) inhibitors. *WDR62* has shown the oncogenic role across a series of human tumors, including colorectal cancer. We replicated the findings using a replication set for each treatment, which proved the power of the proposed approach in providing

ASHG 2022 Annual Meeting Poster Abstracts

reproducible results.

Support: U10CA180821, U10CA180882, U24CA196171, <https://acknowledgments.alliancefound.org>;
U10CA180888 (SWOG); Lilly, Genentech, and Pfizer; ClinicalTrials.gov Identifier: NCT00265850

Cancer Posters - Wednesday

PB1138. Impact of polygenic risk, pathogenic variants, and family history on detection of breast cancers within and outside the national mammography screening programme on 117,252 women over 1992-2019

Authors:

N. Mars^{1,2}, S. Kerminen², M. Pirinen^{2,3,4}, E. Salminen^{5,6}, K. Aaltonen⁷, T. Meretoja^{8,9}, S. Heinavaara¹⁰, FinnGen, E. Widen², S. Ripatti^{2,4,1}; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²Inst. for Molecular Med. Finland, FIMM, HiLIFE, Univ. of Helsinki, Helsinki, Finland, ³Helsinki Inst. for Information Technology HIIT and Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland, ⁴Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland, ⁵Dept. of Clinical Genetics, HUSLAB, HUS Diagnostic Ctr., Univ. of Helsinki and Helsinki Univ. Hosp., Helsinki, Finland, ⁶Dept. of Med. Genetics, Univ. of Helsinki, Helsinki, Finland, ⁷Dept. of Clinical Genetics, Helsinki Univ. Hosp., Helsinki, Finland, ⁸Breast Surgery Unit, Comprehensive Cancer Ctr., Helsinki Univ. Hosp., Helsinki, Finland, ⁹Univ. of Helsinki, Helsinki, Finland, ¹⁰Finnish Cancer Registry and Cancer Society of Finland, Helsinki, Finland

Abstract Body:

Polygenic risk scores (PRS) for breast cancer are effective for stratifying women for their breast cancer risk. They show potential for risk-stratified screening, but there is limited evidence on the impact of PRSs on interval and screen-detected breast cancer in population-based data. Similarly, although family history and pathogenic variants (PV) are applied for risk-tailored screening in selected individuals, limited data about their utility on a national level exists. In Finland, women are biennially screened nationwide with mammograms between ages 50 and 69. Using FinnGen, combining the national Mass Screening Registry for breast cancer (780,091 screening invitations 1992-2019) with imputed genotypes and the Finnish Cancer Registry, we perform a comprehensive assessment of the impact of a genome-wide PRS (PRS-CS, 1,079,089 variants), Finnish-enriched pathogenic variants, and family history on breast cancer detection within and outside the screening program for 117,252 screening-aged women (12,965 cases).

The higher the PRS, the higher was the positive predictive value (PPV) of breast cancer screening: at age 50, the PPV in the highest vs lowest PRS deciles were 24.1% and 6.0%, and at age 60, 49.0% and 14.7%. Similar PPVs were detected for carriers of PVs and for family history, and combinations of these three risk factors led to PPVs up to 55.2%. No excess of benign breast lesions was detected with any of the risk factors. All performed similarly for in situ and invasive cancers. A high PRS (>90%) conferred a particularly high risk for bilateral breast cancer (HR 4.85, 95%CI 3.28-7.17), with similar risks for PVs and positive family history. At all ages, women with a high PRS with a negative screening finding had an elevated risk of being diagnosed with an interval breast cancer, and a screen-detected cancer in the next screen. This indicates that women with high PRS could benefit from more frequent screens. The proportion of women with a lymph node-positive breast cancer at diagnosis was similar for the risk factors (range 41.8-42.9% for interval-detected, 29.2-32.3% for screen-detected).

Our findings demonstrate benefits of risk-based screening in breast cancer. Risk-tailored screening has long been employed for subgroups of women with family history and/or PVs. Our real-life data provide evidence that it would be beneficial to use PRSs - the risk factor most common of the three - together with PVs and family history for optimal risk-based screening. Although these risk factors were compelling for risk stratification within the Mass Screening Registry, the optimal strategies for a cost-effective, risk-tailored breast cancer screening remain to be studied.

Cancer Posters - Thursday

PB1139. Implementation of new ClinGen/CGC/VICC recommendations for classification of oncogenicity of somatic variants using AI

Authors:

Y. Einhorn, O. Shani; Genoox, Palo Alto, CA

Abstract Body:

Introduction: The ClinGen, CGC and VICC recently published new recommendations for the classification of pathogenicity of somatic variants in cancer, in order to create a set of standards when classifying the oncogenicity of a somatic variant. While the detailed evaluation criteria will increase classification consistency between different labs, the process is time-consuming and holds some computational challenges. Here, we present a novel AI-based oncogenicity classification engine for somatic variants based on these new recommendations and integrated it into Franklin (franklin.genoox.com), an open-access platform for variant interpretation.

Methods: An AI oncogenic classification engine was developed and integrated into Franklin for implementing the new recommendations and removing their computational and technical challenges. We evaluated the classifications against the dataset which was classified by experts during the process of creating these recommendations, which included 94 variants in 10 cancer-related genes.

Results: Comparing the results shows high concordance without any strong conflict between the experts and Franklin. 100% (34/34) of the Oncogenic/Likely-Oncogenic variants were classified as Oncogenic/Likely-Oncogenic, 100% (7/7) of the B/LB variants were classified as B/LB/VUS-Leaning-Benign (VUS-B). 74% (23/31) of the VUS variants were classified as VUS, VUS-Leaning-Oncogenetic, or VUS-B. Of the 8 VUS conflicts, 2 were classified as B/LB and 6 were classified as Oncogenic/Likely-Oncogenic, suggesting that there might be additional evidence to be considered.

Conclusions: Herein we demonstrated how Franklin can aid in following the new oncogenic classification recommendations, a task that is challenging, and time-consuming. To the best of our knowledge, this is the only solution that currently exists today.

Cancer Posters - Wednesday

PB1140. Implementation, utilization, and diagnostic outcomes of a technology-enabled telehealth-based program for individuals to access multi-cancer early detection testing

Authors:

S. Weissman; Genome Med., South San Francisco, IL

Abstract Body:

Background: Multi-cancer early detection (MCED) is a novel method to screen for many different cancers by detecting circulating tumor DNA in blood. Upon its commercial availability in 2021, we established two nation-wide telehealth programs that rely on digital tools to provide access to MCED in an efficient and scalable manner: 1) a patient-initiated program (PIP) and 2) an employer benefit program (EBP). Qualification for MCED was assessed via a digital tool for both programs, but the EBP required a consultation with a genetic counselor (GC) for individuals <50 years of age.

Purpose: We investigated the implementation, utilization, and diagnostic outcomes of a technology-enabled telehealth program that provides access to MCED.

Methods: We performed a chart review of consecutive patients who presented for MCED from June 2021 - April 2022 in two separate programs - 1) an EBP for individuals aged 50 or older and those that were 18-49 years of age and had a GC consultation to determine eligibility based on personal and family history; 2) a PIP for individuals aged 40 or older through an automated questionnaire that assessed eligibility based on cancer risk factors. We extracted patient sex, age, risk criteria met, personal cancer history, test completion, and MCED results.

Results: 99% (4909/4963) of people completed the screening process by finishing the questionnaire or having a GC consultation. 59% (2919/4909) and 41% (1990/4909) came through the PIP and EBP, respectively. 59% were male (n=2896) and the mean age was 54 (SD=12). Overall, 78% (3810/4909) qualified for MCED. Of those, 78% (2960/3810) qualified based on age alone (50 or older). The remaining individuals (n=850) qualified because of a family history of cancer (790, 93%), smoking history (128, 15%), HIV infection (9, 1.1%), and/or previous solid organ transplant (5, 0.6%). Through the EBP, 23% (459/1990) of individuals met with a GC and of those, 55% (253/459) qualified for MCED. To date, 76% (n=2910) of individuals completed MCED, 12% (n=462) have testing pending, and 11% (n=438) never completed or canceled testing. 18 of the 2910 individuals with a completed test had a cancer signal detected (0.61%); mean age was 63 years (SD=8.7) and the majority were male (78%; 14/18). 50% (9) were diagnosed with cancer, 22% (4) were likely false positives, and 5 (28%) are still undergoing a clinical work-up.

Conclusion: MCED testing offered through a nation-wide telemedicine program had high utilization and completion of both the digital tool and testing. This suggests that a tech-enabled telehealth model can be used successfully to provide access to MCED with minimal barriers.

Cancer Posters - Thursday

PB1141. Inference on the genetic architecture of breast cancer risk

Authors:

Y. Yasui¹, W. Letsou², F. Wang³, C. Im⁴, Y. Sapkota⁵, Z. Wang², S. Mirzaei Salehabadi¹, J. Baedke⁶, W. Moon⁵, Q. Liu⁴, L. Robison⁶, J. Martinez⁷; ¹St Jude Children's Res. Hosp., Memphis, TN, ²St. Jude Children's Res. Hosp., Memphis, TN, ³St. Jude Children's Res. Hosp., Memphis, TN, ⁴Univ. of Alberta, Edmonton, AB, Canada, ⁵St Jude Children's Res. Hosp., Memphis, TN, ⁶St. Jude Children's Res. Hosp., Memphis, TN, ⁷The Polytechnic Univ. of Catalonia, Barcelona, Spain

Abstract Body:

What are the major determinants of women's risk of developing breast cancer? Rare mutations such as those in the *BRCA1/2* genes, common alleles identified by genome-wide association studies, or non-genetic factors? The largest population-based twin study of cancer, the Nordic Twin Study of Cancer (NTSC), with 3,933 breast cancer cases among 21,054 monozygotic and 30,939 dizygotic female twin pairs, provided exceptional data from which we can draw inferences regarding this question. Specifically, there were three key observations in NTSC: (#1) the average lifetime breast cancer risk, approximately 8%, does not differ by twin zygosity; (#2) the mean time interval between diagnoses when both twins develop disease (i.e., the disease concordance) also does not differ by twin zygosity; but, (#3) conditioned on one twin's developing disease, the incidence rate in the co-twin is approximately 1% per year if the pair are monozygotic and 0.5% per year if they are dizygotic. Assuming that non-genetic risk factors are shared similarly between twins regardless of their zygosity, we can draw two conclusions from these observations. First, (#1) and (#3) imply that the major determinant of risk is in the germline DNA: conditioned on the development of any breast cancer in one twin (not limited to "familial breast cancer"), the incidence rate in the co-twin is several-fold higher than the average risk (about 8% in lifetime) if the pair is monozygotic, and only half as much higher if dizygotic. Second, the two-fold difference in the conditional incidence rate between zygositys (#3) seemingly conflicts with the equality of the mean inter-twin disease intervals in disease concordance between zygositys (#2). This can be resolved, however, if rare variants, not common variants, in the germline DNA elevate the risk. That is, the two-fold difference in the conditional incidence rate between zygositys reflects the 50:50 chance of a dizygotic co-twin's carrying or not carrying one of the high-risk rare variants, and only when (like monozygotic twins) both twins carry the variant and develop disease (disease concordance), the inter-twin disease interval can be calculated (hence the equality of the mean interval between zygositys). These conclusions are consistent with the hypothesis of Peto and Mack, which states that only a subgroup of women carry the risk and that this risk is high, and that of McClellan and King, which states that breast cancer risk is defined by genetic heterogeneity with many rare underlying mutations. Our presentation details the deductive reasoning that derives these conclusions and draws a critical inference regarding breast cancer etiology.

Cancer Posters - Wednesday

PB1142*. Integrating GWAS and 3D chromatin interactome data to identify multi-cancer risk genes in hormone-related cancers.

Authors:

S. Rivera^{1,2}, J. Beesley¹, H. Sivakumaran¹, M. Moradi Marjaneh^{1,3}, K. Hillman¹, S. Nair¹, S. Edwards^{1,2}, J. French^{1,2}; ¹QIMR Berghofer Med. Res. Inst., Brisbane, Australia, ²Queensland Univ. of Technology, Brisbane, Australia, ³UK Dementia Res. Inst. at Imperial Coll., London, United Kingdom

Abstract Body:

Breast, endometrial, ovarian and prostate are hormone-related cancers that account for over 15% of all cancer deaths. To date, GWAS has identified 150 breast, 16 endometrial, 37 ovarian and 167 prostate cancer-associated genetic variants. In this study, we observed that those variants are often in close proximity suggesting common mechanisms underlying these cancers. We identified 45 multi-cancer risk regions (MCRRs) defined as genomic regions that contain variants associated with risk of two or more hormone-related cancers. Disease-associated variants frequently fall in DNA regulatory elements, such as enhancers, that can lie up to 1Mb from their target gene promoter. Therefore, we hypothesized that within MCRRs, genetic variants associated with different cancer types modulate the expression of common target genes through altered tissue-specific regulatory elements. To enrich for chromatin interactions between gene promoters and enhancers within MCRRs in relevant cell types, we performed promoter capture HiC (PCHiC) in twelve immortalised 'normal' and cancer cell lines. To identify MCR genes we integrated PCHiC interactions and risk-associated SNPs within MCRRs. We identified 78 candidate MCR genes associated with at least two hormone-related cancers. This list includes established cancer driver genes such as MYC and CCND1, but also >20 genes with little or no reported role in cancer. One example is MLLT10, a candidate breast and ovarian cancer risk gene. MLLT10 encodes a transcription factor that interacts with DOT1L, which methylates H3K79, a histone mark associated with active transcription. Reporter assays show risk SNPs induced MLLT10 promoter activity, suggesting increased MLLT10 contributes to cancer risk. CRISPR-based functional studies are currently underway to elucidate the role of MLLT10 in breast and ovarian cancer development. We anticipate that some of the identified multi-cancer risk genes may provide new drug targets for future prevention or treatment of hormone-related cancers.

Cancer Posters - Thursday

PB1143. Integrative data analysis to uncover genes and pathways underlying the aggressiveness of lung cancer brain metastasis and the overall prognosis

Authors:

R. Asakereh^{1,2}, Q. Zhang^{1,3}, P. Shoostari^{1,2,4}, ¹Univ. of Western Ontario, London, ON, Canada, ²Children's Hlth.Res. Inst., Victoria Hosp., London Hlth.Sci. Ctr., London, ON, Canada, ³Univ. Hosp., London Hlth.Sci. Ctr., London, ON, Canada, ⁴Ontario Inst. for Cancer Res., Toronto, ON, Canada

Abstract Body:

Introduction. Recent advances in cancer diagnosis and treatment have resulted in better management of the mortality caused directly by tumors at primary sites. However, this on the other hand, has increased the proportion of observed metastatic cases among cancer patients. In particular, we observe a relatively high portion of brain metastasis with primary lung cancer. The ability to predict the prognosis of the primary tumor and the prediction of how rapidly the metastasis may occur, would result in providing better treatment decisions. As a step towards this goal, we need to identify the specific molecular mechanisms underlying fast/slow metastasis in different patients. Current studies in this area have several drawbacks, including neglecting the heterogeneity of the tumor cells, and/or focusing on a limited number of genes rather than a whole transcriptome analysis. In our study, we have addressed these drawbacks and uncovered several genes and molecular processes affecting the prognosis of the lung tumor brain metastasis.

Methods. We have collected a dataset which contains the spatially profiled gene expression of three different organs (lung, brain, and lymph nodes) in 35 patients using Nanostring Digital Spatial Profiler. We followed a data integration method based on Multi-Omics Factor Analysis (MOFA) to discover the molecular sources of heterogeneity in the patients and applied robust statistical tests to identify the factors that contribute to prognosis differences between patients. Then we identified a set of highly relevant genes and performed KEGG pathways analysis to identify the underlying pathways.

Results. Our analysis indicated that there are three major categories of pathways likely to be implicated in lung cancer patients with fast brain metastasis. They include (a) immunity-related pathways, (b) stemness related pathways such as focal adhesion and ECM receptor interaction, and (c) metabolic pathways such as Warburg effect and oxidative phosphorylation. The tumor associated immune related pathways are twofold: (1) the impaired immune system in tumor suppression such as neutrophil extracellular traps; and (2) the enhanced ability of the tumor cells to escape the immune surveillance, such as antigen processing and presentation.

Conclusions. Our analyses predict target pathways underlying the brain metastasis in lung cancer patients and establish a means for the prediction of the lung tumor prognosis using molecular experiments.

Cancer Posters - Thursday

PB1144. Investigating the link between cholesterol-lowering genes and risk of skin cancers

Authors:

J. Ong, J-C. Dusingize, C. M. Olsen, M. Law, N. Pandeya, R. E. Neale, D. C. Whiteman, S. Macgregor; QIMR Berghofer Med. Res. Inst., HERSTON, Australia

Abstract Body:

Epidemiologic studies have yet to resolve whether alterations in serum lipids levels influence skin cancer risk. The evidence from observational studies and clinical trials are not concordant. To clarify whether there is genetic evidence supporting a potential causal link, we adopted a series of Mendelian randomization (MR) approaches to evaluate the effects of genetically predicted low-density lipoprotein-cholesterol (LDL-C) levels on the risk of developing several types of common skin cancers, namely cutaneous melanoma, basal cell carcinoma (BCC) and squamous cell carcinoma (SCC).

In our study, we considered two different approaches. The first approach utilizes an array of summary-data based MR techniques, where we relied on LDL-C GWAS summary statistics identified from the Global Lipids Genetics Consortium (GLGC) and GWAS summary statistics for melanoma, BCC and SCC directly obtained from the latest European GWAS meta-analyses of more than half a million participants. The second approach is based on conventional two-stage-least square estimates. Here we calculated the polygenic risk scores (PRS) for LDL-C and performed a 2SLS MR analyses within the UK Biobank, including a stratified analysis specifically on statin-free users. The MR association between the KC PRS and each cancer site was estimated using logistic regression adjusted for standard covariates such as age, sex and genetic principal components.

Based on the pooled estimate derived from the 77 LDL-associated variants, there was limited evidence of an effect of genetically-predicted LDL-C on the risk of developing skin cancer (OR, 95% CI: 0.99, 0.92-1.05 for melanoma, 1.03, 0.96-1.10 for BCC and 1.06, 0.97-1.16 for SCC). In supplementary analyses, the ORs and their 95% CI were 1.03, 0.98-1.07 for melanoma, 1.01, 0.99-1.03 for BCC and 1.01, 0.96-1.07 for SCC. Results did not change when we restricted our analyses to variants in lipid-lowering drug targets gene regions or when we evaluate statin-free users within the UK Biobank cohort. Taken together findings from both summary-based and individual-level MR analyses, our study does not support a potential role of LDL-C-lowering genetic variants on the risk of developing skin cancer.

Cancer Posters - Wednesday

PB1145. Isolated multinodular goiter in a family with DICER1 heterozygous germline variant

Authors:

P. Dalal^{1,2}, G. Leone^{1,2}, R. Rabenn³, A. Parikh⁴, L. Ponsky⁵, C-H. Wu¹; ¹Dept. of Genetics and Urology, Case Western Reserve Univ. and Univ. Hosp., Cleveland, OH, ²Northeast Ohio Med. Univ., Rootstown, OH, ³Univ. Hosp., Cleveland, OH, ⁴Univ Hosp. Case Med Ctr., Cleveland, OH, ⁵Dept. of Urology, Case Western Reserve Univ. and Univ. Hosp., Cleveland, OH

Abstract Body:

Background

Germline pathogenic variant of DICER1 (OMIM *606241, 14q32.13) is a monogenic cause of multinodular thyroid goiter 1, with or without Sertoli-Leydig cell tumors (OMIM #138800) and pleuropulmonary blastoma (PPB) (OMIM #601200). The inheritance pattern is autosomal dominant with incomplete penetrance. The other manifestations include GLOW (OMIM #618272) and embryonal rhabdomyosarcoma 2. GLOW encompasses global developmental delay, lung cysts, overgrowth and Wilms tumor. Embryonal rhabdomyosarcoma (OMIM #180295) presents with uterine cervix embryonal rhabdomyosarcoma and variant carriers also presented with multinodular goiter, Sertoli-Leydig cell tumors, and Wilms tumor. Most tumors onset before the age of 40, with PPB presenting infants and children less than 6 years. 32% of women and 13% of men with a DICER1 variant will be diagnosed with multinodular thyroid goiter and/or have undergone a thyroidectomy by age 20. We present a family with confirmed DICER1 pathogenic variant c.5528-2del who presents with development of thyroid goiters.

Case Report

A 44 year old female is referred to genetics with a history of total thyroidectomy due to development of multinodular goiters in adolescence. Family history is significant for lung cancer in the paternal grandmother, colon cancer in her maternal mother, and lymphoma in her brother. The patient is otherwise healthy without other concerning signs or symptoms. The patient's 14-year-old son has a past medical history of aortic dilation, autism, cardiac murmur, and multinodular thyroid goiter. The patient's 12-year-old daughter with the same paternity has a past medical history of confirmed nontoxic, multinodular goiter. Both children plan to undergo a thyroidectomy. The mother, the 14-year-old son, and the 12-year-old daughter have a heterozygous DICER1 c.5528-2del variant. This variant has not been reported in any literature or database before and is isolated to these three individuals. Serial workups for other DICER1 syndromes for these 3 individuals are negative.

Discussion

We present a family with confirmed DICER1 variant of c.5528-2del who presents with development of only thyroid goiters without other typical DICER1 presentations. We would like to present this specific variant, c.5528-2del, to the genetic community as a likely association with isolated multinodular thyroid goiter. Our report may be useful for future genotype-phenotype correlations of DICER1. The family plans to actively monitor for the development of other DICER1 presentations by following with multidisciplinary care.

Cancer Posters - Thursday

PB1146. Large-scale whole genome sequencing reveals that APOBEC mutagenesis is a common process in normal human small intestine

Authors:

Y. Wang¹, P. S. Robinson^{1,2}, T. H. H. Coorens^{3,1}, L. Moore¹, H. Lee-Six¹, A. Noorani¹, M. Sanders¹, R. Katainen⁴, R. Heuschkel⁵, R. Brunton-Sim⁶, R. Weston⁷, D. Read⁷, B. Nobbs⁷, R. C. Fitzgerald⁸, K. Saeb-Parsy⁹, I. Martincorena¹, P. J. Campbell¹, S. Rushbrook^{6,10}, M. Zilbauer^{5,2}, S. J. A. Buczacki¹¹, M. R. Stratton¹; ¹Wellcome Sanger Inst., Hinxton, United Kingdom, ²Dept. of Paediatrics, Univ. of Cambridge, Cambridge, United Kingdom, ³Broad Inst. of MIT and Harvard, Boston, MA, ⁴Applied Tumor Genomics Res. Program, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, ⁵Dept. of Paediatric Gastroenterology, Hepatology and Nutrition, Addenbrooke's Hosp., Cambridge, United Kingdom, ⁶Norfolk and Norwich Univ. Hosp., Norwich, United Kingdom, ⁷NIHR Clinical Res. Network: East of England, Addenbrooke's Hosp., Cambridge, United Kingdom, ⁸The Early Cancer Inst., Dept. of Oncology, Univ. of Cambridge, Cambridge, United Kingdom, ⁹Dept. of Surgery and Cambridge NIHR BioMed. Res. Ctr., BioMed. Campus, Univ. of Cambridge, Cambridge, United Kingdom, ¹⁰Norwich Med. Sch., Univ. of East Anglia, Norwich, United Kingdom, ¹¹Nuffield Dept. of Surgical Sci., Med. Sci. Div., Univ. of Oxford, Oxford, United Kingdom

Abstract Body:

Somatic mutations in normal tissues can reveal the evolution of the genomic landscape during life and provide insights into the mutagenic processes that may contribute to cancer development. Although we now have a comprehensive perspective on the landscape of somatic mutations in cancers, due to technological limitations, we are still in the process of understanding the diversity of somatic mutation patterns in normal cells. Here, as part of a wider survey of human tissues, we have whole genome sequenced 342 laser-microdissected normal epithelial crypts from the small intestines of 39 individuals, and examined the somatic mutational burden, mutational signature exposure, and cancer driver mutation landscape. We find the frequent presence of single-base substitution mutational signatures SBS2 and SBS13, thought to be caused by the APOBEC family of cytidine deaminase enzymes.

APOBEC mutagenesis is a common source of mutations across the broad spectrum of human cancer types. However, it is unusual in previous studies of normal human tissues; in particular, the common presence of APOBEC mutational signatures in the small intestine differs from other segments of the gastrointestinal tract including oesophagus, stomach and colon, where it is rare. APOBEC3A/B are previously thought to be the major contributing enzymes of APOBEC mutagenesis, but bulk and single-cell RNA sequencing data of small and large intestine didn't indicate any elevated expression of APOBEC3A/B in the small intestine. We extract mutational signatures with extended context, and find APOBEC mutations are enriched in APOBEC1/3A motif. Together with the fact that the small intestine has a much higher APOBEC1 expression level than colon, this indicate that the common RNA editing enzyme APOBEC1 might have also contributed to the elevated APOBEC mutagenesis level in the small intestine.

Through sampling multiple crypts from the same individual together with phylogenetic reconstruction, we find that APOBEC mutagenesis is episodic in vivo, can happen throughout life (from as early as 4 years of age), but only a single or a very small number of episodes are found during the lifetime of an individual in any single crypt. Localised clusters of SBS2/13 mutations (kataegis) are also commonly found. Crypts with SBS2/13 often have immediate crypt neighbours without SBS2/13, suggesting that the underlying cause of SBS2/13 is controlled by cell-intrinsic factors rather than due to a widely distributed microenvironmental exposure. Overall, our findings substantially further define the features of this biologically important and distinctive mutational process.

Cancer Posters - Wednesday

PB1147. Leveraging network architecture and transfer learning for precision oncology

Authors:

S. Yi; The Univ. of Texas at Austin, Austin, TX

Abstract Body:

Leveraging multi-omics and network biology holds great promise for the next generation of precision oncology; yet, due to tumor heterogeneity, this is challenging. Towards this end, here we are presenting our two recent efforts, leveraging network architecture and transfer learning for improved precision oncology. In the first study, we built comprehensive cancer-specific miRNA regulatory networks across 30 cancer types to systematically analyze the effect of mutations on miRNA related pathways. 3,518,261 mutations from 9,819 samples were mapped to miRNA-gene interactions (mGI), and mutations in miRNAs versus in their target genes show a mutually exclusive pattern in almost all cancer types. We find that driver mutations play their roles by altering RNA binding energy and the expression of target genes. We provide this data resource (mGI-map) through a user-friendly, open-access web portal. Together, our results will facilitate novel non-coding biomarker identification and therapeutic drug design. The second effort deals with rational drug combination design. Numerous studies have been conducted to produce clinically relevant pharmacological response forecasts by integrating modern machine learning algorithms and several data types. Insufficient patient numbers and lack of knowledge of the molecular targets for each drug under study limit their use. As a proof of concept, we use single-cell RNA-seq based transfer learning to contextualize patients' tumor cells in terms of their more similar cell lines with known susceptibility to drug combinations. Our objective is to maximize the translational potential of in-vitro assays for identifying synergistic drug combinations and prioritizing them for clinical use. Consistent findings in a cohort of breast cancer patients corroborated our understanding of the disease's molecular subtypes. To aid in creating personalized treatments and data-driven clinical trials, we identified the most prevalent match options and prioritized synergistic combinations based on tumor compositions at various resolution levels.

Cancer Posters - Thursday

PB1148. Leveraging somatic cancer mutation data to predict the pathogenicity of germline missense variants

Authors:

B. Haque^{1,2}, M. Curtis^{1,2}, A. Pan², G. Costain^{1,2,3,4}; ¹Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, ²Genetics and Genome Biology, The Hosp. for Sick Children, Toronto, ON, Canada, ³The Ctr. for Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, ⁴Div. of Clinical and Metabolic Genetics, The Hosp. for Sick Children, and Dept. of Paediatrics, Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Background: Germline missense variants are often classified using current guidelines as having uncertain clinical significance. New scalable and easy-to-implement strategies are needed to generate supporting evidence for pathogenicity that can improve the interpretation of rare missense variants. We hypothesized that publicly available somatic cancer mutational data from tumor sequencing studies can be used to develop and test a new method of predicting the pathogenicity of germline missense variants. **Methods:** Somatic cancer missense mutations were extracted from the Cancer Hotspots database, and then annotated with *in silico* tools and public databases (e.g., ClinVar, gnomAD) using computational pipelines. Cancer mutations were identified that had been reported and classified in the ClinVar database as germline missense variants related to rare Mendelian disorders. By assessing the relationship between somatic cancer and germline variants in shared amino acid residues, a statistical model can be developed to estimate pathogenicity of novel missense variants. **Results (preliminary):** Among the 2,456 cancer mutations from Cancer Hotspots, 698 were reported as germline variants in ClinVar. The odds ratio for these germline variants having a pathogenic or likely pathogenic classification was 24.30 (95% CI, 20.78-28.41, $p < 0.001$), compared to all other germline missense variants in these same genes that are not in Cancer Hotspots. These and other data will be used as a training set for a supervised learning model for estimating the probability of pathogenicity of novel missense variants. **Conclusion:** Germline missense variants that are also known to be somatic cancer driver mutations are enriched for pathogenic/likely pathogenic classifications. These findings support the use of somatic cancer data as novel supporting evidence for germline variant classification to facilitate improved diagnosis of rare disorders.

Cancer Posters - Wednesday

PB1149. Liquid biopsy for the detection of actionable human variants in the blood of cancer-diagnosed dogs: Opening the door to comparative oncology and therapeutic development.

Authors:

I. Chorny, K. M. Kruglyak, J. A. Tynan, P. Nakashe, G. E. Hernandez, L. M. McLennan, J. M. Rafalko, J. Chibuk, A. L. McCleary-Wheeler, A. Flory, D. S. Grosu, D. W. Y. Tsui; PetDx, La Jolla, CA

Abstract Body:

Introduction: Characterizing actionable variants in humans and dogs can expand the knowledge base of comparative oncology and support the development of novel treatments for the benefit of both species. Blood-based liquid biopsy testing has opened new opportunities to noninvasively profile the cancer genomes of both species and aid in these translational efforts. The purpose of this study was to examine the ability of liquid biopsy testing to identify genomic variants common in human cancers in the blood of dogs with cancer diagnoses. **Methods:** In this study, the Cancer Gene Census (CGC) in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (v91) was interrogated for canine orthologs. A targeted next-generation sequencing somatic panel was then developed that incorporated a subset of the small genomic alterations from the CGC for which a canine ortholog was identified. The panel also incorporated a broader set of oncogenes and tumor suppressor genes (TSGs), including *TP53*, the most common TSG mutated in cancer. This panel was applied to plasma and tissue samples from a cohort of over 300 client-owned dogs with histologically confirmed diagnoses of a variety of cancer types. **Results:** High homology (>85%) was observed between the human and canine genomes, with even higher levels of homology (>90%) noted between human and canine oncogenes and TSGs. Approximately one third of cancer-diagnosed dogs were found to have a variant in both tissue and pre-surgical plasma that mapped to an actionable variant in the human genome or to the *TP53* gene (which is a human prognostic marker). Of those, ~10% had an ortholog of a variant targeted by current human therapies, including variants in *AKT1*, *HRAS*, *KRAS*, *NRAS*, and *PIK3CA*. **Conclusion:** This study demonstrates the technical feasibility of blood-based detection of canine orthologs of variants that are actionable in human cancers. The ability to noninvasively profile the genomic landscape of canine cancers opens the possibility of targeted treatment selection based on liquid biopsy. Many targeted therapies have received regulatory approval or are under development in human oncology; however, such therapies are lacking in veterinary medicine. Liquid biopsy is uniquely positioned to aid in large scale safety and efficacy studies to support therapeutic development to benefit both humans and dogs alike.

Cancer Posters - Thursday

PB1150. Loss of function of *neurofibromin* affects *afadin* -6 in neurofibromatosis type-1.

Authors:

M. Sulaiman¹, A. Dodson², T. Helliwell²; ¹Ahmadu Bello Univ., Zaria, ZARIA, Nigeria, ²Univ. of Liverpool, Liverpool, U.K, United Kingdom

Abstract Body:

Introduction Neurofibromatosis type 1 (NF1) is a common autosomal disorder in human with incidence of 1/3000 people worldwide. The *NF1* gene is mapped to chromosome 17, (17p21). Its protein product, *neurofibromin* negatively regulate *Ras*. Therefore, loss of function of *neurofibromin* may promote cell proliferation. The *afadin* (*AF-6*) protein is a multi-domain protein that contains two potential *Ras* binding domains and is mapped to 6q27 of human chromosome. *AF-6* has also been described as a constituent of a cell-cell adhesion system named; *Nectin*, *Afadin* and *Ponsin* (NAP), which is express at adherens junctions.

Aim of the Study This study was undertaken in order to demonstrate the likely consequences of instability of chromosome 17 on chromosome 6 and the functional integrity of cell-cell junction in familiar, sporadic and plexiform neurofibromas.

Material and Methods Prior informed patient consent from familial (9 cases), sporadic (7 cases) and plexiform neurofibromas (6 cases) was obtained, and the study has an ethical approval (06/1505/137) granted by the Liverpool Research Ethics Committee. Standard immunohistochemistry method was used. Briefly, routine histological procedures (haematoxylin and eosin staining) was first carried out followed by staining using cell markers and Rabbit anti-AF-6 polyclonal Zymed (Cat # 433280 staining on both experimental and control tissues.

Result Weak to moderate membranous and nuclei immunolocalization of the AF-6 were observed in endothelial cells of all the familial and the plexiform forms. Similarly, the *AF-6* moderately localize to the membrane of the endothelial cell of 57% of the sporadic neurofibroma studied.

Conclusion From this study, we suggest that the instability of chromosome 17 (*neurofibromin*) may affect chromosome 6 (*AF-6*) in familiar, plexiform and sporadic neurofibroma with the attendant loss of cell-cell adhesion and subsequent unregulated cell proliferation

Cancer Posters - Wednesday

PB1151. Low CD2 expression in sentinel lymph nodes of early breast cancer patients is associated with postoperative tumor recurrence or metastasis

Authors:

S. Kang, H. Lee, J-H. Kim, T. Park; Ajou Univ. Sch. of Med., Suwon, Korea, Republic of

Abstract Body:

Recurrence and metastasis are associated with increased mortality and poor outcomes in early breast cancer (EBC) patients. To predict loco-regional recurrence or distant metastasis, various prediction models based on clinicopathological characteristics or genetic profiles have been suggested. However, most of those systems focused on the primary tumor, and prediction of recurrence in EBC is still limited. The sentinel lymph nodes (SLNs) are the first few lymph nodes into which a tumor drains and an important barrier for protecting against metastasis. Since the immune microenvironment plays a key role in tumor progressions and metastasis, evaluation of SLNs could provide not only the state of metastasis to the axillary lymph nodes but also additional information on recurrence of EBC. In this study, we reviewed clinical data of EBC patients who had an operation with SLNs biopsy and selected patients with loco-regional recurrences or distant metastasis during follow-up period (at least 5 years). We analyzed clinicopathological characteristics of patients and investigated immune gene expression of SLNs by using the nCounter® PanCancer Immune Profiling Panel. We further performed immunohistochemistry for the most significantly altered genes. Among 1615 cases of EBC patients, 60 cases (3.7%) revealed loco-regional recurrences or distant metastasis. A total of 24 cases (12 SLNs of patients with recurrence and matched 12 SLNs of those without recurrence) were subjected to gene expression analysis and revealed a significant decrease in the expression of 21 genes. Based on gene expression levels and fold change, we selected 6 candidate genes (CD2, BLNK, ATF2, IGF2R, CXCL16, and IFITM1). We performed immunohistochemistry in a total of 119 cases (54 SLNs of patients with recurrence and matched 65 SLNs of those without recurrence) and found CD2 and ATF2 were significantly decreased in SLNs of EBC patients with recurrence. The pattern of immune gene expression of SLNs, especially C2 and ATF2 could be an additional prognostic marker for predicting recurrences of EBC.

Cancer Posters - Thursday

PB1152. Massively parallel sequencing-based panel strategy for microsatellite instability testing

Authors:

J. Styk^{1,2,3}, J. Budiš^{1,2,4}, J. Radvanský^{1,5,6}, M. Kucharík^{1,2}, O. Pös^{1,2}, W. Krampfl^{1,2,5}, N. Forgáčová^{1,5,6}, T. Sedláčková^{1,2}, P. Janega⁷, I. Lojová^{5,6}, T. Adamík¹, M. Nemeč¹, J. Turňa^{1,4,5}, V. Repiská^{3,8}, T. Szemes^{1,2,5}; ¹Comenius Univ. Sci. Park, Bratislava, Slovakia, ²Geneton Ltd., Bratislava, Slovakia, ³Inst. of Med. Biology, Genetics and Clinical Genetics, Faculty of Med., Comenius Univ., Bratislava, Slovakia, ⁴Slovak Ctr. of Scientific and Technical Information, Bratislava, Slovakia, ⁵Dept. of Molecular Biology, Faculty of Natural Sci., Comenius Univ., Bratislava, Slovakia, ⁶Inst. for Clinical and Translational Res., BioMed. Res. Ctr., Slovak Academy of Sci., Bratislava, Slovakia, ⁷Inst. of Pathological Anatomy, Faculty of Med., Comenius Univ., Bratislava, Slovakia, ⁸Medirex Group Academy n.p.o., Nitra, Slovakia

Abstract Body:

Somatic changes within short tandem repeats (STRs) result in microsatellite instability (MSI) that serves as a diagnostic biomarker for Lynch syndrome-associated tumors and a predictive biomarker for the efficacy of immune checkpoint inhibitors. Up to date, no alternative high-throughput screening strategy to gold-standard methods has been proposed. Along with advances in massively parallel sequencing (MPS) technologies and their ever-increasing cost-effectiveness, it could thus represent a new era of MSI monitoring. However, using them with the homopolymer repeats may be difficult due to technical and methodological drawbacks associated with FFPE samples and circulating cell-free DNA analyses. We designed an easy-handling MSI analysis workflow to enrich 17 homopolymer marker sites (8 - 27 bp) obtained from genomic regions associated with MSI across various malignancies. Unlike other MSI capturing strategies, we focus the sequencing on only the microsatellite regions of interest. The proposed assay was standardized for sequencing using the MiSeq platform. Sequencing data were annotated and genotyped with the bioinformatic tool Dante, not requiring mapping of reads to a reference genome. Our preliminary results clearly show that optimized PCR conditions allow to use FFPE blocks as a template, which are widely utilized in translational research settings. Although there was some variation within individual markers, this MPS-based method produced the same overall MSI status as obtained with the traditional method. We also show that biases resulting from the repetitive nature of homopolymers could be alleviated by pre-analytical optimization. We believe that this strategy could be used in practice as a straightforward and cost-effective assay. However, the analysis of cell-free DNA fragments and other laboratory and statistical methodologies are needed to improve the proposed MSI workflow's utility. The presented work was supported by the Slovak Research and Development projects (APVV-18-0319) and the OP Integrated Infrastructure within projects with ITMS codes 313011V578 and NFP313010Q927, all co-financed by the European Regional Development and Fund.

Cancer Posters - Wednesday

PB1153. Mechanistic insight into HIF1 α -Notch1 axis in chronic myeloid leukemia patients and its diagnostic implications

Authors:

V. Singh, R. Singh, R. Kushwaha; King George's Med. Univ., Lucknow, India

Abstract Body:

Chronic myeloid leukemia (CML) is potentially fatal blood cancer, but its diagnosis needs more attention in the search for molecular markers. However, this study explored the relationship of HIF1 α with the redox system, Krebs cycles, notch1, and other regulatory proteins and better understood the pathophysiology and clinical relevance in CML patients, where the molecular mechanism of this axis is still not clear. This study included CML patient samples (n-60; 60: blood; 10: bone marrow tissues) and compared them with healthy controls (n-20; blood) at the clinical and molecular levels. Here, we found that p210, p190BCR/ABL1 translocation is common in all blast crisis phases of CML. Redox factor and Krebs oncometabolite levels were upregulated, which consequentially led to upregulation and stabilization of HIF1 α that in turn lead to upregulation of their downstream genes (Notch2,4, Ikaros, SIRT1, Foxo-3a,p53, etc.). HIF1 α also downregulates ubiquitin proteasomal factors/apoptotic factors, which trigger its degradation by proline hydroxylation. the results also revealed that HIF1 α has an antagonist relationship with the notch1 pathway that plays the tumor-suppressive role in CML patients. Thus, this study has proven that Notch1 might be a potential diagnostic marker along with some other factors in CML patients. Overall this study concluded that HIF1 α /Notch1 axis plays an important role in the pathogenesis of leukemia, and might open the gate for the future diagnostic markers for CML.

Cancer Posters - Thursday

PB1154. Metabolic Profiling of Primary and Metastatic Breast Cancer Cells and New Treatment Approaches.

Authors:

L. Boccuto¹, D. Ivankovic^{1,2}, B. A. Moffitt¹, B. Dyar¹, N. Khulordava¹, A. Larkins¹, J. Benjock¹, J. Spencer¹, K. Henrie¹, S. Loehr¹, M. A. Learmonth², G. M. DeRosa², A. G. Powell², Z. E. Sanders², M. A. Treaster²; ¹Clemson Univ., Clemson, SC, ²Anderson Univ., Anderson, SC

Abstract Body:

Breast cancer is the second most common type of cancer overall and the first among women, 20-30% of cases develop metastases. More than 154,000 U.S. women are estimated to have metastatic breast cancer, and about 40,000 die every year because of this disease. In preliminary studies, we noted significant abnormalities in the metabolic profile of breast cancer cells, and we demonstrated that a natural compound such as marjoram can correct such abnormalities in cancer cells without disrupting normal cells. This study aims to compare the metabolic profiles of metastatic and non-metastatic cancer cells to one of normal breast tissue, and then assess the efficacy of candidate compounds at correcting metabolic abnormalities in cancerous cells. The purpose of this study is to identify metabolic abnormalities underlying the metastatic transformation of cancer cells, novel biomarkers for early screening, new molecular targets for treatment, and to investigate the efficacy and side effects of candidate therapeutic compounds. Results gathered from the metabolic profiles of metastatic and non-metastatic breast cancer, as compared to the one from normal breast tissue, were investigated and assessed via the Biolog Phenotype Mammalian Microarray (PM-M) platform. This approach has allowed us to analyze the capacity of the cells to produce energy in different metabolic environments and after exposure to over 700 compounds, including energy sources, hormones, cytokines, and other metabolic effectors. The assay has revealed significant differences between the metabolic profiles of metastatic and non-metastatic cancer, both compared to normal breast tissue cells. We have selected the PM-M arrays with the highest number of compounds showing significant differences between metastatic and non-metastatic cancer. We have tested the efficacy of one drug commonly used for chemotherapy in metastatic breast cancer, doxorubicin (Adriamycin), and a natural compound that proved to be effective on cancer cells, marjoram. The two compounds were tested on both primary and metastatic breast cancer cells, as well as on control cells from the normal breast tissue, and their efficacy at correcting metabolic abnormalities was assessed, as well as their toxicity and metabolic side effects. By identifying biomarkers for early diagnosis or screening, new novel treatment approaches can be discovered, and the efficacy of candidate drugs can be assessed. The results of our study may lead to innovative approaches to the characterization, management, and treatment of metastatic breast cancer.

Cancer Posters - Wednesday

PB1155. Metagenomics and metabolomics of persistent high-risk Human Papillomavirus (hrHPV) infections and cervical cancer

Authors:

C. Adebamowo^{1,2,3}, S. Adebamowo^{1,3}, A. Adebayo⁴, P. Achara⁵, B. Alabi⁶, R. Bakare⁷, A. Famooto², K. Obende⁸, R. Offiong⁹, O. Olaniyan¹⁰, ACCME Research Group as part of the H3Africa Consortium; ¹Univ. of Maryland Sch. of Med., Baltimore, MD, ²Inst. of Human Virology Nigeria, Abuja, Nigeria, ³Greenebaum Comprehensive Cancer Ctr., Baltimore, MD, ⁴Asokoro District Hosp., Abuja, Nigeria, ⁵Federal Med. Ctr., Keffi, Nigeria, ⁶Wuse Gen. Hosp., Abuja, Nigeria, ⁷Univ. of Ibadan, Ibadan, Nigeria, ⁸Garki Hosp., Abuja, Nigeria, ⁹Univ. of Abuja, Abuja, Nigeria, ¹⁰Natl. Hosp., Abuja, Nigeria

Abstract Body:

Background: Persistent hrHPV infections is a non-oncolytic infection of the basal epithelium at the transition zone of the cervical epithelium and a necessary cause of cervical cancer. Evidence suggests significant roles for the vaginal microenvironment in establishment of persistence of the infection and progression to cervical cancer. **Methods:** We conducted shotgun metagenomics sequencing and global metabolomics profiling of the cervical secretions from 84 women. Some 68 of these women were tested and categorized based on HPV results as negative (N) or positive (P) at 3 time points: baseline, 12 months, and 24 months, using DEIA/LiPA₂₅ into NNN (13), NNP (16), NPP (10), PPP (12), PPN (15) and PNN (1), while 16 women had either invasive squamous cell carcinoma (15) or papillary adenocarcinoma (1). **Results:** There was higher diversity of bacteria and lower *Lactobacillus* abundance comparing the PPP category to the NNN category. The most frequently identified genus in PPP category were *Gardnerella*, *Prevotella*, *Atopobium*, *Ureaplasma*, *Sneathia* and *Streptococcus*. Cervical cancer samples showed significantly decreased bacterial diversity, absence of *Lactobacillus spp.*, and relatively high abundance of *Cutibacterium*. Our discriminative analysis showed that *Prevotella* prevalence and abundance discriminated between NNN and PPP categories. We found *Mageeibacillus indolicus*, a recently discovered bacteria in 4 members of the PPP group. The commonest HPV types in the cervical cancer group were types 16 and 18, while HPV 82 was the most prevalent hrHPV genotype in the other categories. The top five enriched pathways in the NNN group compared to the PPP group were: *Folate transformations II, sucrose degradation III (sucrose invertase), lipid IVA biosynthesis, super pathway of thiamin diphosphate biosynthesis III (eukaryotes)* and *Guanosine ribonucleotides de novo biosynthesis* mainly contributed by the *Lactobacillus spp.* and unclassified strata. The top enriched gene family in the NNN group: *3.1.22.4 Crossover junction endodeoxyribonuclease, 2.7.1.69 Protein-N(pi)-phosphohistidine—sugar phosphotransferase* and *1.7.1.7 GMP reductase* similarly contributed by diverse *Lactobacillus spp.* and unclassified. We identified HPV integration sites in 6/20 cervical cancer cases and 1/12 of the PPP cases. **Conclusions:** Persistent hrHPV infection was associated with vaginal dysbiosis and reduced abundance of *Lactobacillus spp.* Several bacterial genera and species, gene families and pathways whose abundance was significantly different between the different groups in this study and hold promise as potential biomarkers.

Cancer Posters - Thursday

PB1156*. Methylation-sensitive restriction enzyme sequencing (MRE-seq)-based early detection of cancer using deep neural network

Authors:

H. Kwon¹, S. Shin², N. Min³, H. Kim⁴, S-Y. Yun¹, D. Park¹, T-W. Joo¹, Y. Lim¹, Y-H. Kim¹, K. Lee¹, H. KIM³, M-S. Jeong¹, K. Lee³, K. Kim⁵, S-W. Um², C. Ahn⁴, S. Lee³; ¹Eone-Diagnomics Genome Ctr., Inc., Incheon, Korea, Republic of, ²Samsung Med. Ctr., Seoul, Korea, Republic of, ³Eone Diagnomics Genome Ctr. Inc., Incheon, Korea, Republic of, ⁴Bucheon St. Mary's Hosp., Bucheon, Korea, Republic of, ⁵Gangnam Major Hosp., Seoul, Korea, Republic of

Abstract Body:

Background: Among the various genomic features of blood, methylation patterns could be useful not only for determining whether cancer is present but also for revealing its origin. The purposes of this prospective study are to evaluate the diagnostic performance of methylation-sensitive restriction enzyme digestion followed by sequencing (MRE-seq) using cell-free DNA (cfDNA) in the diagnosis of cancer and to investigate tissue-of-origin (TOO) using a deep neural network (DNN).

Patients and methods: We developed a deep-learning-based prediction model using demethylated-sequence-depth patterns from 63,266 sites through MRE-seq. One hundred and ninety-one patients with stage I-IV cancers (95 lung cancers and 96 colorectal cancers) and 126 noncancer participants were enrolled in this study.

Results: We found an overall area under the ROC curve (AUC) of 0.978 with a sensitivity of 78.1% at a specificity of 99.2% for colorectal cancer and an AUC of 0.956 with a sensitivity of 66.3% at a specificity of 99.2% for lung cancer. In colorectal cancer, the sensitivity was 76.5% in stage I, 76.2% in stage II, 78.3% in stage III, and 83.3% in stage IV at a specificity of 99.2%. In lung cancer, the sensitivity was 48.5% in stage I, 44.4% in stage II, 78.9% in stage III, and 76.0% in stage IV at a specificity of 99.2%. The true positive rates of the TOO model were 94.4% (85/90) and 89.9% (80/89) for colorectal and lung cancers, respectively.

Conclusions: The global hypomethylation pattern obtained by MRE-seq was useful in the diagnosis of cancer with high accuracy and in the determination of TOO through DNN. The role of MRE-seq needs to be validated for various solid tumors in future multicenter studies.

Cancer Posters - Wednesday

PB1157. Mitonuclear genotype remodels the metabolic and microenvironmental landscape of Hürthle cell carcinoma

Authors:

E. Liu¹, I. Ganly¹, F. Kuo¹, V. Makarov², Y. Dong¹, J. Park¹, Y. Gong¹, A. N. Gorelick³, J. A. Kanuf², E. Benedetti⁴, J. Tait-Mulder⁵, L. G. T. Morris¹, J. A. Fagin¹, A. M. Intlekofer¹, J. Krumsiek⁴, P. A. Gammage⁵, R. Ghossein¹, B. Xu¹, T. A. Chan², E. Reznik¹; ¹Mem. Sloan Kettering Cancer Ctr., New York, NY, ²Cleveland Clinic, Cleveland, OH, ³Massachusetts Gen. Hosp., Boston, MA, ⁴Weill Cornell Med., New York, NY, ⁵Cancer Res. UK Beatson Inst., Glasgow, United Kingdom

Abstract Body:

Hürthle cell carcinomas (HCCs) display two exceptional genotypes: near-homoplasmic mutation of mitochondrial DNA (mtDNA) and genome-wide loss of heterozygosity (gLOH). To understand the phenotypic consequences of these genetic alterations, we analyzed genomic, metabolomic, and immunophenotypic data of HCC and other thyroid cancers. Both mtDNA mutations and profound depletion of citrate pools are common in both HCC and other thyroid malignancies, suggesting that thyroid cancers are broadly equipped to survive tricarboxylic acid cycle impairment, whereas metabolites in the NADH-dependent lysine degradation pathway were elevated exclusively in HCC. The presence of gLOH was not associated with metabolic phenotypes but rather with reduced immune infiltration, indicating that gLOH confers a selective advantage partially through immunosuppression. Unsupervised multimodal clustering revealed four clusters of HCC with distinct clinical, metabolomic, and microenvironmental phenotypes but overlapping genotypes. These findings chart the metabolic and microenvironmental landscape of HCC and shed light on the interaction between genotype, metabolism, and the microenvironment in cancer.

Cancer Posters - Thursday

PB1158. Modulating the immune system not to reject an allograft

Authors:

D. McDaniel, L. S. McDaniel; Univ Mississippi Med Ctr, Jackson, MS

Abstract Body:

Introduction: Organ transplantation is a major therapeutic approach in end stage disease with organ dysfunction. Mechanical devices to replace or for restoration of organ structure/function remain a challenge because many biological functions cannot be replicated. The immune system plays a central role in the recovery process including restoration of organ function after allograft transplantation. Organ or tissue chip development following tissue engineering through reprogramming of immune system criteria has been proposed as the next generation of regenerative medicine and transplantation. **Methods:** Current approaches in regenerative medicine includes cell-based therapeutics; combination of cells and scaffolds to stimulate tissue repair *in vivo* and building of tissues *ex vivo* for implantation. Training stimulatory pathways of immunity that induce macrophage polarization, antigen presentation, T cell activation and cytokine production are included. **Results:** Previously, we have shown that human recipients can tolerate their transplanted organ when infusing the donor bone-marrow cells, prior to the kidney transplantation. This is known the induction of tolerance through mixed chimerism. However, this approach required a specialized conditioning regimen to allow allograft survival with less or minimized immunosuppressants. Furthermore, during organ procurement (recovery) stimulatory signals released in the donor organ due to ischemia-reperfusion injury, encounter trained macrophages to secrete pro-inflammatory cytokines leading to allograft rejection. We reported the AIF-1-TLR2 interactions favor the induction of tolerance under blockade with anti-AIF-1 through the secretion of IL-10 by graft-infiltrating macrophages. **Conclusions:** Advances in immune modulation through reprogramming of immune system arranges the foundation to train the immunity in organ transplantation not to reject through epigenetic modification of the immune microenvironment.

Cancer Posters - Thursday

PB1160. Molecular signature predicts immunotherapy resistance in gynecologic cancer patients treated with immune checkpoint inhibition.

Authors:

Y. Ni¹, H. Al-Sudani², A. Soliman³, P. G. Rose¹, H. Mahdi⁴; ¹Cleveland Clinic, Cleveland, OH, ²Einstein Med. Ctr. Montgomery, Philadelphia, PA, ³Case Western Reserve Univ., Cleveland, OH, ⁴Univ. of Pittsburgh Med. Ctr., Pittsburgh, PA

Abstract Body:

Immunotherapy with immune checkpoint inhibitors (ICI) has emerged as a promising option in other solid tumors like lung and urothelial cancers and melanoma. In gynecological (GYN) cancers, the ICI response rates range from 11-17% in the recurrent setting. Various immune signatures predictive of resistance to immune ICI have been described in multiple solid cancers, but still under-investigated in GYN cancer.

For 49 GYN cancer patients included in our study, without transcriptome signature, immune-related toxicity was the only clinical predictor of ICI treatment response ($p=0.008$). The objective clinical response was the only predictor of progression-free survival (ICI-PFS, $p=0.0008$) and overall survival (ICI-OS, $p=0.01$). Commonly used ICI marker PD-L1 expression was negatively correlated with progression-free survival (ICI-PFS) ($p=0.0019$). We performed transcriptome and signaling pathway enrichment analyses based on ICI treatment responses and the survival outcome, and further estimated immune cell abundance using 547 gene markers. Our data revealed that TGF-beta regulated signaling pathway was noted to play an important role in immunotherapy failure. Using our 6-genes TGF-beta score, we observed longer ICI-PFS associated with lower TGF-beta score (8.1 vs. 2.8 months, $p=0.046$), which was especially more prominent in ovarian cancer (ICI-PFS 16.6 vs. 2.65 months, $p=0.0012$). Further, abundant immunosuppressive cells like T-regulatory cells, eosinophils, and M2 macrophages were associated with shorter ICI-OS and correlated positively with *CD274* (PD-L1) and *CTLA4* expressions. Interestingly, patients who failed immunotherapy also have higher *CD47* expression ($p=0.03$). Higher *CD47* correlated with lower cytotoxic T-cells ($p < 0.001$) and dendritic cells ($p=0.038$) as well as lower *CD274* (PD-L1) and *CTLA4* expression but positively correlated with higher TGF- signature ($p=0.006$). With an integrated random forest prediction model, besides immunotherapy treatment cycle and toxicity, TGF-beta score and *CD47* together with *CD274* (PD-L1) expression are the top important contributors predicting immunotherapy response. Our study provides insight on the potential role of TGF-beta and *CD47* in mediating immunotherapy resistance and cross-talking to immunosuppressive environment in GYN cancer. If validated in a larger cohort, can identify patients who likely to fail ICI and benefit from targeting this pathway to enhance the response to ICI.

Cancer Posters - Wednesday

PB1161. Mosaic chromosomal alterations detected in HIV infected men and its relationship to non-Hodgkin lymphoma.

Authors:

S. Khan¹, S-H. Lin¹, W. Zhou¹, D. W. Brown¹, C. Vergara², S. Wolinsky³, O. Martinez-Maza⁴, J. Margolick², J. Martinson⁵, S. K. Hussain⁶, M. J. Machiela¹, E. A. Engels¹; ¹Natl. Cancer Inst., Rockville, MD, ²Bloomberg Sch. of Publ. Hlth., The Johns Hopkins Univ., Baltimore, MD, ³Feinberg Sch. of Med., Northwestern Univ., Chicago, IL, ⁴UCLA AIDS Inst. and Jonsson Comprehensive Cancer Ctr. at UCLA, Los Angeles, CA, ⁵Univ Pittsburgh GSPH, Pittsburgh, PA, ⁶Univ. of California Davis Comprehensive Cancer Ctr., Sacramento, CA

Abstract Body:

Men living with HIV (MLWH) are at increased risk of infection and non-Hodgkin lymphoma (NHL); however, limited clinical markers are available for identifying men most susceptible to these outcomes. Mosaic chromosomal alterations (mCAs), the age-related acquisition of large acquired chromosomal alterations in a clonal subset of leukocytes, are markers of genomic instability and associated with increased susceptibility to infections and hematologic malignancies. As mCAs have not previously been characterized in MLWH, we investigated the frequency and distribution of mCAs among MLWH as well as potential elevated risk of NHL conferred by mCAs. We scanned for mCAs in blood-derived DNA from 3,139 MLWH and 2,840 HIV- men from the Multicenter AIDS Cohort Study (MACS) using raw genotyping array intensity data. The proportion of MLWH and HIV- with autosomal mCAs was similar (121 [3.9%] vs. 102 [3.6%]); however, mosaic loss of the Y chromosome (mLOY) was less common in MLWH than in HIV- men (44 [1.4%] vs. 83 [2.9%]). Logistic regression models adjusting for age and smoking found no evidence for an association of HIV with autosomal mCAs (OR=1.19, 95% CI=0.98-1.58, P-value=0.22), but an inverse association with mLOY (OR=0.60, 95% CI=0.40-0.88, P-value=0.01). No differences in median length of mCAs (31.2 vs 35.7 Mb) or mean cell fraction impacted (7.48% vs 7.65%) were observed between HIV- and MLWH. The genomic distribution of mCAs differed at select chromosomal regions, with MLWH more likely to harbor 14q copy number neutral-loss of heterozygosity (CN-LOH) events (P-value=3.5×10⁻³) and less likely to have chromosome 1 gains (P-value=0.03) and 11 losses (P-value=0.03). A total of 177 (5.6%) MLWH developed incident NHL during follow-up, with 5 cases occurring in MLWH with autosomal mCAs and 3 occurring in MLWH with mLOY. MLWH with mCAs who subsequently developed NHL had higher median cell fractions relative to NHL-free men (7.6% vs. 2.3% in MLWH). Unadjusted Poisson regression models showed no evidence for an increased incidence of NHL among MLWH associated with autosomal mCA (IRR=0.81, 95% CI=0.29-1.78) or mLOY (IRR=1.51, 95% CI=0.37-3.97). There were no common genomic regions impacted by autosomal mCAs in the NHL cases, and positions were similar to NHL-free men in the MACS cohort. Neither autosomal mCAs nor mLOY were associated with mortality in MLWH when examined with models adjusting for age and smoking. We find no overall association between HIV and autosomal mCAs nor between mCAs and NHL risk among MLWH. Our findings suggest limited clinical utility of mCAs in MLWH, but some relationships (e.g., enrichment of 14q CNLOH in MLWH) merit further investigation.

Cancer Posters - Thursday

PB1162. Multigene deregulation 12q12 renal cancer-susceptibility locus.

Authors:

L. Andrade Costa¹, A. Souza¹, E. Zier¹, L. Jessop², T. Myers², S. Chanock², D. Mole³, L. Colli¹; ¹Univ. of São Paulo, Ribeirão Preto, Brazil, ²Natl. Cancer Inst., Bethesda, MD, ³Univ. of Oxford, Oxford, United Kingdom

Abstract Body:

Renal Cancer (RCC) is the seventh most common type of cancer in the United States, and most of these tumors do not respond to radiation and chemotherapy treatments. A Genome-Wide Association Study (GWAS) identified 13 regions related to the risk of developing RCC. In this study, we analyzed the *12q12* region, a region that has already been shown to have two SNPs increasing the expression of *BHLHE41*. However, eQTL data suggests other genes in the region could also be deregulated by those SNPs. It is well established that a region may have multiple functional SNPs, but the interaction between genes is rare. Our objective is to evaluate *12q12* gene interaction and how this could impact RCC susceptibility. For this, we are using CRISPR (CRISPRi) to inhibit the expression of two SNPs associated with the *12q12* region, rs7132434 and rs10842708, in renal cancer cell lines (UO-31 and 786-O) and analyze changes in *BHLHE41*, *SSPN*, *ITPR2*, and *RASSF8* expression. On the 786-O cell line, which is homozygous for protective alleles, the blockade of SNPs does not change the expression of any *12p12* evaluated genes. On the other hand, on UO-31 lineage (risk alleles), the inhibition of both SNPs was able to decrease the expression of *BHLHE41*, and, in a lower effect, it also decreased *SSPN* and *ITPR2* expression. Previous work has shown that *BHLHE41* and *SSPN* have an oncogenic effect on renal cancer development. Here, our project showed evidence for multiple gene deregulation impacting cancer susceptibility in renal cancer.

Cancer Posters - Wednesday

PB1163. Multi-modal characterization of ultra-rare germline genetic variants driving breast cancer risk in the indigenous Arab population

Authors:

H. Chu^{1,2}, A. Al-Sulaiman³, M. Al-Jumaan³, S. Han^{4,2}, S. Y. Camp^{1,2}, R. Gillani^{5,2,6,7}, Y. Al Marzooq³, F. Almulhim³, C. B. Vatte³, A. Alnimer³, A. Almuhanna³, E. M. Van Allen^{1,2,8}, A. Al-Ali⁹, S. H. Aldubayan^{1,2,10,11}; ¹Dept. of Med. Oncology, Dana-Farber Cancer Inst., Boston, MA, ²Cancer Program, The Broad Inst. of MIT and Harvard, Cambridge, MA, ³Coll. of Med., Imam Abdulrahman bin Faisal Univ., Dammam, Saudi Arabia, ⁴Harvard Med. Sch., Cambridge, MA, ⁵Dept. of Pediatric Oncology, Dana-Farber Cancer Inst., Boston, MA, ⁶Dept. of Pediatrics, Harvard Med. Sch., Boston, MA, ⁷Boston Children's Hosp., Boston, MA, ⁸Ctr. for Cancer Genomics, Dana-Farber Cancer Inst., Boston, MA, ⁹Coll. of Med., Imam Abdulrahman bin Faisal Univ., Dammam, MA, ¹⁰Div. of Genetics, Brigham and Women's Hosp., Boston, MA, ¹¹Coll. of Med., King Saud bin Abdulaziz Univ. for Hlth.Sci., Riyadh, Saudi Arabia

Abstract Body:

The indigenous Arab population is underrepresented in genomic studies and the germline predisposition to breast cancer in this population is largely unknown. Using a deep-learning-based variant caller, we performed systematic exome-wide characterization of germline single-nucleotide polymorphisms (SNPs) and indels from 215 unrelated Arab female breast cancer patients unselected for family history and 1290 ancestry-matched female controls. Pathogenic variants in 143 established cancer-predisposition genes were found in 15.8% (95% Confidence Interval [CI]: 11.2%-21.4%; binomial exact) of the patients with 13.02% (95%CI: 8.83%-18.27%) of the patients carrying at least one clinically actionable variant, including a founder frameshift variant in BRCA2 (c.2808_2811del, p.Ala938ProfsX21, rs80359351) previously unreported in this population which was present in 5.12% (95%CI: 2.58%-8.97%) of the patients. A gene-burden analysis of ultra-rare pathogenic and loss of function variants was then performed which showed BRCA2 to be the most significantly enriched gene after adjusting for false-discovery rate (alpha=0.05) (Odds Ratio [OR]: 82.6, 95%CI: 13.44-1739.6, q-value<0.001, p-value<0.001, Fisher's exact), followed by three nominally enriched genes BRCA1 (OR:10.19, 95%CI: 2.47-49.23, q=0.065, p=0.002), SDHB (OR >3.52, q=0.065, p=0.002), and PTEN (OR >1.73, q=0.346, p=0.02). A validated germline copy number variant (CNV) caller was used to analyze 824 cancer-related genes and identified pathogenic CNVs in 4.19% (95%CI: 1.93%-7.80%) of the patients with 2.79% (95%CI: 1.03%-5.97%) of the patients carrying germline MYC partial duplications, which was significantly associated with cancer spreading to lymph node (OR:11.30, 95%CI: 1.42-281.05, p=0.02) and metastasis (OR:11.51, 95%CI: 1.29-83.46, p=0.03). Overall, carriers of pathogenic variants in the expanded cancer genes (mean age: 43.7, standard deviation [SD]: 8.7) presented on average 5.2 years earlier than non-carriers (mean age: 48.8, SD: 10.1, p=0.004, Mann-Whitney U-test), and are almost four times (OR: 3.69, 95%CI: 1.04-13.93, p=0.045, Fisher's Exact) more likely to become metastatic than non-carriers. Collectively, our study identified a highly actionable BRCA2 founder variant in the Arab population with available first-line targeted therapeutic intervention, highlighted the major drivers of breast cancer risk and heritability in this population, and alleviated existing healthcare disparity by providing evidence for clinical actions in cancer patients from an underrepresented population.

Cancer Posters - Thursday

PB1164. Multiomic sequencing analysis with whole genome association study at chromosome 16 of wilms tumor patients.

Authors:

C-S. Chang, K. Eiko; Augusta Univ., Augusta, GA

Abstract Body:

Wilms tumor (WT) is the most common childhood renal tumor and affects approximately 1 in 10,000 children in Europe and North America. Chromosome 16q deletion (del) or loss of heterozygosity (LOH) has been associated with recurrence and adverse prognosis in WT. Through a Genome Wide Association Study (GWAS) on a biallelic test and a filtration of Minor Allele Frequency (MAF) of 0.01, we further discovered that a block of SNPs with the lowest p value of $3.45E-08$ on the SNP of rs2632824 at 16P11.2 found in this study. Therefore, we focused on searching for the genes implicated in the adverse event at this region of Chromosome 16P11.2. We performed the isoform detection of 27 WT's RNASeqs from the SRA database, together with conducting the RNA-seq analysis. Differentially expressed genes were analyzed by Tuxedo suite pipeline from each group. Through an isoform detection analysis by calculating the percent spliced in index (PSI) counts, we found the alternative splicing events mainly with an intron retention detected on genes of *C16orf58* (RUSF1) and *SLC5A2* at Chromosome 16P11.2 from the tumor group. These two genes, especially *SLC5A2*, were statistically up-regulated in the tumor group when compared with the healthy controls from SRA as well. An ITGA gene family are located in the upstream region, a series of TP53 target gene (*TP53TG*) family, *NR 158162* and *LOC390705* as well as *ENPP7P13* are located at the downstream region of Chromosome 16P11.2 that potentially indicates the importance of the risk loci previously identified by Exome-Seq data for the disease phenotype of Wilms tumor by a mechanism of the enhancer regulation from coding and no-coding RNAs at this region. These findings derived from the multiomic sequencing studies can provide a likelihood of being applied as an evidence of genetic biomarkers initiated at the quantitative trait loci for susceptibility as well as an insight of the actionable genes for a therapeutic drug target at chromosome 16P11.2 in pediatric Wilms tumor patients.

Cancer Posters - Wednesday

PB1165. Multi-omic single-cell profiling of endogenous and engineered T cells in patients undergoing CAR T cell therapy for high-grade glioma

Authors:

H. Natri¹, M-I. Chung¹, L. Peter¹, D. Alizadeh², B. Badie², C. Brown², **N. Banovich¹**; ¹Translational Genomics Res. Inst. (TGen), Phoenix, AZ, ²City of Hope Beckman Res. Inst. and Med. Ctr., Duarte, CA

Abstract Body:

Chimeric Antigen Receptor (CAR) T cell therapy has emerged as a promising approach to treating cancer, including glioblastoma (GBM) and other gliomas. T cell exhaustion due to persistent antigen exposure in the tumor microenvironment remains a major limitation to CAR T efficacy in the treatment of solid tumors. Enrichment of less differentiated memory T cells for manufacturing may enhance CAR T product proliferation, persistence, and anti-tumor efficacy. A better understanding of the cellular and transcriptional programs driving T cell fitness can inform CAR T manufacturing to improve clinical outcomes. To this end, we present a comprehensive multidimensional characterization of CAR T cells and endogenous immune cells from 62 patients with recurrent or refractory malignant glioma who participated in a Phase I study on memory enriched CAR T cells targeting a glioma-associated antigen IL13R α 2. Each patient received infusions of CAR T cells produced from T cell subsets enriched either for naive, stem-like memory, and central memory T cells (Tn/mem, n=22) or only central memory T cells (Tcm, n=40). T cells from patient leukapheresis used as starting material in CAR T manufacturing and the resulting infusion products were characterized using single-cell multi-omic profiling of gene expression, surface protein expression, and T cell receptor clonality. Annotation of clusters of 80,986 CAR T cells and 32,319 leukapheresis T cells shows that the dataset consists of a mixture of memory, activated effector, and exhausted cytotoxic (CD8+) and helper (CD4+) T cells. Single-cell data recapitulate differences in surface phenotypes observed between Tn/mem and Tcm products in flow cytometric analyses, with enrichment of cytotoxic T cells in Tn/mem products reflecting differences in manufacturing protocols. Higher expression of exhaustion markers in cells used in manufacturing was associated with poorer expansion *ex vivo*, demonstrating the quality of the patient-derived starting material as one of the drivers of CAR T fitness. Tn/mem products showed higher levels of proliferation *ex vivo* and expressed higher levels of cytotoxic markers, likely underlying higher anti-tumor efficacy. While the fitness of the patient-derived T cells remains a limiting factor in CAR T therapy, our results demonstrate that the manufacturing protocol largely drives CAR T fitness and anti-tumor efficacy. Further developing culture protocol can help reduce cell stress and improve long-term outcomes. These findings will inform strategies for improving clinical outcomes of CAR T cell therapy in GBM and other cancers.

Cancer Posters - Thursday

PB1166. Multiplex ligation dependent probe amplification versus fluorescent in situ hybridization for screening *RB1* copy number variations in Egyptian patients with retinoblastoma.

Authors:

o. Eid¹, H. El Zomor², A. Mohamed³, H. El-Bassyouni⁴, H. Afifi³, M. El-Ayadi⁵, S. Sadek⁶, s. hammad⁷, S. Sherine², R. Mahrous¹, I. Fadel¹, K. Refaat¹, M. Afifi², A. Shelil², O. Ziko⁸, A. Abdel Azeem⁹, A. El-Haddad²; ¹Natl. research centre - Egypt, Giza, Egypt, ²Children's Cancer Hosp., 57357, Cairo, Egypt, ³Natl. Res. Ctr., Cairo, Egypt, ⁴Natl. Res. Ctr., Cairo, Cairo, Egypt, ⁵Children's Cancer Hosp., 57357, Egypt, Cairo, Egypt, ⁶Children's Cancer Hosp., 57357, Egypt, Cairo, Egypt, ⁷Natl. research center, cairo, Egypt, ⁸Ain Shams Univ., Cairo, Egypt, ⁹Res. Inst. of Ophthalmology, Cairo, Egypt

Abstract Body:

Background: Retinoblastoma (RB) is the most common primary intraocular malignant tumor in children. RB is mostly caused by biallelic mutations in *RB1* and occurs in hereditary and non-hereditary forms according to the “two-hit” theory. *RB1* mutations comprise point mutations, indels, large deletions, and duplications. Genetic testing is essential for the comprehensive treatment and management of patients with RB. **Aim:** the aim was to evaluate *RB1* copy number variations (CNVs) using MLPA versus FISH assays in group of Egyptian patients with RB. **Results:** Twelve patients from the 72 patients (16.67%) showed an *RB1* deletion, abnormal methylation status, or both. The detection rate of *RB1* deletions in our cohort was 5.5% by FISH, which agreed with reported detection rates. However, the detection rate increased to 15.27% using MLPA; this assay detected partial gene deletions not detected by FISH. Two patients who had an *RB1* deletion exhibited mosaicism by FISH, one showed 30%-40% *RB1* deletion mosaicism while the other showed only 15% *RB1* deletion mosaicism. MLPA detected deletion in the 1st patient. Also, MLPA assay detected hyper-methylation *RB1* at intron 2 in two patients; one was associated with almost entire *RB1* deletion. The assay also detected hypo-methylated *RB1* at intron 2 in a patient and was associated with a partial *RB1* deletion. **Conclusion:** Our results suggested MLPA is a fast, reliable, and powerful method and should be used as a first-line screening tool for detecting *RB1* CNVs in patients with RB. Moreover, MLPA is advantageous as it evaluates the methylation status/inactivation of *RB1*, not possible by FISH.

Cancer Posters - Wednesday

PB1167. Multi-population GWAS analysis identified functional variants in ever- and never-smoking lung cancer.

Authors:

Y. Li, X. Xiao, J. Xia, J-R. Li, C. Cheng, Y. Han, G. Fernandes, S. Slewitzke, S. Rosenberg, J. Byun, C. Amos, INTEGRAL-ILCCO lung cancer consortium; Baylor Coll. of Med., Houston, TX

Abstract Body:

Introduction Remarkable differences have been identified in both clinical and molecular epidemiology studies between ever- and never-smoking lung cancer. Most of previous GWAS studies were conducted in European and Asian population, and African-American population is under-represented. A multi-population lung cancer Genome-wide Association Study (GWAS) stratified by smoking status has the potential to identify novel variants. **Method** We conducted a multi-population (European, Asian and African-American) GWAS on ~9,000,000 high-quality imputed SNPs in 44,823 ever-smokers and 20,074 never-smokers, respectively. Extensive functional analysis was conducted using approaches including functional annotation analysis, colocalization analysis, and DNA damage assay. **Results** Five independent novel loci, including *GABRA4* from ever-smoking and *LRRC4C* and *LCNLI* from never-smoking lung cancer, were identified with association effect in two or three populations ($P < 5 \times 10^{-8}$). Besides the novel findings, we also validated the lung cancer risk effect for known variants at *VTTIA* and *ACVR1B* in never-smoking women in African-American for the first time. Multiple lines of evidence were obtained suggesting the variants may potentially increase cancer risk through excessive DNA damage level (*GABRA4*) or cis-regulation of gene expression (*LCNLI*). **Conclusion** Our study highlighted the heterogeneity in genetic architecture between ever- and never- smoking lung cancer. It also demonstrated that inclusion of African population in the multi-population GWAS is crucial for better understanding of genomic and environmental variations underpinning lung cancer. This integrated study enhanced our knowledge about the complicated genetic architecture in lung cancer and has a potential to contribute to precision medicine in lung cancer treatment and prevention.

Cancer Posters - Thursday

PB1168*. Multi-stage germline exome sequencing study of 17,546 men with aggressive or non-aggressive prostate cancer identifies genes that may inform clinical gene panel testing.

Authors:

B. Darst¹, E. Saunders², T. Dadaev³, X. Sheng⁴, S. Chanock⁵, S. Berndt⁶, Y. Wang⁷, A. Patel⁷, D. Albanes⁸, S. Weinstein⁸, V. Gnanapragasam⁹, G. Giles¹⁰, T. Nguyen-Dumont¹¹, R. Milne¹⁰, M. Pomerantz¹², J. Schmidt¹³, R. Travis¹³, T. Key¹³, K. Stopsack¹⁴, L. Mucci¹⁵, W. Catalona¹⁶, K. Hetrick¹⁷, K. Doheny¹⁸, R. MacInnis¹⁰, M. Southey¹⁹, R. Eeles²⁰, F. Wiklund²¹, Z. Kote-Jarai³, D. Conti²², C. Haiman⁴; ¹Fred Hutchinson Cancer Ctr., Seattle, WA, ²Inst. of Cancer Res., Sutton, Surrey, United Kingdom, ³The Inst. of Cancer Res., London, United Kingdom, ⁴Univ. of Southern California, Los Angeles, CA, ⁵Natl. Cancer Inst, Rockville, MD, ⁶Natl. Cancer Inst., Rockville, MD, ⁷American Cancer Society, Atlanta, GA, ⁸Natl. Cancer Inst., Bethesda, MD, ⁹Univ. of Cambridge, Cambridge, United Kingdom, ¹⁰Cancer Council Victoria, Melbourne, Australia, ¹¹Monash Univ., Clayton, Australia, ¹²Dana Farber Cancer Inst., Boston, MA, ¹³Univ. of Oxford, Oxford, United Kingdom, ¹⁴Harvard Univ., Cambridge, MA, ¹⁵Harvard Sch. of Publ. Hlth., Boston, MA, ¹⁶Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, ¹⁷Johns Hopkins Univ, Baltimore, MD, ¹⁸Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ¹⁹Monash Univ., Melbourne, Australia, ²⁰The Inst. of Cancer Res., Sutton, United Kingdom, ²¹MEB, Stockholm, Sweden, ²²Univ of Southern California, Los Angeles, CA

Abstract Body:

BACKGROUND: Gene panel testing for prostate cancer can be used to identify men at high risk of aggressive disease. However, due to restricted sample sizes of previous exome sequencing investigations and the focus on candidate genes, evidence is limited for some panel genes, likely rendering them incomplete. We conducted a large case-only exome sequencing study to define genes with rare coding variation that predispose to aggressive prostate cancer. **METHODS:** This study included 9,185 aggressive cases (6,033 died due to prostate cancer and 1,730 had metastatic disease) and 8,361 non-aggressive cases of European ancestry from 19 international studies. Stage 1 samples (n=5,545) had whole-exome sequencing, and stage 2 samples (n=12,001) had targeted exome sequencing for 1,459 genes selected based on stage 1 results and previous evidence. Gene burden analyses were meta-analyzed across stages 1 and 2, aggregating rare (MAF<1%) deleterious variants to compare carrier frequencies in non-aggressive prostate cancer cases to 1) aggressive and 2) metastatic cases using logistic regression models adjusted for age, study, and principal components of ancestry. **RESULTS:** In addition to genes previously associated with aggressive prostate cancer (*BRCA2*, *ATM*, and *NBN*), we identified robust (P<0.002) but not exome-wide significant potentially novel prostate cancer susceptibility genes *MMP19* (involved in reproduction and metastasis), with carriers having 2.8-fold higher odds of prostate cancer death (95% CI=1.53-5.05, P=8x10⁻⁴); *PKD2L2* (involved in fertility), with carriers having 3.5-fold higher odds of prostate cancer death (95% CI=1.76-7.04, P=5x10⁻⁴); and *SMPD1* (involved in converting sphingomyelin to ceramide), with carriers having 5.3-fold higher odds of metastatic disease (95% CI=1.85-14.98, P=0.002). In a focused set of 36 prostate cancer gene panel genes and previously curated candidate prostate cancer DNA repair genes, we found strong evidence (P<0.05) of association with aggressive or metastatic disease for *BRCA2*, *ATM*, *NBN*, *MSH2*, *XRCC2*, *SLX4*, *MRE11A*, *POLK*, *POLH*, and *MSH5*. We observed intermediate evidence ([P<0.2 and OR>1] or [OR>1.8] for aggressive or metastatic disease) for *CHEK2*, *MLH1*, *PALB2*, *TP53*, *BRCA1*, *RAD51D*, *BARD1*, *GEN1*, *XPC*, *FAM175A*, *LIG4*, *POLD1*, *EIL2*, and *RDM1*. Deleterious variants in the 36 genes were carried by 15.7% of aggressive and 17.7% of metastatic cases compared with 11.7% of non-aggressive cases. **CONCLUSION:** These findings support the importance of rare genetic variation in aggressive prostate cancer risk and provide evidence for additional susceptibility genes that may be beneficial to include on prostate cancer gene panels.

Cancer Posters - Wednesday

PB1169. Next Generation SP DNA Sample Prep and DLS Labeling Readies Optical Genome Mapping Workflows for Adoption at Scale

Authors:

H. Sadowski, K. Pham, V. Alexiadis, M. Valdez, M. Yadav, Y. Zhang, L. Anna, C. Proskow, A. Files, R. Nieto, K. Fernandez, M. White, M. Oldakowski; Bionano Genomics, San Diego, CA

Abstract Body:

Movement through the Bionano optical genome mapping (OGM) workflow from biological sample to genome-wide structural variation analysis requires the isolation of ultra-high molecular weight (UHMW) genomic DNA (gDNA) and the Direct Label and Stain (DLS) of this UHMW gDNA for analysis on the Bionano Saphyr platform. We present significant evolution of the protocols for the isolation of UHMW gDNA and the DLS labeling of that isolated DNA towards adoption of OGM at larger scale.

The Bionano Prep SP family of kits and protocols has been designed specifically for the isolation of UHMW gDNA. The next generation workflow improvements presented here make it easier to use at scale both manually and on automation platforms. We present data showing improved robustness across a broader range of difficult starting materials, including bone marrow aspirates (BMAs). We demonstrate higher molecule size and data throughput when these samples are run on the Bionano Saphyr system. We also show how these improvements are adopted onto a liquid handling platform to form the first-ever automated UHMW gDNA isolation and purification solution.

The Bionano Prep DLS chemistry has been the backbone of genome-wide structural variation detection using labeled molecules collected on the Saphyr system. We present significant improvements that pave the way for adoption at scale and automation. The improvements allow for 12 samples to be processed at a single time (versus 6), a reduction in overall assay time of nearly 50%, and the streamlining of the sample-to-answer Saphyr workflow by 1 day.

Overall, these improvements allow for processing more samples through the Saphyr system and provide for a sample-to-answer turnaround time for complex blood cancers of 4 days.

Cancer Posters - Thursday

PB1170. No evidence of association between clonal hematopoiesis and risk of prostate cancer in large samples of European ancestry men.

Authors:

Y. Xu¹, A. Wang¹, K. Xu¹, Y. Yu², T. Dadaev³, E. Saunders³, X. Sheng¹, S. Chanock⁴, S. Berndt⁴, D. Albanes⁴, S. Weinstein⁴, V. Gnanapragasam⁵, G. Giles^{6,7,8}, T. Nguyen-Dumont^{7,8}, R. Milne^{6,7,8}, M. Pomerantz⁹, J. Schmidt^{10,11}, L. Mucci¹², W. Catalona¹³, K. Doheny¹⁴, R. MacInnis^{6,8}, M. Southey^{6,7,8}, R. Eeles^{3,15}, F. Wiklund¹⁶, Z. Kote-Jarai³, C. Huff², A. de Smith¹, D. Conti¹, C. Haiman¹, B. Darst¹⁷; ¹Univ. of Southern California, Los Angeles, CA, ²U.T. MD Anderson Cancer Ctr., Houston, TX, ³The Inst. of Cancer Res., London, United Kingdom, ⁴Natl. Cancer Inst., Bethesda, MD, ⁵Univ. of Cambridge, Cambridge, United Kingdom, ⁶Cancer Council Victoria, Melbourne, Australia, ⁷Monash Univ., Melbourne, Australia, ⁸Univ. of Melbourne, Melbourne, Australia, ⁹Dana-Farber Cancer Inst., Boston, MA, ¹⁰Univ. of Oxford, Oxford, United Kingdom, ¹¹Aarhus Univ., Aarhus, Denmark, ¹²Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ¹³Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, ¹⁴Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ¹⁵The Royal Marsden NHS Fndn. Trust, London, United Kingdom, ¹⁶Karolinska Inst.t, Solna, Sweden, ¹⁷Fred Hutchinson Cancer Res. Ctr., Seattle, WA

Abstract Body:

Little is known regarding the potential relationship between clonal hematopoiesis (CH) of indeterminate potential (CHIP), which is the expansion of hematopoietic stem cells with somatic mutations, and risk of prostate cancer, the fifth leading cause of cancer death of men worldwide. We evaluated the association of age-related CHIP with overall and aggressive prostate cancer risk in two large whole-exome sequencing studies of 86,884 European ancestry men, including 3,579 prostate cancer cases and 77,760 controls from the UK Biobank and 2,770 aggressive and 2,775 non-aggressive prostate cancer cases from a case-only study, Whole-Exome Sequencing Study in Prostate Cancer (WESP). Somatic CHIP variants (SNVs and Indels) were identified based on a list of 74 genes with previously reported mutations in human hematologic cancers. Somatic variant calling was carried out using the GATK Mutect2, and only variants meeting the following criteria were included: 1) minor allele frequencies (MAF) <0.1% in the respective UK Biobank or WESP data; 2) MAF <0.5% in the Genome Aggregation Database (gnomAD); and 3) variant allelic fraction (VAF) >10%. This led to a total of 1663 variants in 61 CHIP genes identified in the UK Biobank and 360 variants in 52 CHIP genes identified in WESP. We found that CHIP, defined by carrier status of CHIP genes individually and in aggregate, was not associated with overall (aggregate OR=1.13, 95% CI=0.88-1.45, P=0.34), aggressive (aggregate OR=0.91, 95% CI=0.73-1.14, P=0.43), or metastatic (aggregate OR=0.90, 95% CI=0.57-1.42, P=0.65) prostate cancer risk or death due to prostate cancer (aggregate OR=0.84, 95% CI=0.65-1.08, P=0.18). Further, CHIP was not associated with germline genetic risk of overall prostate cancer, measured using a polygenic risk score, or carrying germline pathogenic/likely pathogenic/deleterious variants in DNA repair genes, which have previously been found to be associated with aggressive prostate cancer. While findings from this study do not implicate CHIP as a risk factor for prostate cancer, it will be important to investigate other types of CH in association with prostate cancer risk, such as loss of chromosome Y, which is common and heavily influenced by germline genetics.

Cancer Posters - Wednesday

PB1171. Novel Germline Pathogenic Variants of *APC* and *BMPRIA* genes in Algerian patients with Hereditary Polyposis Syndromes

Authors:

F. Cherbal¹, F. Khider¹, A-L. Boumehdi¹, A-W. Damache¹, S. Sabri¹, K. Layada², F. Zebboudj³, H. Mahfouf⁴, M. Maaoui⁵; ¹Molecular Genetics Team, LBCM, Faculty of Biological Sci., USTHB, Algiers, Algeria, ²Gastroenterology services, Univ. Hosp. Mustapha Bacha, Sch. of Med., Univ. of Algiers-1, Algiers, Algeria, ³Mohamed El Kolli Publ. Hosp., Academic Med. Oncology Services, Sch. of Med., Univ. of Algiers-1, Rouiba, Algeria, ⁴Mohamed El Kolli Publ. Hosp., Academic Med. Oncology Services, Sch. of Med., Univ. of Algiers1, Rouiba, Algeria, ⁵Bachir Mentouri Publ. Hosp. , Academic Gen. Surgery Services, Sch. of medicine, Univ. of Algiers-1, Kouba, Algiers, Algeria

Abstract Body:

Background Hereditary polyposis syndromes (HPS) are a group of rare, inherited syndromes that can be present in various age groups of patients (from children to adults). Patients diagnosed with HPS have a high risk to develop colorectal cancer (CRC) and extracolonic cancer at early age. The present study focuses on two HPS: Familial Adenomatous Polyposis (FAP) and Juvenile Polyposis Syndrome (JPS). **Methods** The patients and their families were referred from 2012 to 2019 through six main medical oncology services and gastroenterology and general surgery services of public hospitals of Algiers. Clinical and pathological information was extracted from medical records of the patients with particular attention to the age at diagnosis, gender, history of CRC and instance of extracolonic cancers. Family histories of HPS were obtained from interviews, pedigrees and chart review of the index cases. We screened by PCR-direct sequencing the entire exon 15 of *APC* gene in 44 patients with strong family history of hereditary polyposis and two patients were analyzed by NGS using a cancer panel of 30 hereditary cancer genes (Color genomics). **Results** The analysis of DNA samples by PCR-Sanger sequencing of 44 patients revealed that 16 FAP patients carried out nine (09) distinct germline pathogenic variants in the *APC* gene. Out of these nine pathogenic variants, the novel *APC* variant c.2544dup detected in young FAP female index case and her sister has never been reported in individuals with FAP. The rare germline pathogenic variant c.4728dup has been detected in two unrelated FAP patients; hence, the codon 1577 could be a new mutational hot spot in Algerian FAP families. Interestingly, the recurrent *APC* germline pathogenic variant c.3927_3931del had occurred *de novo* in two unrelated FAP patients. NGS analysis revealed a novel *APC* germline pathogenic variant c.1605dup in young FAP proband with a strong history of FAP and 16 of his relatives were FAP affected along 3 generations. Subsequently genetic testing was performed in four relatives of this index case and they were found to carry the novel pathogenic variant c.1605dup. NGS analysis also revealed a novel *BMPRIA* pathogenic variant c.1474-1G>C in young JPS patient. We performed *in-silico* analysis for this novel variant and the results showed an alteration of the wild type acceptor site and an activation of a cryptic acceptor site, respectively, most probably affecting splicing. **Conclusions** The present study will contribute to the molecular genetics characterization of hereditary polyposis syndromes in Algerian families that is relevant for clinical management in the areas of genetic testing, early diagnosis, treatment and prevention.

Cancer Posters - Thursday

PB1172. Nucleotide resolution of large tandem duplications in cancer genomes

Authors:

P. Audano¹, F. Menghi¹, E. Liu², C. Beck³; ¹The Jackson Lab., Farmington, CT, ²The Jackson Lab, Bar Harbor, ME, ³The Jackson Lab. and Univ. of Connecticut Hlth.Ctr., Farmington, CT

Abstract Body:

Large scale chromosomal rearrangements are a hallmark of cancer genomes, and often lead to the amplification of proto-oncogenes or truncation of tumor suppressors. The Tandem Duplicator Phenotype (TDP) is common in breast, ovarian, endometrial, and liver cancers, and was first identified using short-read and array technologies. These experiments revealed three tandem duplication (TD) size groups associated with aberrations in *TP53* and *BRCA1* (group 1, ~11 kbp), *CCNE1* (group 2, ~231 kbp), and *CDK12* (group 3, ~1.7 Mbp) and showed TD clustering around genes; together these data suggest errors in DNA break repair underlying duplication formation. The precise mechanisms and full spectrum of variation resulting from TDs remains unknown because both the breakpoints of duplications and the sequence differences in the additional copy were not resolvable to basepair accuracy. We applied modern long-read sequencing technologies including PacBio HiFi and ultra-long ONT to a triple-negative breast cancer (TNBC) clonal pool known to harbor group 1 TDPs, MB436. In this cell line, we found long-read support for 87 of 92 validated duplications where missing validated TDs were small and in accessible loci suggesting they are missing in our clonal pool. Using standard tools, we fully assembled 75 TDs up to 495 kbp (2-7 assembled TDs per normal genome up to 27 kbp). We targeted an additional 17 large unassembled TDs resolved the duplication breakpoint as well as inter-duplication variants up to 20 kbp from breakpoints. By determining duplication breakpoints to basepair accuracy and analyzing duplicated sequence for somatic variants, we are constructing a database of mechanistic signatures and identifying mutations in duplicated genes resulting from DNA repair errors.

Cancer Posters - Wednesday

PB1173. OM2BFB: Detecting and elucidating Breakage Fusion Bridge structures in cancer genomes using Optical Mapping data

Authors:

S. Raeisi Dehkordi¹, J. Luebeck¹, N. Jing², I. Tsz-Lo Wong³, G. Xu⁴, L. Krockenberger¹, U. Rajkumar¹, A. Caplin¹, T. Xu¹, C. Coruh⁴, Q. Jin⁵, K. Ledward², K. Turner⁶, A. Pang⁷, J. Law⁴, F. Yue⁵, J. Zhao², P. Mischel³, V. Bafna¹; ¹Univ. of California San Diego, La Jolla, CA, ²Harvard Univ., Boston, MA, ³Stanford Univ., Palo Alto, CA, ⁴Salk, La Jolla, CA, ⁵Northwestern Univ., Chicago, IL, ⁶Boundless Bio, San Diego, CA, ⁷Bionano Genomics, San Diego, CA

Abstract Body:

Focal copy number amplification (fCNA) of oncogenes is a significant driver of cancer pathogenicity. Key fCNA mechanisms include Extrachromosomal DNA (ecDNA) formation and Breakage-fusion-bridge (BFB) cycles. Unlike ecDNA, BFB architectures have not been systematically explored. BFB formation starts with a telomere break, followed by anaphase bridge formation between sister chromatids. Multiple cycles of fusion bridge break to generate inverted duplications during cytokinesis, and subsequent fusions of broken ends in sister chromatids leads to rapid amplification of large (megabase sized) genomic intervals. BFBs can arise early during carcinogenesis, enable oncogene amplification and resistance to targeted therapy (e.g. HER2+ tumors), and have been implicated in neurodevelopmental disorders. These findings highlight the need for improved BFB detection and reconstruction--a challenge for short read data.

Optical Mapping technology provides physical maps that can be assembled into ultra-long OM assemblies (N50 ~ 50 Mbp), that overcome the challenge of analyzing large duplicated segments and fold-backs characteristic of BFB. We describe OM2BFB, a computational method for detection and reconstruction of BFB structures using OM data. OM2BFB rapidly filters for candidate BFBs, then enumerates candidate BFB structures and scores them using a novel likelihood formulation. In simulations, OM2BFB detected and reconstructed BFBs with nearly perfect recall.

We applied OM2BFB to 65 focal amplifications (28 samples), including cancer cell lines, and also breast cancer samples with brain metastases (PDX and primary). It detected and reconstructed 24 putative BFB structures in 16 different samples. On HCC827, the automated reconstruction including amplification of EGFR was identical to the previously manually reconstructed structure. While HER2 (chr17q) amplification dominated BFB structures (7 of 24 cases), BFB amplicons were found on 11 chromosomes, including 12 oncogenes. The structures often included translocation from other chromosomes, and other rearrangements. Telomere loss is a defining characteristic, but >90% percent of detected BFBs were closer to the centromere. Analysis of non-BFB structures revealed cases of ecDNA reintegration into chromosomes, while a few cases suggested ecDNA originating from BFB. BFB fCNA were lower in copy number (mean CN:27) and involved larger genomic segments (~2.2 Mbp) compared to ecDNA. Together, the results suggest that OM2BFB is a useful tool for elucidating the unique role of BFB in oncogene amplification.

Cancer Posters - Thursday

PB1174. Oncological biomarkers identified by NGS: characterization of the colombian population tumors from 2019 to 2022

Authors:

L. Parada Niño¹, M. Gálvez Bermudez¹, S. Bello Uyaban¹, M. Latorre Quintana¹, M. Lasso Carlosama²; ¹Gencell - Genuino Group, Bogotá, Colombia, ²Colina Clinic, Bogotá, Colombia

Abstract Body:

Justification: Cancer is one of the main causes of mortality worldwide and had an incidence of 182/1.000 habitants in Colombia, in 2020. Biotechnological advances such as next-generation sequencing (NGS), have allowed the analysis of tumor genomic profile, using liquid biopsy (cell-free circulating tumor DNA [cfDNA]) or tumor tissue, for the identification of oncological biomarkers that provide prognostic, predictive, and therapeutic information, also allowing the identification of germline variants. **Objective:** To describe the oncological biomarkers identified by NGS in liquid biopsy or tumor tissue samples, in a cross-section of Colombian patients with a histopathological diagnosis of cancer. **Methods:** Retrospective cross-sectional and descriptive observational study, in subjects with diagnosis of cancer, who had molecular testing by NGS in liquid biopsy or tumor tissue, and normalized with a germline sample, from May 2019 to April 2022. Participants have informed consent, according to the Helsinki's declaration. Using a NGS panel of less than 700 genes, genetic variants were identified including somatic and germline variants, as well as microsatellite instability status (MSI), and tumor mutational burden (TMB). **Results:** A total of 45 subjects were included, with a mean age of 62,8 years, 54,3% women, and 17,7% cases documented a family history of cancer. Molecular studies were performed for lung (20%), prostate (15.5%), breast (15.5%), pancreas (8.9%) and colon cancer (8.9%), among others. The analyses performed in liquid biopsy represented 75,55% of the cases. Clinically significant somatic variants were identified in 94% of cases, with an average of 3,6 variants per case. The most prevalent mutated genes were *TP53*, *AR*, *MAP3K1*, and *PTEN*. Tumor tissue cases represented 24,11% of the cohort. Biologically relevant somatic variants were identified in 81,8% of these tumors, where loss-of-function variants were predominant (90%). Mutations were more frequently found in *TP53*, *APC*, and *SMAD4* genes. MSI stable was identified in 100% of the cases. TMB was between 0,4 to 24,92 m/Mb for liquid biopsy and within 0,1 to 8,24 m/Mb in tumor tissue. Germline variants were reported in 6,6% of the cases, identified in *BRCA1*, *MUTYH*, *NF1*, and *BMPRIA* genes. **Conclusions:** This study provides important epidemiological data for Colombian cancer population, where NGS is an useful tool to detect biomarkers to potentially direct therapeutic schemes.

Cancer Posters - Wednesday

PB1175. Online implementation of combined polygenetic protein prediction algorithm for oncology (C3PO): Tumor mutation contributions to cancer hallmarks based on proteomics.

Authors:

G. R. G. Caryotakis, B. H. Song, T. H. West, M. H. Bailey; Brigham Young Univ., Provo, UT

Abstract Body:

The hallmarks of cancer, as defined by Hannahan and Weingberg (2011 and 2022), classify 13 essential characteristics linked to the development and progression of human cancer. Molecular contribution to cancer hallmarks can be measured using different next generation technologies, i.e., whole exome sequencing, RNA-sequencing, and high through proteomics measures using LC-MS-MS. However, little work has been done to establish the cumulative effect of tumor mutations on protein variation globally. Using 1061 samples collected by the Clinical Proteomic Tumor Analysis Consortium (CPTAC), we built a polygenic risk model to predict the genomic contribution to variable protein levels. Using genomics alone, we can correlate scores with protein data ranging from correlation coefficients of 0.03-0.90. While it may be difficult to predict all protein-levels, we reasoned that consistent changes across many proteins in a pathway might reveal biologically meaningful results. Here we outline a web tool developed in R Shiny, which provides sample-specific profiles for all 13 cancer hallmarks. Our bioinformatics approach empowers users to link any oncological genomic data to protein levels. Our tool can be used for both individual cancer types and cancer in general. When applied to the CPTAC lung adenocarcinoma (LUAD) cohort we observed distinct shifts in hallmarks that are correlated with smoking status- specifically the ‘non-mutational epigenetic reprogramming’ hallmark. Hallmark variations were mostly attributed to HDAC6, HDAC2, and SUV39H2 protein activity ($p=4.7 \times 10^{-6}$, $CI=-0.77 - -0.33$; $p=1.2 \times 10^{-3}$, $CI=0.15-0.57$; $p=3.4 \times 10^{-4}$, $CI=0.80-2.42$; respectively). Due to the modular nature of our tool, future iterations may include expanded datasets such as transcriptomic or methylomic data. Additionally, more specific cancer types can be included to improve predictions. We anticipate this will help identify prominent hallmark patterns in tumors leading to improved outcomes in modern precision medicine.

Cancer Posters - Thursday

PB1176. Ovarian sex cord stromal tumor is a likely additional phenotype of *BAP1*-Tumor Predisposition Syndrome

Authors:

M. Abdel-Rahman, M. Y. Hussein, C. Ingalls, C. Porteus, A. Ansari, L. Senter, C. M. Cebulla, L. Byrne; The Ohio State Univ., Columbus, OH

Abstract Body:

Objective: Germline mutations in the tumor suppressor gene *BAP1* are associated with the hereditary tumor predisposition syndrome, *BAP1*-TPDS (OMIM 614327), with predisposition to four main cancers: uveal melanoma, mesothelioma, cutaneous melanoma, and renal cell carcinoma. However, other cancers are likely part of the clinical phenotype. Here, we report the potential association of ovarian stromal sex cord (OSSC) cancers with *BAP1*-TPDS. **Methods:** A total of 3321 unselected cancer patients enrolled through the Total Cancer Care (TCC) Protocol, at the Ohio State University Comprehensive Cancer Center and were assessed for germline *BAP1* variants by secondary analysis of research exome sequencing. Exome sequencing was carried out through the Oncology Research Information Exchange Network (ORIEN) AVATAR program. Selected *BAP1* variants were confirmed by direct sequencing. Additionally, subjects with germline *BAP1* variants identified through routine clinical testing were enrolled in a separate protocol (NCT04792463). Cascade testing for at risk family members was then performed for P/LP or suspicious variant of uncertain significance (VUS). **Results:** We identified 15/3321 missense VUSes in the tested cohort of unselected cancer patients. Fourteen of these were located outside the ubiquitin carboxy-terminal hydrolase (UCH) domain of BAP1 (amino acids:1-240). One of these VUS (c.86T>G, p.V29G) was located in the active UCH domain of BAP1 and was identified in a female patient with an OSSC cancer. The variant is not reported in the general population and predicted deleterious by multiple computational tools. The patient had a family history consistent with *BAP1*-TPDS including multiple mesotheliomas, as well as other cancers. The ovarian tumor of this patient showed evidence of biallelic inactivation of BAP1. We reviewed our clinical cohort of germline *BAP1* P/LP patients and their positive family members. These included 100 individual (31 males and 69 females) from 50 families. We identified three with ovarian cancer; one had OSSC cancer, one had papillary serious ovarian carcinoma and the pathology was not available on the third. A literature search identified two additional OSSC cancers in patients with germline *BAP1* P/LP. Furthermore, OSSC tumors were the most common spontaneous cancer observed in murine models of germline heterozygous *BAP1* P/LP variants. **Conclusions:** Given the rarity of OSSC in general population, our results in addition to published data, strongly suggest that it is likely part of the clinical phenotype of *BAP1*-TPDS. Validation studies in independent cohorts are warranted.

Cancer Posters - Thursday

PB1177. *PALB2*-mutated human mammary cells display a broad spectrum of morphological and functional abnormalities beyond defective DNA damage response.

Authors:

R. Winqvist^{1,2}, H. Tuppurainen¹, N. Laurila¹, M. Nätyinki¹, L. Eshraghi^{1,3}, A. Tervasmäki¹, L. Erichsen⁴, C. Storgaard Sørensen⁴, K. Pylkäs^{1,2}, H. Peltoketo¹; ¹Univ Oulu, Oulu, Finland, ²Northern Finland Lab. Ctr., Oulu, Finland, ³Garvan Inst. of Med. Res., Sydney, Australia, ⁴Univ Copenhagen, BRIC, Copenhagen, Denmark

Abstract Body:

Heterozygous mutations in any of three major genes, *BRCA1*, *BRCA2* and *PALB2*, are associated with high-risk hereditary breast cancer susceptibility frequently seen as familial disease clustering. *PALB2* is a key interaction partner and regulator of several vital cellular activities of *BRCA1* and *BRCA2*, and thus it is required for DNA damage repair and alleviation of replicative and oxidative stress. Little is however known about how *PALB2*-deficiency affects cell function beyond that, especially in the three-dimensional setting, but also about its role during early steps of malignancy development. To answer these questions, we have generated biologically relevant isogenic cell lines with mutations that are comparable to certain clinically important *PALB2* defects. We show in a non-cancerous background how both mono- and biallelically *PALB2*-mutated cells exhibit gross spontaneous DNA damage and mitotic aberrations. Furthermore, *PALB2*-deficiency disturbs three-dimensional spheroid morphology, increases the migrational and invasional capacity of the cells, and broadly alters their transcriptome profiles. TGFB signaling has been enhanced in *PALB2*-mutated cells and its inhibition partially rescues the cells. Independent of DNA damage, expression of several genes associated with cell adhesion and migration has also been triggered in *PALB2*-mutated cells. The obtained results indicate comprehensive changes in cell function upon *PALB2* mutations, even in the presence of half a dosage of wild type *PALB2*, and demonstrate how *PALB2* mutations may predispose their carriers to malignancy.

Cancer Posters - Wednesday

PB1178. Pan-Cancer Analysis introduces HLF as a tumor suppressor gene and a promising biomarker for prognosis and targeted therapy in human tumors

Authors:

M. Ahmadi¹, S. Ghaderian²; ¹Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, ²Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of

Abstract Body:

Hepatic leukemia factor (HLF) is considered a clock-dependent transcription factor. Even though prior research has linked the aberration of the *HLF* gene to cancer development, no pan-cancer assessments of the *HLF* gene have been published. Therefore, we employed several databases to demonstrate the possible functions of the *HLF* gene in various cancer types in this investigation. We analyzed the expression level of the *HLF* gene across TCGA cancers using the UALCAN, GEPIA2, and GSCA databases. We further evaluated the association of the *HLF* gene with the overall survival of cancer patients. We also assessed the genetic alterations, the methylation level of the promoter region, protein-protein interaction of the *HLF* gene, and the correlation between HLF expression and immune cell infiltration in various tumors utilizing the cBioPortal, DNMIVD, GeneMANIA, and TIMER2 databases, respectively. Besides, data from the GSCA database was used to investigate the possible roles of the HLF gene and the relationship between HLF expression and drug sensitivity. We determined that the *HLF* gene was downregulated in tumor tissues of a wide range of cancers compared to corresponding normal tissues. We also detected a significant association between abnormal expression of *HLF* and patient survival, genetic mutations, and immune infiltration. In addition, pathway analysis indicated that *HLF* affected Apoptosis, Cell Cycle, EMT, and PI3K/AKT signaling pathways. Finally, we showed that abnormal expression of the *HLF* gene could reduce the sensitivity to several anti-tumor agents and small molecules in the GDSC dataset. Our findings propose that lower expression of the *HLF* gene mediates carcinogenesis and cancer progression and could be recognized as a valuable biomarker for prognosis, immunotherapy, and target treatment in various cancers.

Cancer Posters - Thursday

PB1179. Pan-cancer polygenic risk distinctly contributes to prediction of heterogeneous subsequent malignancies across treatment profiles in survivors of childhood cancer.

Authors:

C. Im¹, N. Sharafeldin², Y. Yuan¹, Y. Sapkota³, Z. Wang³, M. A. Arnold⁴, G. T. Armstrong³, M. M. Hudson³, K. K. Ness³, L. L. Robison³, S. Bhatia², J. P. Neglia⁵, L. M. Turcotte⁵, Y. Yasui³; ¹Univ. of Alberta, Edmonton, AB, Canada, ²Univ. Alabama at Birmingham, Birmingham, AL, ³St. Jude Children's Res. Hosp., Memphis, TN, ⁴Univ. of Colorado, Aurora, CO, ⁵Univ. of Minnesota, Minneapolis, MN

Abstract Body:

Subsequent malignant neoplasms (SMNs) contribute substantially to late morbidity/mortality among long-term survivors of childhood cancer. Given growing evidence of common biological hallmarks for cancers including indications of a shared genetic basis (pleiotropy), we explored the impact of pan-cancer polygenic risk factors on overall SMN risk in 10,137 childhood cancer survivors. Participants in the Childhood Cancer Survivor Study and St. Jude Lifetime Cohort Study (European genetic ancestry) with genotype data, medical record-abstracted treatment exposures, and pathology-confirmed SMNs were analyzed. Top-performing published genome-wide polygenic risk scores (PRSs) for specific cancers and “any cancer” in the Michigan Genomics Initiative (MGI) study (Cancer-PRSWeb; prsweb.sph.umich.edu) were computed, where pan-cancer PRSs came from UK Biobank GWAS (N>300K). Hazard ratios (HRs) for any SMN incidence were estimated by Cox regression, with age, sex, batch/cohort, fine-scale ancestry, and primary cancer diagnosis/treatment (e.g., radiation therapy [RT], alkylating agents, anthracyclines, epipodophyllotoxins) as covariates. A total of 1,022 SMNs were observed among 922 survivors (total N=10,137; 242,962 person-years of follow-up), including subsequent breast (25.5% of SMNs, N=237) and thyroid (23.6%; N=218) cancers and basal cell carcinomas (BCCs; 21.8%, N=201). Since the top pan-cancer PRS in MGI (lassosum with 179 variants, PRS₁₇₉) performed similarly across the two cohorts and combined sample, we report combined results. PRS₁₇₉ was significantly associated with pan-SMN risk (HR per PRS SD=1.12, P=6.2x10⁻⁴). This association was consistently attenuated among survivors who received any RT (top vs. bottom PRS₁₇₉ quartiles, HR=2.13, P=4.5x10⁻⁴ if no RT vs. HR=1.19, P=0.13 if any RT), while anthracyclines enhanced risk (HR=2.79, P=1.1x10⁻³ if no RT, anthracyclines >0 mg/m²). Top PRSs for breast cancer, thyroid cancer, and BCC in MGI were significantly associated with corresponding specific SMN risks (HRs=1.29 to 1.37, P=5.9x10⁻¹⁴ to 1.5x10⁻⁴). Although these cancer subtypes were common (~60% of SMNs), pan-cancer PRS₁₇₉ remained significantly associated with pan-SMN risk (HR per PRS SD=1.11, P=4.1x10⁻³) while the three cancer-specific PRSs did not (P>0.06) in a model with all four scores. In summary, pan-cancer polygenic risk factors can predict overall SMN risk among childhood cancer survivors despite the heterogeneity of SMNs, depending on treatments received. Further research to refine pan-SMN genetic risk predictors reflecting common carcinogenic hallmarks and their effect modification by treatments is needed.

Cancer Posters - Wednesday

PB1180. Parent-of-origin-aware genomic analysis in the management of patients and families with genetic disease

Authors:

K. Schrader^{1,2}, V. Akbari¹, V. Hanlon², K. O'Neill², K. Dixon¹, Y. Shen², A. Sharma², J. Senz¹, Y. Wang¹, D. Chan², A. Fok², Q. Hong¹, D. Regier², A. Virani¹, L. Lefebvre¹, F. Feldman², M. Marra^{2,1}, S. Sun², S. Yip^{1,2}, P. Lansdorp^{1,2}, S. Jones^{1,2}; ¹Univ. British Columbia, Vancouver, BC, Canada, ²BC Cancer, as part of the Provincial Hlth.Services Authority, Vancouver, BC, Canada

Abstract Body:

Parent-of-origin-aware genomic analysis using Oxford Nanopore Technologies long-read sequencing combined with Strand-seq is a novel breakthrough method that enables assignment of variants to each biological parent with >99% accuracy using only the blood sample of the child. We demonstrate the analytic validity and clinical utility of this method of parent-of-origin-aware genomic analysis in *SDHD* pathogenic variant carriers with known parental segregation. Pathogenic variants in *SDHD* have parent-of-origin effects with disease penetrance for high lifetime risk for paragangliomas and pheochromocytomas that are dependent on transmission of the pathogenic variant through the male gamete. Hence knowledge of the parent-of-origin of *SDHD* pathogenic variants is essential when advising on management with significant clinical implications for lifelong surveillance. When parent-of-origin of the *SDHD* pathogenic variant is unknown, patients may undergo potentially unnecessary lifelong surveillance or forego necessary surveillance due to their uncertain risk estimate. Potentially unnecessary surveillance directly and indirectly burdens both the patient and the health care system and includes physical, practical, psychological, and other burdens. Beyond management of the patient, parent-of-origin-aware genomic analysis may help inform variant curation, recurrence risks, and efficiently direct cascade genetic testing throughout the family. While the added dimension to genetic testing will improve patient-centered care, there will be ethical considerations related to inference of a parent's germline variant status with >99% accuracy. Although obligate carrier status is not a new concept, routinely predicting carrier status with high accuracy could have significant duty-to-warn implications, especially in the setting of actionable diseases and secondary findings. These issues will need to be considered as the applications of parent-of-origin-aware genomic analysis expand and as uptake of this technology transforms precision medicine for both patients and families with genetic disease.

Cancer Posters - Thursday

PB1181. Pathogenic genetic variants from highly connected cancer susceptibility genes confer the loss of structural stability and further link with diabetes

Authors:

M. Hossain¹, N. Ferdous², M. Mohini³, E. Md. Tabassum Hossain⁴, M. Islam⁵, A. Mohiuddin²; ¹Natl. Inst. of Biotechnology, Savar, Bangladesh, ²Mawlana Bhashani Sci. and Technology Univ., Tangail, Bangladesh, ³Mawlana Bhashani Sci. and Technology, Tangail, Bangladesh, ⁴Natl. Inst. of Biotechnology, Tangail, Bangladesh, ⁵Mawlana Bhashani Sci. and Technology Univ., Savar, Bangladesh

Abstract Body:

Genetic polymorphisms in DNA damage repair and tumor suppressor genes have been associated with increasing the risk of several types of cancer. Analyses of putative functional single nucleotide polymorphisms (SNP) in such genes can greatly improve human health by guiding the choice of therapeutics. In this study, we selected nine genes responsible for various cancer types for gene enrichment analysis and found that BRCA1, ATM, and TP53 were more enriched in connectivity. Therefore, we used different computational algorithms to classify the nonsynonymous SNPs which are deleterious to the structure and/or function of these three proteins. Our study demonstrated that V1687G and V1736G variants of BRCA1, I2865T and V2906A variants of ATM, V216G and L194H variants of TP53 are major mutations with pathogenic impact and are likely to have a greater impact on destabilizing the proteins. To stabilize the high-risk SNPs, we performed mutation site-specific molecular docking analysis and validated using molecular dynamics (MD) simulation and molecular mechanics/Poisson Boltzmann surface area (MM/PBSA) studies. Additionally, SNPs of untranslated regions of these genes affecting miRNA binding were characterized. Besides, we have collected the human blood sample (n = 750) from the cancer-affected people and performed a population genetics study. We performed the PCR-RFLP method to observe the frequency of the genotype and found significantly associated with the genes BRCA1 (P = .001), ATM (P = .016), and TP53 (P = .011) of cancer affected patients. Interestingly, the metadata suggests that the genotypes are also significantly associated with diabetes patients (n = 370). Furthermore, we have found a strong relationship between cancer and diabetes from the exploration of systems biology and metadata. Hence, this study will assist in developing precision medicines for cancer types related to these polymorphisms and diabetes.

Cancer Posters - Wednesday

PB1182. Pathogenic Variants in Adult-Onset Cancer Predisposition Genes Identified by Pediatric Tumor Genomic Testing: Prevalence, Tumor Features, and Germline Testing Outcomes

Authors:

N. Oak¹, R. B. McGee¹, L. Harrison¹, K. Xu², A. K. Blake¹, R. Nuccio¹, R. Mostafavi¹, S. Lewis³, L. Taylor¹, M. Kubal¹, A. Ouma¹, S. Hines-Dowell¹, K. E. nichols¹; ¹Dept. of Oncology, St. Jude Children's Res. Hosp., Memphis, TN, ²Ctr. for Applied Bioinformatics, St. Jude Children's Res. Hosp., Memphis, TN, ³Dept. of Hematology, St. Jude Children's Res. Hosp., Memphis, TN

Abstract Body:

Background: Genomic sequencing of pediatric tumors is uncovering pathogenic variants (PVs) in adult-onset cancer predisposition genes (aoCPGs), some of which are germline in origin. Nevertheless, it remains uncertain whether aoCPG PVs contribute to pediatric tumor formation and unclear how germline aoCPG data inform the care of children with cancer. To address these questions, we examined the prevalence and spectrum of tumor aoCPG PVs, tumor molecular features, and germline testing and clinical management outcomes among children and young adults with cancer who underwent genomic testing at our institution between 2017-2019.

Results: Tumor genomic reports from 1,023 patients were queried for the presence of PVs in 25 pre-selected aoCPGs. Thirty-eight patients (4%) had 39 tumors harboring 41 PVs affecting 13 aoCPGs. Analysis of available whole genome sequencing data from 28 of these 39 (72%) tumors revealed mutational signatures consistent with defective DNA repair, including homologous recombination (HR) in tumors with *ATM*, *BRCA2*, *PALB2*, *CHEK2* PVs (n=8 of 18, COSMIC signature SBS3), mismatch repair in tumors with *MSH2*, *MSH6*, *PMS2* PVs (n=5 of 5, SBS15, SBS26), and base excision repair in tumors with *MUTYH* PVs (2 of 3, SBS36). Notably, only seven of these 15 mutation-signature positive tumors harbored genomic lesions affecting (n=6) or presumably affecting (n=1) the second aoCPG allele. We analyzed HR deficiency (HRD) using scarHRD, which incorporates information on genomic loss of heterozygosity, telomeric allelic imbalance and large-scale state transitions. Two tumors (both with biallelic *ATM* inactivation) had significantly higher HRD scores compared to tumors with monoallelic HR gene inactivation (Wilcoxon p-value=0.019, n=16) or in tumors without any HR gene PVs (p=0.012, n=212). Thirty patients completed germline analysis with 25 (83%) testing positive and 10 exhibiting consistent family histories of cancer. Despite high uptake of germline testing, only five patients' care changed based on germline findings: a young adult with *PMS2* and *AXIN2* variants for whom colon cancer screening was initiated and four children for whom germline data was used to direct donor selection for stem cell transplantation.

Conclusion: In sum, 4% of pediatric tumors harbored aoCPG PVs, with at least 36% (15 of 41 total PVs) demonstrating an associated DNA mutation signature. Among 18 tumors with HR gene PVs, only two exhibited evidence of HR deficiency. Further studies are warranted to determine whether or how aoCPG PVs contribute to pediatric tumor formation. There was unanimous parental interest in germline testing of aoCPGs despite limited direct clinical impact.

Cancer Posters - Thursday

PB1183. Phylogeny-aware detection of single-nucleotide variants and mode of evolution from cancer single-cell DNA sequencing data.

Authors:

M. Edrisi¹, M. Valecha², S. Chowdary³, S. Robledo⁴, H. Ogilvie¹, D. Posada², H. Zafar⁵, M. Li¹, L. Nakhleh¹; ¹Rice Univ., Houston, TX, ²Univ. of Vigo, Vigo, Spain, ³Indian Inst. of Technology Kanpur, Kanpur, India, ⁴Univ. of Houston, Houston, TX, ⁵Indian Inst. of Technology Kanpur, Kanpur, India

Abstract Body:

The growth of a tumor can be regarded as an evolutionary interplay between the mutations and the selective pressure in the tumor environment. Cancer evolutionary biology studies the timing and type of the mutations, how they accumulated, and the selective sweeps that occurred during this evolutionary process by reconstructing the tumor phylogenetic tree. With the advent of single-cell DNA sequencing (scDNA-seq), it is now possible to infer the phylogenetic tree of a tumor at a single-cell resolution. This promising opportunity comes with the burden of elevated error rates, allelic dropout, and nonuniform coverage in the scDNA-seq data that complicates the mutation detection required for phylogeny reconstruction. Here we report on two methods we developed: Phylovar (previously published) and MoTERNN. We developed Phylovar to overcome the challenge of mutation detection from scDNA-seq data by incorporating the evolutionary history of the cells. Later, we developed MoTERNN to infer the evolutionary pattern of tumor growth from scDNA-seq data.

In Phylovar, we followed up on phylogeny-aware mutation detection for single-nucleotide variations (SNVs), which was previously shown to be an effective approach to overcome the technical noise in single-cell sequencing protocols by the methods SCIΦ and scVILP. In this approach, the phylogeny of single cells plays the role of a nuisance parameter constraining the SNV detection method. Despite being accurate, these methods are not scalable to the extensive genomic breadth of single-cell whole-genome and whole-exome sequencing data. Phylovar extended the phylogeny-guided variant calling approach to sequencing datasets containing millions of loci while being as accurate as its preceding methods.

In MoTERNN, we focused on determining the mode of evolution of tumor cells from their inferred evolutionary history and SNV profiles acquired from scDNA-seq data. Here, we employed recursive neural networks that capture tree structures to classify the evolutionary history of tumor cells into one of four evolutionary modes—linear, branching, neutral, and punctuated. We trained our model, MoTERNN, using simulated data in a supervised fashion and applied it to the phylogenetic tree obtained from scDNA-seq of a triple-negative breast cancer patient. It is worth mentioning that we used Phylovar to obtain this phylogenetic tree and the SNV profiles of the cells as the input to MoTERNN. MoTERNN identified a punctuated mode of evolution for this patient which was in agreement with the results of the original study.

Cancer Posters - Wednesday

PB1184. Potential misrepresentation of inherited breast cancer risk by common germline alleles.

Authors:

W. Letsou¹, F. Wang¹, W. Moon¹, C. Im², Y. Sapkota¹, L. Robison¹, Y. Yasui^{1,2}; ¹St. Jude Children's Res. Hosp., Memphis, TN, ²Univ. of Alberta, Edmonton, AB, Canada

Abstract Body:

Hundreds of common variants have been found to confer small but significant differences in breast cancer risk, supporting the polygenic additive model of inherited risk. This widely accepted model is at odds with twin data indicating highly elevated risk in a subgroup of women. Using a novel closed-pattern-mining algorithm called Chromosome Overlap, we provide evidence that rare variants or haplotypes may underlie the association of breast cancer risk with common germline alleles. Our method consists in iteratively pairing chromosomes from affected individuals and looking for noncontiguous patterns of shared alleles without exhaustive enumeration. We applied Chromosome Overlap to haplotypes of genotyped SNPs from 9,011 female breast cancer cases from the UK Biobank (UKBB) at three topologically associating domains containing well-established common-allele “hits” for breast cancer risk near *FGFR2*, *CCND1*, and *CASC16*. A total of 181,034 UKBB women of “white British” ancestry were used to assess the discovered haplotypes, and 55,346 cases and controls of European ancestry in the Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) case-control study were used for replication. To account for differences between the UKBB Axiom Array and the DRIVE OncoArray, imputation and phasing for all samples was done using the TOPMed Imputation Server at UKBB genotyped SNPs. After using Chromosome Overlap to mine closed haplotype patterns among case chromosomes on 50–60 SNPs within ~100 kb of (but not necessarily including) the three GWAS hits, we found three respective strongly linked common haplotypes ($LD\ r^2 \geq 0.7$, $D' \geq 0.99$) with similar effect sizes (hazard ratios: 1.2–1.3); to find rare haplotypes, we performed a second phase of the analysis in adjacent ~400-kb regions containing ~100 SNPs each on chromosomes carrying the common haplotype. Out of twenty rare (frequency < ~0.1%) risk haplotypes of large effect identified in UKBB at $p < 1.0 \times 10^{-5}$, four (hazard ratio: 4.22–20.2) were subsequently replicated in DRIVE (odds ratio: 2.13–11.9) at $p < 0.05$. This level of replication was found to be robust by permutation analysis. Because we had low power to detect such rare haplotypes and were limited to UKBB genotyped SNPs, our successful replication of four of them suggests there could be many similar haplotypes which underlie the association of other GWAS hits with disease risk. Our results are also consistent with the genetic heterogeneity of breast cancer risk and suggest a novel type of “synthetic association” wherein common risk alleles on a rare risk haplotype may misrepresent disease risk through their tagging of many “false positive” haplotypes.

Cancer Posters - Thursday

PB1185. Potential subtype-informative genetic risk variants and genes identified from a case-case analysis for breast cancer

Authors:

X. Sun^{1,2}, X. GUO³, A. Reiner¹, G. Watt¹, J. Bernstein¹, X. Shu¹; ¹Mem. Sloan Kettering Cancer Ctr., New York, NY, ²Zhejiang Chinese Med. Univ., Hangzhou, China, ³Vanderbilt Univ. Med. Ctr., Brentwood, TN

Abstract Body:

Background: Breast cancer, comprised of subtypes with distinct pathological and molecular features, is highly heterogeneous. Compared to overall breast cancer, the understanding of genetic architecture for the subtypes have lagged. **Methods:** We obtained summary statistics data of genome-wide association studies (GWAS) conducted in intrinsic-like breast cancer subtypes, including luminal A-like, luminal B-like, luminal B/Her2-negative-like, Her2-enriched-like and triple negative breast cancer (TNBC) from the Breast Cancer Association Consortium (BCAC, total 106,278 cases and 91,477 controls). We applied linkage disequilibrium score (LDSC) regression to infer single-nucleotide variant (SNV) based heritability for each subtype and calculated estimates of pairwise genetic correlations among the subtypes. We then conducted case-case comparisons among the intrinsic-like breast cancer subtypes using the newly developed case-case GWAS (CC-GWAS) approach. The conditional and joint multiple-SNV analysis was further conducted for the identified variants, as well as a gene-based analysis across the genome using MAGMA. **Results:** LDSC analysis revealed a moderate to high genetic correlation among the five intrinsic-like subtypes (ranging from 0.52 to 0.88). We identified 13 significant risk loci from the pairwise case-case comparisons across the five intrinsic-like subtypes ($P < 5 \times 10^{-8}$) and additional 8 loci with a suggestive association ($P < 5 \times 10^{-7}$), all in close proximity to previously reported breast cancer susceptibility loci (<500 Kb). Among them, 12 loci were independent from previously reported breast cancer susceptibility variants in the conditional analysis (Bonferroni corrected $P < 0.05$). Eight of the 12 loci were identified from the comparisons between TNBC and luminal A-like cancers, followed by TNBC versus luminal B/Her2-negative-like (n=1) and Her2-enriched-like versus luminal A-like (n=1). Two loci identified from the comparisons between TNBC and luminal A-like were also found in the comparisons of TNBC versus luminal B-like (8q22.3) and TNBC versus luminal B/Her2-negative-like (11q22.3). The gene-based analysis also identified 32 significant risk-associated genes after the Bonferroni correction, of which 19 genes were specific to TNBC when compared to luminal A-like cancer. **Conclusion:** This study showed the complexity in the heterogeneity of breast cancer genetic susceptibility and provided new insights into the etiology of breast cancer especially for TNBC.

Cancer Posters - Thursday

PB1187. Prevalence of mutations in hereditary cancers of Czech patients.

Authors:

M. Koudová¹, L. Cerna¹, S. Chvojka¹, M. Sekowska¹, M. Urbanova¹, M. Bittoova¹, F. Zembol¹, B. Honysova¹, H. Dohnalova¹, A. Puchmajerova¹, D. Stejskal²; ¹GENNET s.r.o, Prague 7, Czech Republic, ²GENNET s.r.o., Praha 7, CZ

Abstract Body:

Introduction: Czech patients fulfilling clinical criteria for hereditary cancer (HC) are tested by NGS panel CZEKANCA (PMID: 29649263, Seq-Cap EZ Choice, Roche) that includes 226 genes associated with HC. Methods: Total 7286 DNA samples - patients (4729) and healthy relatives (2557) were tested. The most indications were for the breast, ovarian and colorectal HC. We have developed a bioinformatic pipeline using local installation of Ensembl genomic database for annotation and own variant database CheckBase for data handling and clinical reporting. Clinical interpretation was prioritised according to primary indication, revision based on IACR* standards, ACMG guidelines (PMID: 25741868) and the HGVS genetic nomenclature. Detection of exon deletions and duplications and larger gene rearrangements was performed by analyzing the coverage of NGS data on the target areas of the CZEKANCA panel. Positive results were validated by direct sequencing and CNV findings by MLPA. Results: Clinical important variants (class 5 and 4)* were detected in 20,1% samples and VOUS (class 3)* in 15,6% samples. The most variants were detected in genes BRCA1 (18%), BRCA2 (12%), CHEK2 and MUTYH (7%), NBN (6%), ATM and PALB2 (4%), NF1 (3%), MLH1, FH, FANCA, FANCM (all 2%). Conclusion: These results highlight the importance of extending the examination to other cancer susceptibility genes, because *BRCA1/2* gene mutations are responsible only for 1/3 of the heritable mutations in HC.

Cancer Posters - Wednesday

PB1188*. Prevalence of suspected pathogenic or likely pathogenic germline *TP53* variants in population sequencing databases: A genotype-first analysis.

Authors:

K. De Andrade¹, N. Strande², J. Kim¹, J. Haley², J. Hatton¹, M. Frone¹, P. Khincha¹, G. Thone², U. Mirshahi², J. Dove², A. Levine³, D. Stewart¹, D. Carey², S. Savage¹; ¹Natl. Cancer Inst., Rockville, MD, ²Geisinger, Danville, PA, ³Simons Ctr. for Systems Biology, Inst. for Advanced Study, Princeton, NJ

Abstract Body:

Pathogenic or likely pathogenic (P/LP) germline *TP53* variants are associated with Li-Fraumeni syndrome (LFS), a hereditary cancer predisposition disorder characterized by early-onset cancers. While the true prevalence of LFS is unknown, population-based studies suggest that the prevalence of P/LP germline *TP53* variants in the general population may be higher than previous estimates based on ascertainment of severely affected families. Genotype-first approaches can overcome selection biases present in most traditional phenotype-based studies. Analysis of individuals unselected for specific traits allows the detection of carriers of germline variants that would not otherwise be identified. We identified suspected germline *TP53* variants in two electronic health records (EHR)-linked sequencing cohorts, UK Biobank (n=200,600) and Geisinger healthcare system (n=175,449). Variants with allele fraction \geq 30% in exonic regions were classified based on ClinVar and InterVar. A total of 56 individuals from Geisinger (56/175,449; 0.03%) and 22 from UK Biobank (22/200,600; 0.01%) were detected with suspected P/LP germline *TP53* variants; including 35 females and 43 males. Median age at first visit was 64 years for females (range 18-80) and 62 years for males (range 22-80). Sixty-six individuals were cancer-free at least until the age of 60. Forty-seven individuals were diagnosed with a first cancer (median age at diagnosis 67 years; range 22-84) and 15 had developed multiple primaries on analysis of last available EHR data. Hematologic cancers were the most common first diagnoses. There were 54 unique P/LP variants, with p.R181H being the most frequent (9/78; 11.5%). Three of the nine p.R181H carriers had been diagnosed with cancer (bladder, prostate, and thyroid). We corroborate conservative population-based prevalence estimates of P/LP germline *TP53* variants in the range of 1:5,000 individuals of predominant European ancestry. However, the median ages at cancer diagnosis in the UK Biobank and Geisinger cohorts were 30 years higher than in LFS cohorts. This survival bias suggests that most of the *TP53* variants detected in our study may be associated with low cancer penetrance. Despite using an allele fraction cut-off \geq 30%, the relatively high number of hematologic cancers suggests there could be some somatic *TP53* variants occurring due to clonal hematopoiesis of indeterminate potential whose allele fractions resemble that of germline origin. Genotype-first analyses are valuable in identifying suspected P/LP germline *TP53* variants associated with variable cancer penetrance and with non-traditional LFS phenotypes.

Cancer Posters - Thursday

PB1189. Primary myelofibrosis and progression to acute myeloid leukemia in a young patient with variant in *CARL*

Authors:

A. Tamayo Palacio¹, N. Huerta Bolfeta¹, C. Afanador², G. C. Ramirez², S. Villa-Perez³, L. Gaviria Jaramillo³, G. Vasquez Palacio⁴; ¹Hosp. Infantil de México Federico Gómez, Ciudad de México, Mexico, ²Laboratorio Integrado de Med. Especializada LIME, Facultad de Med., Univ. de Antioquia, Medellín, Colombia, ³Hosp. San Vicente Fundación, Medellín, Colombia, ⁴Laboratorio Integrado de Med. Especializada LIME, Facultad de Med., Univ. de Antioquia, Medellín, Colombia

Abstract Body:

Introduction: Myeloproliferative neoplasms (MPNs) are clonal disorders characterized by increased proliferation of erythrocytes, megakaryocytes and mature granulocytes. Among the Philadelphia negative (Ph-) myeloproliferative neoplasms are Polycythemia Vera, Primary Myelofibrosis and Essential Thrombocythemia. In recent years, somatic mutations have been identified in the JAK2, MPL and CALR genes, which are associated with the progression of Ph- MPNs and are currently considered fundamental markers for the diagnosis of these neoplasms.

Case report: We present the case of a 35-year-old female patient with no relevant history, who was initially treated in Panama for abnormal uterine bleeding, severe anemia (Hb 3 g/dL) and pancytopenia. Gynecological alterations were excluded and the patient decides to return to her hometown. At the San Vicente Fundación Hospital, in Medellín, Colombia, the bone marrow (BM) biopsy showed findings suggestive of myelofibrosis and cells suggestive of blasts and immunohistochemistry was compatible with acute myeloid leukemia (AML) associated with grade 3 fibrosis. Myelogram: 9% blasts, hypercellular cord with dysplastic changes in erythroid and myeloid lines. Infectious causes were excluded and molecular studies were requested for BCR-ABL, JAK2, MPL and CALR. Negatives for BCR-ABL1, MPL and JAK2 (V617F), it also presented a 5bp deletion in exon 12 (rs56241661) in JAK2, and an 11bp deletion in exon 9 in CALR

Materials and methods: Clinical evaluation of the patient, bone marrow biopsy, immunohistochemistry, immunophenotype, qRT-PCR for BCR-ABL1, JAK2 and capillary electrophoresis in MPL and CALR. Discussion: CALR located at 19p13.2, codes for the calreticulin protein, which has functions in calcium homeostasis, chaperone activity in protein and glycoprotein folding, apoptosis and immunogenic cell death. Variants in CALR have been correlated with the origin of MPN, especially primary myelofibrosis (PMF) and essential thrombocytosis (ET); currently being the second most important marker in these pathologies. The 11 bp deletion in CALR is a frameshift variant affecting the KDEL domain, which is essential for the retention of calreticulin in the endoplasmic reticulum. The variant identified in JAK2 needs further study. Therefore, the detection of the variant in CALR, still awaiting the sequencing result, is remarkable for the assessment of prognosis, the course of the disease and its management.

Cancer Posters - Wednesday

PB1190. Prognostic Value of Bile Acid Transporter SLC10A1 Expression in Hepatocellular Carcinoma

Authors:

H. Chen; West Windsor-Plainsboro High Sch. South, West Windsor, NJ

Abstract Body:

Background/Objectives: Hepatocellular carcinoma (HCC) is the most common primary liver malignancy that occurs predominantly in patients with underlying chronic liver disease and cirrhosis. Emerging data have implicated that the expression and function of hepatic bile acid uptake transporter NTCP (sodium-taurocholate co-transporting polypeptide) are altered in hepatocellular carcinoma. NTCP, encoded by the SLC10A1 gene (solute carrier family 10 member 1), is a key transport protein involved in the enterohepatic recirculation of bile acids. We aimed to systematically analyze SLC10A1 expression and its prognostic role in HCC using various open databases. Methods: SLC10A1 expression in HCC was assessed using UALCAN and GEPIA database. The promoter *methylation* levels were also examined by *UALCAN*. Correlation between SLC10A1 expression and patient survival was evaluated with OncoLnc. SLC10A1 genetic alterations in HCC were explored using cBioPortal. Results: SLC10A1 expression is significantly down-regulated in the clinic-pathological characteristics (cancer stages, patient's race, gender, age, weight, tumor grade, nodal metastasis status, and histological subtypes) examined in HCC patients compared to normal counterparts. SLC10A1 promoter methylation level in HCC was higher than that in normal liver tissue. Clinically, low expression of SLC10A1 was correlated with shorter overall survival in HCC patients ($P= 0.0038$). Among HCC cases with SLC10A1 genetic alterations, mutations were the most common type of alteration (data from TCGA Liver Hepatocellular Carcinoma, PanCancer Atlas). Conclusions: These observations indicate that SLC10A1 may function as a potential tumor suppressor gene. In HCC patients, SLC10A1 expression levels may also serve as a prognostic predictive marker.

Cancer Posters - Thursday

PB1191*. Proteome studies of breast, prostate, ovarian, and endometrial cancers implicate plasma protein regulation in cancer susceptibility.

Authors:

I. Gregga¹, B. Prins², A. Butterworth², P. Pharoah³, S. Gayther⁴, A. Manichaikul⁵, H. Im⁶, S. Kar², J. Schildkraut⁷, H. Wheeler⁸; ¹Loyola Univ. Chicago, Chicago, IL, ²Univ. of Cambridge, Cambridge, United Kingdom, ³Univ Cambridge, Cambridge, United Kingdom, ⁴Cedars Sinai Med. Ctr., Los Angeles, CA, ⁵Univ Virginia, Charlottesville, VA, ⁶Univ. of Chicago, Chicago, IL, ⁷Emory Univ., Augusta, GA, ⁸Loyola Univ Chicago, Chicago, IL

Abstract Body:

Predicting protein levels from genotypes for proteome-wide association studies (PWAS) may provide insight to the mechanisms underlying cancer susceptibility. Here, we focused our PWAS analysis on breast, endometrial, ovarian, and prostate cancers and their subtypes using GWAS summary statistics from several large consortia comprising 237,483 cases and 317,006 controls. We applied three sets of plasma protein prediction models to the cancer summary statistics, performing PWAS to identify protein-cancer associations. In the Atherosclerosis Risk in Communities (ARIC) study, there were 1,276 protein prediction models trained in 7,213 European American individuals available to use as our discovery set. We used additional protein prediction models from INTERVAL (592 proteins trained in 3,301 European individuals) and TOPMed-MESA (300 proteins trained in a population of 975 European, Chinese, African, and Hispanic individuals) to follow up associations discovered with the ARIC models. We found 93 significant protein-cancer associations at an FDR threshold of 0.05 with the ARIC models. Our top 19 ARIC results are significant at a more stringent $p < 5.6E-06$, accounting for multiple testing, and 9 of these proteins had models available to test in our additional model sets. Seven were significant in INTERVAL and/or TOPMed-MESA with the same direction of effect ($p < 5.5E-07$): MSMB-prostate, PLG-prostate, TNFRSF6B-prostate, GSTM1-breast, ARL3-prostate, SERPINA3-prostate, and LAYN-breast. LAYN is associated with breast, ER-positive breast, and prostate cancer at $FDR < 0.05$, suggesting pleiotropy. To further support our results, we applied Bayesian colocalization analysis to the cancer GWAS and ARIC cis-pQTL summary statistics, focusing on our top findings. Testing both one causal variant and multiple-causal variant model assumptions, we found colocalized SNPs for MSMB-prostate (Posterior Probability, PP=1), PLG-prostate (PP=1), LAYN-breast (PP=0.88), SERPINA3-prostate (PP=0.65), and GSTM1-breast (PP=0.62). SNPs in LAYN, SERPINA3, and GSTM1 did not reach genome-wide significance for cancer in the original GWAS, highlighting the power of PWAS for novel locus discovery. In summary, we have identified several loci that likely act through plasma protein regulation to affect cancer susceptibility. PWAS and colocalization are promising methods to identify the molecular mechanisms associated with common variant risk alleles underlying complex traits.

Cancer Posters - Wednesday

PB1192. Rare germline variants in Deubiquitylating enzymes (DUBs) in young women with breast cancer

Authors:

L. Grimaldo¹, A. Mohar², N. Reynoso-Noverón³, L. Sánchez-Zamorano⁴, L. Gómez-Flores-Ramos⁵; ¹The Natl. Inst. of Publ. Hlth.of Mexico, Mexico City, Mexico, ²Inst. of BioMed. Res., Natl. Autonomous Univ. of Mexico, Mexico City, Mexico, ³Res. division, Natl. Cancer Inst., Mexico, Mexico, ⁴Population Hlth.Res. Ctr., Natl. Inst. of Publ. Hlth., Mexico, Mexico, ⁵CONACYT, Mexico; Population Hlth.Res. Ctr., Natl. Inst. of Publ. Hlth., Mexico City, Mexico

Abstract Body:

Breast cancer in young women (BCYW) represents nearly 20% of cases of BC in Mexico and Latin America. Studies reveal that BCYW exhibit a higher genetic susceptibility and a specific genomic signature compared to older women. The evidence shows that deubiquitylating enzymes (DUBs) participate in tumorigenesis, have oncogenic or tumor suppressor activity, and are proposed as therapeutic targets. Whether they contribute to BCYW is unclear. This study aimed to search for rare germline variants in DUBs and describe their pathogenic potential. We analyzed whole-exome 100X data from a clinical cohort of 115 BCYW (women diagnosed with BC <40 years old) using the cloud-based platform PeCanPie. We identified three novel mutations in DUBs. First, *CYLD* rs773205473 (c.1274T>A; p.Ile425Lys), a missense mutation located in exon 13, the C-terminal USP domain, is responsible for its DUB activity. The *CYLD* rs773205473 is classified as pathogenic according to the Varsome database. *CYLD* regulates cell proliferation and apoptosis pathways, acting as a tumor suppressor gene implicated in the epithelial-mesenchymal transition in BC. The second variant, *USP6* rs149134900 (c.4029T>A), is a stop gain mutation located in exon 29 in its USP domain. The SNV *USP6* rs149134900 is classified as likely-pathogenic/VUS according to Varsome. *USP6* can initiate tumorigenesis by inducing the production of matrix metalloproteinases following NF-kappa-B activation. It is a protooncogene associated with the aneurysmal bone cyst, nodular fasciitis, pancreatic adenocarcinoma and BC development. Finally, *USP7* rs753160107 (c.3203-6C>T) is an intronic variant located in the C-terminus domain and classified as benign in Varsome. *USP7* forms a complex with DAXX that regulates the intrinsic MDM2 ligase activity, consequently regulating p53 through proteasomal degradation. None of the variants we report are in the Clinvar database, and their global population MAF is <0.0001. This is the first time *CYLD* rs773205473, *USP6* rs149134900, and *USP7* rs753160107 are reported in Mexican YWBC, making them potentially relevant for early-onset BC. The three patients carrying DUBs rare variants were diagnosed with BC before the age of 30 with Luminal-B invasive ductal carcinoma. All of them had a family history of cancer.

Cancer Posters - Thursday

PB1193. Rare variant association study in lymphoid cancers

Authors:

S. Ralli^{1,2}, S. J. Jones¹, S. Leach¹, A. Brooks-Wilson^{1,2}; ¹Canada's Michael Smith Genome Sci. Ctr., BC Cancer, Vancouver, BC, Canada, ²Simon Fraser Univ., Burnaby, BC, Canada

Abstract Body:

Genetic factors have been identified in lymphoid cancer using genome-wide association and familial studies, with effect sizes ranging from small for common variants to large for rare variants. However, a substantial fraction of the heritability remains unexplained for lymphoid cancers, leading to the question of 'missing heritability,' which might be attributed to rare variants with small effect sizes. We aim to identify genes and pathways with rare germline variants associated with lymphoid cancers using exome sequencing.

Forty multiple-case lymphoid cancer families were exome sequenced, and one affected individual with the earliest age of onset or rarest type of lymphoid cancer was chosen from each family. The ethnicities for lymphoid cancer cases were determined by PCA using KING, and 38 affected individuals of European ethnicity were selected for the association study. Non-Finnish Europeans from gnomAD exomes (N = 56,885) or ExAC (N=33,370) were the controls. High-quality rare deleterious variants with a read depth ≥ 10 were selected from cases and controls for the rare variant association study to be carried out by TRAPD under a dominant model at gene and pathway levels. TRAPD performs a Fisher's exact test to analyze if there is a significantly higher burden of rare deleterious variants in cases than controls for each gene and pathway.

Analysis of genes with two or more rare variants identified five putatively pathogenic germline variants observed in four genes, *INTU*, *PEX7*, *EHHADH*, and *ASXLI*, associated with lymphoid cancers.

Reactome's innate and adaptive immune systems showed rare variant association at the pathway level. Identifying the pathways and genes related to lymphoid cancers will enable us to understand the biology of these cancers and potentially design new prevention and treatment avenues.

Cancer Posters - Thursday

PB1194. Relative involvement of human papillomavirus constitutional genomic instability: Cellular sequela syndromes.

Authors:

E. McGhee, K. Kemp, J. Vadgama; Charles R. Drew Univ. of Med. and Sci., Los Angeles, CA

Abstract Body:

Persistent infections involving high-risk human papillomavirus (HPV) in perinatal transmission is being established as an etiological agent for constitutional genetics in Mendelian syndromes. Recent data indicates that several polymorphisms of key regulators from DNA damage pathways are associated with Mendelian Syndromes. The long arm of chromosome 11q has received much attention as a high frequency of epigenetic deletions of various sites, observed in HPV-long term infection that led to cervical cancer, which is the fourth most common cancer in women, and the fourth leading cause of cancer death indicating the presence of putative tumor suppressor genes, and its potential association with Jacobsen syndrome. To better understand the importance of HPV and constitutional genomic instability, we investigated genetic and epigenetic alterations of tumor suppressor genes such as the ATM and p53 genes in human keratinocytes transfected with HPV16 E6/E7 oncoproteins (16-MT). In this study we examined epigenetic changes and chromosome imbalances, using target CRISPR, NGS, FISH, and arrayCGH. We used several genetic probes, mapping to terminal regions of 11q where the ATM gene is located, genomic instability of chromosomes 13, 14, 15, 20, and 22 are highly methylated regions seen in HPV infected human cells indicating an association with Robertsonian syndrome. This study show that the ATM gene located on chromosome 11q is deleted in these human cells, relating to chromosome deletions, that are recurrent abnormalities in cervical cancer, implicating the loss of tumor suppressor genes as a significant mechanism that drive cells to become genomic unstable, and a possible association to Jacobsen syndrome. Our findings suggest that a delivery detection system involving perinatal transmission of HPV, such as gene editing- CRISPR/Cas9 may be effective in targeting HPV proteins in human cells and may also provide therapeutic benefits.

Cancer Posters - Wednesday

PB1195. Resolving clone- and haplotype-specific copy number variation and DNA methylation in heterogeneous tumors with nanopore sequencing.

Authors:

S. Aganezov, J. Beaulaurier, E. Harrington, S. Hickey, S. Juul; Oxford Nanopore Technologies, New York, NY

Abstract Body:

Large-scale somatic genomic copy number variations (CNV) accumulate during cancer progression, resulting in a tumor comprised of collections of cells, or clones, with distinct CNV profiles. Untangling intra-tumor heterogeneity and inferring clone- and haplotype-specific CNV profiles is important for cancer research and can help inform treatment.

We present a computational workflow to infer clone- and haplotype-specific cancer CNV profiles by processing long nanopore reads obtained with high-throughput bulk sequencing of a tumor and matching normal sample. The workflow first identifies and phases heterozygous germline single nucleotide polymorphisms (SNPs) in the normal sample. Nanopore reads from the tumor sample are then haplotagged with presence/absence of the phased germline SNPs. Both the overall and the haplotype-specific read counts from the tumor are then tallied over fixed-size bins tiled across the reference genome. Finally, we use the state-of-the-art HATCHet matrix factorization algorithm to process the total- and allele-specific read counts and get integer clone- and allele-specific copy number profiles, as well as the clonal cellular fractions of the tumor sample.

To test the proposed approach, we have nanopore sequenced to ~100x average read-depth coverage a COLO829 tumor and a COLO829BL matching normal cell lines. Previous NGS-based bulk and single-cell analyses of the COLO829 cell-line have revealed it to be highly aneuploid and heterogeneous. We show that the proposed nanopore-based workflow identifies clone- and haplotype-specific cancer CNV profiles in concordance with previously published results, with regions of heterogeneity in full agreement with earlier bulk and single-cell studies. Inferred CNV profiles and their clonal fractions are further supported by observed allele-frequencies of somatic SNPs. We demonstrate the stability of the obtained results across lower tumor sample sequencing coverage levels, with CNVs remaining consistent down to 30x tumor coverage, putting the proposed approach on par with the industry standard NGS-based experiments. Notably, because of the unique ability of long nanopore reads to retain single-molecule methylation signals, we were further able to identify haplotype-specific differentially methylated regions both within the tumor sample, as well as in a tumor vs normal comparison, thus shedding light on acquisition/loss of DNA modifications during tumor growth.

These results demonstrate the ability to resolve clonal and subclonal CNVs in structurally aberrant heterogeneous cancers, while also revealing the previously inaccessible allele-specific tumor methylation.

Cancer Posters - Thursday

PB1196. Revisiting tumor evolution in multifocal hepatocellular carcinoma with different clonal origins

Authors:

X. Tang¹, Y. Shao², H. Wang¹, D. Li³, K. Ding⁴; ¹State Key Lab. of Genetic Engineering and Collaborative Innovation Ctr. for Genetics and Dev., Sch. of Life Sci., Fudan Univ., Shanghai, China, ²Dept. of Cardiothoracic Surgery, The First Affiliated Hosp. of Chongqing Med. Univ., Chongqing, China, ³Hepatobiliary and Pancreatic Cancer Ctr., Chongqing Univ. Cancer Hosp., Chongqing, China, ⁴Med. Genetic Inst. of Henan Province, Henan Provincial People's Hosp., Henan Key Lab. of Genetic Diseases and Functional Genomics, Henan Provincial People's Hosp. of Henan Univ., People's Hosp. of Zhengzhou Univ., Zhengzhou, China

Abstract Body:

Multifocal hepatocellular carcinoma (MF-HCC) accounted for more than half of HCCs with uncharacterized molecular mechanisms. A full delineation of the dynamic changes of tumor heterogeneity and clone/subclone components longitudinally could enhance understanding of the formation and development of MF-HCC. Here, we investigated the clonal evolution of MF-HCCs by inferring the order of occurrence of critical genomic events. Whole-exome was sequenced in 74 tumor samples from spatially distinct regions in 36 resected foci and adjacent non-cancerous tissues in ten MF-HCC and one MF-preneoplastic patient. Inter-focal tumor samples showed a significantly higher heterogeneity than that intra-tumoral samples. Early intrahepatic metastasis was evident when the cell number was about 10^4 - 10^5 size of the primary site. Furthermore, MF-HCCs showed a more aggressive developing history in spending approximately 150-300 days from initiation to the metastatic time, significantly shorter than the previously reported 5-10 years cost. We also illustrated the mutational footprints in the preneoplastic lesions of multicentric occurrence and uncovered dominant preneoplastic clones in multicentric carcinogenesis. The portrayed tumor heterogeneity and distinct evolutionary narratives provide significant implications for the personalized drug development and therapeutic strategies for MF-HCC.

Cancer Posters - Wednesday

PB1197. Rewired phospho-signaling networks by genetic variants in Clear Cell Renal Cell Carcinoma

Authors:

S. Mao, M. Wei; Shanghai Jiao Tong Univ., Shanghai, China

Abstract Body:

Clear cell renal cell carcinoma (RCC), the most common type of renal cancer, has been considered insensitive to traditional therapies such as chemotherapy and radiation. Despite the discoveries of novel agents targeting cellular pathways and advance in immunotherapy, response to these treatments has been noticed to be effective in only a small part of patients with late-stage cancers. Therefore, continuous development of new methods to identify biological candidate targets will provide critical clues for future improved therapeutic strategy design. Phosphorylation has been known to provide signal transduction and impact protein-protein interactions, especially in the field oncogenic pathology. Genetic variations exert influence on the phosphorylation sites of proteins so that the phospho-signaling networks were affected in human body. Here, we hypothesized that human missense single nucleotide variants (SNVs) in the phosphorylation sites that were differentially phosphorylated in RCC could modulate kinase signaling interactions and altered the process of RCC. In this study, we first summarized a list of 794 candidate functional phosphorylation sites based on previous literatures. Then, we analyzed 691,052 single nucleotide substitutions from the exome sequencing data of 314 ccRCC patients from the UK Biobank dataset. The phosphorylation associated SNVs (pSNVs) were defined as harmful SNVs that occurred in sites on our list. Our results demonstrated that there were 1012 pSNVs in 280 RCC-associated phosphorylation sites and the significances of these functional pSNVs were evaluated by comparison to the gnomAD SNV frequency data. Significance scores were calculated by assigning various weights to the genes with multiple pSNVs altered kinase binding motifs or replaced central phospho-residues. Genes with frequent pSNVs and significance scores broadly expressed and enriched in functions of RNA splicing, hypoxia and cell cycle regulation signaling, resulting in rewired proteins. The most significant genes (CDK4, MAPK3 et. al) were also shown to have potential associations with RCC. In addition, we evaluated the prognosis prediction effects of pSNVs and constructed a machine learning model based on the DNA sequencing and clinical data of ccRCC samples in the Cancer Genome Atlas (TCGA) database. This model was tested using UKBiobank data set and an independent data set including 880 Chinese ccRCC patients. In sum, our findings successfully identify multiple candidate cell signaling genetic factors contributing to ccRCC development and provide important clues to future mechanistic and translational studies of RCC.

Cancer Posters - Thursday

PB1198. Risk of cancer in persons with a single Fanconi anemia gene variant: A population-based study of MyCode Cohort

Authors:

B. Altintas¹, J. Kim¹, J. Haley², DiscovEHR Collaboration, D. Carey², D. Stewart¹, L. McReynolds¹; ¹Natl. Cancer Inst., Rockville, MD, ²Geisinger, Danville, PA

Abstract Body:

Fanconi anemia (FA) is a bone marrow failure and cancer predisposition syndrome. The risk of malignancies in heterozygous carriers of a single pathogenic FA gene variant has been a question in the field. Our group previously studied the relatives of patients with FA enrolled in the National Cancer Institute (NCI) cohort. Now, we investigate a population-based cohort using a genome-first approach to quantify the cancer risk in heterozygous individuals, potentially identify unrecognized patients with FA and uncover new phenotypes.

We analyzed exome sequencing data from 175,449 individuals enrolled in the MyCode Cohort of Geisinger. Variants in 22 FA genes were identified and annotated by using ANNOVAR, snpEFF and ClinVar. We determined pathogenicity based on ClinVar classification when available, and InterVar classification when there was no ClinVar entry. Phenotypes were extracted from the linked electronic health record (EHR) system and the tumor registry at Geisinger using ICD-10 codes.

At least one FA single nucleotide variant was identified in 7,132 individuals (4.06%). The most frequently affected genes were *FANCN/PALB2* (0.9%), *FANCM* (0.5%), *FAND1/BRCA2* (0.4%) and *FANCA* (0.3%). The rate of heterozygous carriers of a copy number variant within or encompassing one of the FA genes was 0.34%; most commonly in *FANCA* (0.15%).

We identified 28 potentially unrecognized patients with FA; 16 were female and 12 were male. Current median age of these patients was 67 years (range: 32-82). None of these patients had a diagnosis of FA available in the EHR. There were seven male patients with truncating variants in *FANCB*; one of them had adenocarcinoma of an unknown primary at age 68 years, another patient developed adenocarcinoma of the prostate at age 64 and mucinous adenocarcinoma of the head of the pancreas at age 69 years. These patients had no report of an expected early onset cytopenia or congenital abnormalities. We also identified four patients with two *PALB2* variants who were older than 40 years and had no severe phenotype. Three of these patients had no reported phenotype, one patient had developed adenocarcinoma of the prostate at age 53 years. Medical records will be reviewed for confirmation of phenotypes or lack thereof. The validation of genotypes in patients with biallelic variants and the estimation of cancer risk in heterozygous carriers are ongoing.

Our data highlight the power of the genotype-first approach in identifying undiagnosed patients with FA and uncovering novel phenotypes that will aid to properly manage their disease. The high numbers of carriers in this study will allow us to better estimate the cancer risk and validate findings from the NCI cohort.

Cancer Posters - Wednesday

PB1199. Serum biomarkers are altered in UK Biobank participants with mosaic chromosomal alterations.

Authors:

A. Hubbard, W. Zhou, D. W. Brown, S-H. Lin, S. J. Chanock, M. J. Machiela; Natl. Cancer Inst., Rockville, MD

Abstract Body:

The age-related clonal expansion of cells harboring mosaic chromosomal alterations (mCAs) is indicative of clonal hematopoiesis. mCAs can be detected from raw genotype intensity data of blood-derived DNA and have been associated with disease, specifically elevated risk of infection and hematologic cancers. The biological mechanisms linking mCAs to disease risk are poorly understood. Studies of clinical serum biomarkers in individuals with mCAs could provide new insight into molecular mechanisms linking mCAs to elevated disease risk. We investigated the association of 30 serum biomarkers with autosomal mCAs, mosaic loss of the Y chromosome (mLOY) and mosaic loss of the X chromosome (mLOX) in 415,144 cancer-free participants aged 38 to 73 years from the UK Biobank. Log-transformed serum biomarker measurements were regressed against each type of mCA with stratification by sex and adjustment for potential confounders including age, age-squared, and smoking. In total, 12,103 participants had autosomal mCAs, 36,921 had mLOY and 10,246 had mLOX. Our analysis did not provide evidence for a relationship between autosomal mCAs or mLOX and C-reactive protein, a common marker of inflammation, but did demonstrate a strong inverse relationship with mLOY ($p=2.07 \times 10^{-11}$). There was not a clear association between mCAs and hormones except for testosterone, where we observed an overall inverse relationship with expanded autosomal mCAs driven by females ($p=0.009$). We also noted a positive relationship between testosterone and mLOY ($p=5.44 \times 10^{-55}$). While there was no evidence of an association between autosomal mCAs or mLOX and glucose or HbA1c, both displayed strong inverse associations with mLOY ($p < 1.96 \times 10^{-10}$) supporting prior inverse associations reported between mLOY and type 2 diabetes. Our analysis identified inverse relationships between autosomal mCAs, and several lipid biomarkers ($p < 3.80 \times 10^{-5}$). The observed lipid associations remained statistically significant with larger observed effect sizes when restricted to expanded autosomal mCAs (cell fraction $> 10\%$). Conversely, several lipid biomarkers demonstrated a positive relationship with mLOX ($p < 5.92 \times 10^{-4}$) and mLOY ($p < 1.85 \times 10^{-42}$). Our results suggest common clinical serum biomarkers associated with disease risk are altered in individuals with mCAs, although associations vary by type of mCA. Future studies with serial biomarker measurements are needed to disentangle whether these observations reflect causal relationships.

Cancer Posters - Thursday

PB1200. Sex differences in the Lung Adenocarcinoma transcriptome: Analysis of TCGA LUAD and GTEx lung tissue.

Authors:

C. Printzis¹, L. Pesce^{1,2}, B. Stranger^{1,2}; ¹Ctr. for Genetic Med., Northwestern Univ., Feinberg Sch. of Med., Chicago, IL, ²Dept. of Pharmacology, Northwestern Univ., Feinberg Sch. of Med., Chicago, IL

Abstract Body:

Lung cancer is one of the leading causes of death among all cancer deaths, and its mortality, morbidity, and incidence are associated with biological sex. Despite the prevalence of lung cancer and well-known environmental risk factors, there is still much to be learned about the underlying genetic mechanisms and tumor biology. We have characterized sex differences in the transcriptomes of Lung Adenocarcinoma (LUAD) from male and female-derived tumors of The Cancer Genome Atlas (TCGA) and determined the functional impact of those differences. We compared these patterns to those observed in lung tissue from the Genotype-Tissue Expression (GTEx) project. While performing this analysis, we learned several lessons about how transcriptome heterogeneity affects the ability of methods like Surrogate Variable Analysis (SVA) to estimate unknown bias without removing the biological signal of interest. In addition to observing how hidden factor adjustments affect differentially-expressed (DE) gene discovery, we performed simulations to evaluate our ability to avoid false positives while maximizing true discoveries when the effect, batch, and noise are defined. If an effect is heterogeneous, indicating that only a portion of the samples display the effect, we observed that all SVD-related corrections remove all effects. Given the inherent heterogeneity of cancer, these methods produce a large fraction of false negatives. We observed that SVA consistently reduced true positive discoveries to a few percent and often to zero except for the largest uniform differences. We analyzed 516 transcriptomes from TCGA LUAD (N = 238 males; N = 278 females) and 578 from GTEx lung, (N = 395 males; N = 183 females). We identified 798 sex-DE genes in LUAD without use of SVA, and 50 with SVA. In GTEx lung, we identified 213 sex-DE genes without SVA, and 500 with SVA. The 50 DE genes from LUAD with SVA tend to be homogeneous sex-DE genes and are mostly chromosome X genes, e.g., XIST. The sex-biased regulation appeared to be different between healthy lung tissues and lung cancer as indicated by the low replication rate across the two sets. Gene set enrichment analysis of sex-DE genes revealed different enrichments between TCGA and GTEx, as well as between male- and female-derived tumors.

Cancer Posters - Wednesday

PB1201. Shared predisposition to cancer and metabolic disorders using large-scale genomic data

Authors:

V. Pascat¹, L. Zudina², A. Ulrich², A. Demirkan², Z. Balkhiarova², J. Maina³, P. Igor², P. Froguel⁴, M. Kaakinen², I. Prokopenko²; ¹Université de Lille, Lille, France, ²Univ. of Surrey, Guildford, United Kingdom, ³Univ. of Lille, NYERI, Kenya, ⁴UMR 8199, Lille, France

Abstract Body:

Common **metabolic conditions**, such as **type 2 diabetes (T2D)** and **hypertension**, are related to a higher risk of comorbidities and notably to several **cancers**, such as rare **pancreatic** and common **postmenopausal breast, prostate, and colorectal**. These diseases' pathophysiology and comorbidity are related to shared biological processes leading to hyperinsulinemia, hyperglycemia, hyperlipidemia, inflammation. However, the **shared genetic effects** between these diseases **have not been studied systematically**. We dissected the genetic variability shared between T2D, hypertension, and the four aforementioned cancers through standard **polygenic scores (PGS)** and then **clustering of respective disease outcomes and endophenotypes associated DNA variants** to define implicated biological pathways. We finally assessed the causality of these associations by doing multivariate **Mendelian Randomization (MR)**. We constructed PGSs using 2,254 variants, effects of which were evaluated on 39 endophenotypes by hierarchical clustering, giving us five biological process groups. Variants were given weights using Z-scores from published Genome-Wide Association Studies (GWAS). We constructed PGSs and tested their effects in 487,410 individuals from the UK BioBank, including 18,676/8,201/1,425/11,825/40,619 people diagnosed with postmenopausal breast/colorectal/pancreatic/prostate cancer/T2D respectively. We applied two-sample MR to investigate the role of T2D, hypertension, and the different pathways from cluster analysis in cancer risk. We found that **breast, colorectal and pancreatic genetic factors** are directly significantly associated with the **risk of T2D** while **prostate cancer** genetic variants are **inversely associated with T2D** (Pval < 0.05/4). Reciprocally, **PGS based on T2D variants** is significantly associated with a **decreased risk of prostate cancer** (Pval < 0.05/7). Upon clustering, PGSs highlighted the role of **higher adiposity in susceptibility to pancreatic cancer**. No relevant association was found after multiple correction between hypertension and cancers. MR showed a **direct causal relationship of colorectal cancer with T2D's metabolic syndrome and insulin resistance cluster** (OR_{mr} = 1.16, SE = 0.11, Pval = 3.01e-3), and highlighted a **protective causal relationship between adiposity and breast cancer**. This association was driven by **BMI, mediated by a later age at menarche** (OR_{mr} = 0.75, SE = 0.10, Pval = 8.00e-6). This study provides evidence for the complex shared genetic predisposition between metabolic disorders and cancers, with shared biological pathways highlighting the role of adiposity.

Cancer Posters - Thursday

PB1202. Single cell RNA sequencing reveals global cellular landscape of appendiceal cancer

Authors:

L. Bui¹, A. Gutierrez¹, X. Cao², J. wang³, F. Meng³, M. Feng³, L. Arvanitis³, R. Mannan³, M. Raof³, N. Banovich¹; ¹Translational Genomics Res. Inst., Phoenix, AZ, ²Beckman Res. Inst., City of Hope, Duarte, CA, ³City of Hope, Duarte, CA

Abstract Body:

Appendiceal cancer is a rare cancer with approximately 1-2 cases per 1 million individuals. Appendiceal cancer develops from the cells of the appendix and contains two types: epithelial and neuroendocrine appendiceal cancer. Appendix epithelial malignancy starts from the lining cells of the appendix that can lead to the accumulation of mucins, which may result in appendix rupture and spread inside the abdominal cavity. Epithelial appendiceal cancer includes invasive adenocarcinomas (low-grade or high-grade mucinous vs. non-mucinous), and appendiceal mucinous neoplasm (low-grade or high-grade). Classification of appendiceal cancer grades can only be done on tissue specimen by a pathologist and diagnosing appendiceal cancer at early stages is challenging due to the lack of a reliable blood or urine test. To comprehensively characterize the cell types, transcriptional profile, and molecular mechanisms driving cellular remodeling in epithelial appendiceal cancer, we performed a single-cell RNA sequencing (scRNA-seq) study on 12 appendix samples (nine cancerous samples from four different classification grades and three healthy controls). Using unsupervised clustering and cell type specific markers, we identified 24 distinct cell types from more than 85,000 cells and studied the changes in transcriptomic profiles of these cell types between cancerous and healthy samples. The cancerous epithelial cells from all different classification grades exhibit tumor characteristics with high expression of the tumor marker CEACAM6. Consistent with the histology classification of the appendix biopsy, we identified a highly expressed MUC5B+/MUC5AC+ cell cluster in the mucinous goblet-cell adenocarcinoma samples. Interestingly, a MUC5B+ cell type is present in low-grade mucinous adenocarcinoma, but not in the low-grade mucinous neoplasm samples, suggesting differences in cellular components of these two malignancy types at early stage. Cancer-associated-fibroblasts (CAFs), the hallmarks of cancerous mesenchymal phenotypes, are found in all cancerous samples, but not in the healthy controls. Additionally, we observed high expression levels of cytotoxic and exhaustion markers in the cancerous CD8+ T cells and NK cells. Taken together, our study provides an insight into the genetic control of appendiceal cancer at the single cell level and valuable resources for the development of biomarkers for appendiceal cancer classification.

Cancer Posters - Wednesday

PB1203. Single cell transcriptomics of Pituitary Neuroendocrine Tumors (PitNETs)

Authors:

M. Brunner, F. Santoni, J. Meylan-Merlini, M. Messerer, R. Daniel, M. Hegi, M. Muriset; CHUV, Lausanne, Switzerland

Abstract Body:

The pituitary gland, a master controller of hormones production and secretion, is a main component of the endocrine system. Unlike neoplastic formation in other organs, Pituitary Neuroendocrine Tumors (PitNETs) are almost exclusively benign adenomas. PitNETs have been extensively characterized by DNA sequencing, bulk RNA and DNA methylation in order to link those features with tumor invasiveness, prognosis and possible relapses. Here, we present the single cell transcriptome analysis on 7 independent tumors for a total of 37440 single cells: a bi-hormonal (GH-PRL) tumor, two corticotropic (POMC) macro-adenomas and 4 non-secreting adenomas.

Characterization of all tumors showed heterogeneous cell populations; on top of tumorigenic cells, structural cells (endothelial and fibroblasts) and immune cells can be detected. In the two most invasive tumors we identified an unexpected small population of proliferative cells (MKI67+, TOPA2+, BIRC5+, PBK+). Intriguingly, BIRC5 and PBK, two genes already known to be implicated in different type of carcinomas, were highly expressed in this cluster, suggesting a possible link between PitNETs and other types of tumors of various origin. In two non-secreting adenomas we identified a group of cells expressing autophagy related genes, supporting the recent hypothesis that autophagy is an important mechanism for tumor homeostasis. This cluster overlaps with a cluster of specialized cells expressing ribosomal genes and these cells are depleted of mitochondrial proteins. The anticorrelation between the expression of ribosomal and mitochondrial related mRNA is observed in all of our tumors suggesting a recurrent structural configuration to optimize energy balance and transcriptional activity. Our results give new perspectives on the comprehension of the structural composition and the dynamic progression of pituitary tumors.

Cancer Posters - Thursday

PB1204. Single Sample Gene Set Enrichment Analysis (ssGSEA) to Identify Dysregulated Pathways in a Biomarker Matched Pilot Study of Mantle Cell Lymphoma.

Authors:

H. Hill, Y. Liu, P. Jain, K. Chen, M. Wang; Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract Body:

Mantle cell lymphoma (MCL) is a rare, incurable type of non Hodgkin lymphoma (NHL) that is disposed to to repeated relapses, therapeutic resistance and disease progression. Identification of novel therapies and combinations of drugs is essential in overcoming the cycle of relapse and resistance.

The MCL MATCH trial (NCT04872413) aims to identify therapeutic agents which are personalized to a patient's somatic aberrations. The trial utilizes targeted gene expression profiling (GEP) and in-vitro cell viability drug screening to identify potentially efficacious treatments in patients who have developed resistance to Bruton's tyrosine kinase inhibitors (BTKi).

GEP of single patient tumor samples is a unique challenge; most studies of gene expression use differential expression between multiple samples of different phenotypes. Here we utilized single sample GSEA (ssGSEA) on our first enrolled patient sample with a targeted gene panel (Nanostring Technologies, Seattle, WA) that interrogated pathways relating to the biological hallmarks of cancer. ssGSEA relies on gene count list ranking within a sample and leverages a permutation test and statistical correction for multiple testing to evaluate significance. The normalized enrichment score (NES) represents the degree to which the genes in a sample are coordinately up or down regulated within a sample.

We found that two pathways were significantly enriched in the sample: Cell Cycle (NES = 2.68, False Discovery Rate (FDR) q-value = 0.00) and DNA damage repair (NES = 2.20, FDR q-value = 0.00). Combined with in-vitro drug screening and evidence from published literature, ssGSEA could serve as a clinically viable technique to match personalized treatments in a precision medicine trial.

Cancer Posters - Wednesday

PB1205. Single-cell identity and state coupled with genome-wide SNV and CNV in primary breast cancer with ResolveOME profiling.

Authors:

J. Zawistowski¹, I. Salas-Gonzalez¹, T. Morozova¹, T. Tate¹, K. Kennedy¹, D. Arvapalli¹, S. Velivela¹, J. Blackinton¹, J. Marks², E. S. Hwang², G. Harton¹, V. Weigman¹, J. West¹; ¹BioSkryb Genomics, Inc., Durham, NC, ²Duke Univ. Med. Ctr., Durham, NC

Abstract Body:

The molecular events governing the transition from ductal carcinoma in situ (DCIS) to invasive breast cancer are still being elucidated, whereby precise definition of these events has the potential provide a therapeutic window of intervention. To simultaneously expose both genomic and transcriptomic underpinnings in primary breast cancer samples and to ascertain intratumoral heterogeneity we utilized ResolveOME to profile single cells from tumor biopsies of DCIS/invasive ductal carcinoma (IDC). While earlier single-cell methods have importantly unified assessment of copy number variation (CNV) and transcriptomics, they do not yield complete and uniform genome-wide coverage for single nucleotide-level data, made possible with ResolveOME's employment of primary template-directed amplification. As input into ResolveOME, we stratified singulated mastectomy samples by epithelial cell adhesion molecule (EpCAM) surface marker expression with FACS. The genomic arm of ResolveOME followed by analysis with BaseJumper software cataloged genome-wide single nucleotide variation (SNV) in 24 single cells expressing either high or low levels of EpCAM, including the identification of oncogenic *PIK3CA* N345K, while identifying cooccurring DCIS/IDC prototypical chromosomal loss of 11q, 13q and 16q/17p harboring tumor suppressor loci. Concurrently, the transcriptomic arm of ResolveOME enabled the calling of cell identity with the Human Cell Atlas transcriptional database, revealing monocytic and fibroblastic infiltration in the biopsy samples in addition to the expected cells of epithelial identity. The coupling of SNV data and transcriptome data critically unveiled phenotypic cell state heterogeneity, whereby an epithelial cell with both *PIK3CA* N345K and associated chromosomal losses manifested with a stemness/fibroblastic gene expression signature characteristic of the EpCAM low clade of patient cells. We are currently expanding ResolveOME profiling to additional DCIS/IDC patient tumors to comprehensively define driver and regulatory SNV while simultaneously distinguishing between infiltration of non-epithelial cell types from instances of epithelial morphing of physiological cell state within a biopsy. Furthermore, we are defining cell lineage at both the CNV and SNV level as additional single cells are sequenced for each patient sample. These data collectively highlight the molecular complexity and heterogeneity even among a small number of biopsied cells, and underscore the criticality of the interplay of DNA/RNA tiers of information when positing oncogenic mechanisms.

Cancer Posters - Thursday

PB1206. Single-cell methylation and 3C sequencing in prostate tumors shows substantial epigenome reprogramming

Authors:

T. Li, Y. Zhang, K. Abuhanna, R-R. Huang, H. Ye, P. Boutros, C. Luo; Univ. of California, Los Angeles, Los Angeles, CA

Abstract Body:

Prostate cancer is associated with genomic rearrangements and genome-wide epigenomic reprogramming. In addition, prostate cancer tumors have substantial clonal heterogeneity that contributes to relapse following primary treatment. Recent single-cell RNA sequencing studies have revealed diverse transcriptomic programs in the tumor microenvironment associated with prostate cancer progression. However, prostate tumor cell-type heterogeneity at the epigenomic level remains poorly understood. We therefore investigated regulatory abnormalities in prostate cancer by simultaneously profiling methylation and chromatin conformation (3C) in primary prostate tumor samples. We demonstrate that methylation and 3C show concordance in resolving distinct tumor and non-tumor cell populations. Differential methylation analysis between cell types revealed heterogeneity in genome-wide methylation patterns. Finally, we used 3C maps to identify structural variants in tumor subpopulations, and identified novel chromatin loops induced by these variants.

Cancer Posters - Wednesday

PB1207. Single-cell multi-omic characterization of MPAL phenotypes reveals patient heterogeneity associated with poor treatment outcomes

Authors:

C. Hayford¹, V. Kennedy², C. D'amato¹, Y. Xue¹, A. George¹, I. Clark³, C. Peretz², E. Tran², K. Fontanez¹, A. Abate², C. Smith²; ¹Fluent BioSci., Watertown, MA, ²UCSF, San Francisco, CA, ³Berkeley, Berkeley, CA

Abstract Body:

Mixed Phenotype Acute Leukemia (MPAL) is a heterogeneous hematologic malignancy associated with poor patient prognosis, characterized by the presence of both myeloid and lymphoid phenotypes. While clinical data suggests that MPAL patients respond better to lymphoid therapy, treatment is not standardized and responses are often variable. Single-cell characterization of MPAL phenotypes provides a method to understand nuances of patient heterogeneity that may contribute to variable treatment responses. Here, we use the Fluent BioSciences' PIPseq T20 3' Single Cell RNA Kit in combination with a curated panel of BioLegend TotalSeq-A antibodies to characterize a single MPAL patient using both gene and cell surface marker expression. We observe significant intra-patient variability at both the RNA and protein levels, with both healthy immune cell types and malignant blasts separating into distinct clusters. Protein expression overlays on the gene expression space reveal MPAL blasts are HLA-DR and CD45, and healthy immune cells display canonical markers. Differential gene expression analysis identifies multiple Zinc finger genes highly expressed in MPAL blasts, while standard cell surface markers complement differentially expressed genes for the healthy immune cells. Interestingly, surface protein expression identifies further heterogeneity in the MPAL blast population, with a subset of cells exhibiting elevated CD30 expression. Identification of heterogeneity within a set of patient blasts provides a mechanism for discovery of potential resistant cell populations that arise after therapy with a single agent. Together, these results provide insight into the potential of single-cell multi-omic characterization to aid in the discovery of new phenotypes, and serve as a complement to traditional genomic and histological approaches prevalent in the field.

Cancer Posters - Thursday

PB1208. Single-cell sequencing of genomic DNA: Assessing clonal populations of tumor cells in endometrial cancer

Authors:

E. Velazquez Villarreal¹, L. D. Gibbs¹, D. Mahinbakht¹, D. M. Da Silva¹, L. D. Roman¹, D. W. Craig², J. D. Carpten¹; ¹Univ. of Southern California, Los Angeles, CA, ²USC, Keck Sch. of Med., Los Angeles, CA

Abstract Body:

Introduction: Tumor clonal populations have been studied using high-resolution sequencing technologies in some tumor types but have been less explored in Endometrial Cancer (EC). This lack of knowledge exists particularly among underrepresented minority patients, where the disease burden is high. This study sought to assess the clonal populations of four ECs (two African Americans, one Hispanic/Latino, and one Caucasian). **Methods:** EC samples represented different histologies. We used single-cell whole genome sequencing of genomic DNA and performed bulk whole-exome sequencing of the same four tumor samples. The 10x ChromiumTM single-cell copy number variation (scCNV) assay was used with frozen tumor sections to collect and sequence single nuclei. Hierarchical clustering and visual inspection of the generated heatmaps were applied to each tumor sample. **Results:** scCNV data was generated for approximately 300 cells per tumor. We assessed on average five different tumor clonal populations based upon scCNV per sample. We were also able to identify somatic mutations from whole exome sequencing data and identified clones that also harbored these mutations. We were able to identify the presence of oncogenic mutations including PTEN, PIK3CA, and TP53 in heterogeneous clonal populations.

Conclusions: Overall, our study reveals the impact of analyzing tumors using high-resolution single-cell sequencing to assess aberrations in different clonal populations within each tumor sample, which could help to make more precise treatment decisions in cancers that represent significant health disparities.

Cancer Posters - Wednesday

PB1209. *SMARCA4* Schwannomatosis: a known cancer susceptibility gene, a new phenotype.

Authors:

F. Chan-Pak-Choon^{1,2}, C. Nogué³, A-S. Chong³, C. Roca³, R. Austin⁴, T. Robertson⁵, K. van Spaendonck-Zwarts⁴, H. Mar Fan⁴, B. Rivera^{3,6}, W. Foulkes^{1,2,6}; ¹Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, ²Lady Davis Inst. and Segal Cancer Ctr., Jewish Gen. Hosp., Montreal, QC, Canada, ³Molecular Mechanisms and Experimental Therapy in Oncology Program, Bellvitge BioMed. Res. Inst. (IDIBELL), L'Hosp. et de Llobregat, Barcelona, Spain, ⁴Genetic Hlth. Queensland, Royal Brisbane and Women's Hosp., Brisbane, Australia, ⁵Pathology Queensland, Faculty of Med., Royal Brisbane and Women's Hosp., Brisbane, Australia, ⁶Gerald Bronfman Dept. of Oncology, McGill Univ., Montreal, QC, Canada

Abstract Body:

Schwannomatosis (SWNT) is a disorder characterized by a predisposition to multiple benign spinal, peripheral and intracranial nerve sheath tumors (schwannomas (SWNs)). Of the 13-25% of SWNT cases that are familial, 48% are caused by germline *SMARCB1* mutations, and 38% by *LZTR1* mutations. The established molecular mechanism of SWNT consists of a germline *SMARCB1* or *LZTR1* mutation, loss of heterozygosity (LOH) of chromosome 22q (loss of wildtype *SMARCB1* or *LZTR1* and *NF2*), and finally a somatic *NF2* mutation which collectively lead to the inactivation of *SMARCB1* or *LZTR1* and *NF2*. Recently, a variant in *DGCR8*, another gene on chromosome 22q, was reported in one kindred and an unrelated individual with SWNs and multinodular goiters.

We report a family, where the proband presented with a spinal SWN at age 30y whilst her mother (deceased) had had 4 peripheral SWNs (right and left arms) at age 50y followed by a glioblastoma (GBM) at age 54y. All the tumors, except for the GBM, showed loss of BRG1 (*SMARCA4*) in 70-80% of cells and loss of INI1 (*SMARCB1*) in the complementary 10-20% of cells. The GBM showed retention of BRG1 and INI1. Whole exome sequencing of the proband's germline (GM) revealed a likely pathogenic (LP) variant, *SMARCA4* (NM_001128844.2):c.1752_1755del, p.(Lys585Argfs*27) (ClinVar ID: 873514). The GM variant was confirmed in all SWNs and the GBM. Furthermore, we identified LOH at the *SMARCA4* locus along with 12-23 Mb of chromosome 19p in all SWNs (acting as the second hits), but not in the GBM. Methylation analyses are ongoing. No LOH on chromosome 22 or alterations within the 22q SWN genes were detected in the germline nor the tumors, suggesting that the SWNs were caused by a deficient *SMARCA4*, and not by the loss of chromosome 22.

SMARCA4 deficiency has not been reported in SWN patients. GM *SMARCA4* mutations are known to cause SCCOHT, AT/RTs and be associated with uterine sarcomas and infantile pulmonary teratoid tumors. Our findings suggest that there might be other mechanisms by which SWNs are formed - possibly pertaining to *SMARCA4* and *SMARCB1* being key members of the SWI/SNF complexes that regulate essential cellular processes, such as those involved in neuronal differentiation.

Funded by CIHR grant (FDN-148390) to WF and Gershman Scholarship to FCPC.

Cancer Posters - Thursday

PB1210. Splicing analysis of endometrial cancer GWAS risk loci: splicing-associated variants reveal *NF1* and *SKAP1* as candidate susceptibility genes.

Authors:

D. Glubb, T. O'Mara, A. Spurdle, D. Canson; QIMR Berghofer Med. Res. Inst., Brisbane, Australia

Abstract Body:

Alternative splicing contributes to the development of many common traits, including cancer. Indeed, splicing analysis of genome-wide association study (GWAS) variants has revealed both likely causal variants and genes at several cancer GWAS risk loci. However, splicing is not well studied in GWAS even though it may explain the functionality of GWAS variants at ~10% of loci.

To assess GWAS variants for splicing effects, we have developed a prioritization workflow using a combination of SpliceAI and HEXplorer splicing prediction tools, supplemented by splicing quantitative trait locus (sQTL) annotations from the Genotype Tissue Expression Project. Application of this workflow to a set of candidate causal variants from 16 endometrial cancer GWAS risk loci revealed four single nucleotide polymorphisms (SNPs) at two risk loci (17q11.2 and 17q21.32) that were predicted to generate alternative transcripts.

Three SNPs at the 17q11.2 risk locus were predicted to increase inclusion of alternative exons in *NF1* transcripts. The exon inclusion associated with the protective (A) allele of rs7502834 was predicted to generate a truncated NF1 protein with altered C-terminal domain. This prediction was supported by an association of the protective allele with expression of the corresponding alternative *NF1* transcript in subcutaneous adipose. This splicing is expected to reduce expression of the canonical *NF1* transcript, and is consistent with a previous transcriptome-wide association study (TWAS) that found decreased *NF1* expression in subcutaneous adipose associated with decreased endometrial cancer risk. Notably, *NF1* haploinsufficiency is protective for obesity, a well-established risk factor for endometrial cancer.

At the 17q21.32 locus, the risk allele (A) of rs2278868 was predicted to generate a nonsense-mediated decay *SKAP1* transcript through exon skipping and expression of this transcript in whole blood was associated with the risk allele. This finding was also consistent with a TWAS association of decreased *SKAP1* expression in whole blood and increased endometrial cancer risk. As *SKAP1* regulates T-cell receptor signaling and enhances conjugation of T-cells and antigen-presenting cells, decreased *SKAP1* expression may reduce T-cell activity and, consequently, impact endometrial tumor immune surveillance.

In summary, our workflow revealed potentially causal GWAS variants that increase alternative splicing of two genes, with tissue-specific effects that may explain their associations with endometrial cancer risk. This approach could be applied to other datasets to prioritize GWAS variants, and their splicing target genes, for causality.

Cancer Posters - Wednesday

PB1211. t(6;9)(p22;q34) translocation with DEK-NUP214 fusion identified in one patient with chronic myeloid leukemia

Authors:

Y. Zhang¹, D. Jeyakumar², L. Lee², K. Dang³, J. Tso³, L. Yang³, J. Chan⁴, X. Zhao¹, F. Quintero-Rivera⁵; ¹Dept. of Pathology and Lab. Med., UC Irvine Hlth., Orange, CA, ²Div. of Hematology/Oncology, Dept. of Med., UC Irvine Hlth., Orange, CA, ³Dept. of Pathology and Lab. Med., Cytogenetics and Genomics Lab., UC Irvine Hlth., Orange, CA, ⁴Dept. of Pathology and Lab. Med., Univ. of California Irvine, UCI Hlth., Irvine, CA, ⁵Dept.s of Pathology and Lab. Med., and Pediatrics, Div. of Genetic and Genomic Med., Univ. of California Irvine, UCI Hlth., Irvine, CA

Abstract Body:

Chronic myelogenous leukemia (CML) is a clonal hematopoietic disorder characterized by t(9;22)(q34;q11.2) translocation, that results in the generation of a shortened chromosome 22 (Philadelphia-Ph⁺ chromosome), and BCR-ABL1 fusion oncogene which leads to the production of a constitutively activated BCR-ABL1 tyrosine kinase. The unregulated tyrosine kinase activity of BCR-ABL1 fusion protein contributes to the immortality of leukemic cells. Cytogenetics clonal evolution in CML, the occurrence of additional chromosomal abnormalities in addition to the Ph⁺ chromosome, is associated with disease progression and is considered a parameter that defines the accelerated-phase (AP) of CML. Translocation (6;9)(p22;q34) is one of the recurrent cytogenetics findings in acute myeloid leukemia (AML), and myelodysplastic syndrome, and is usually associated with poor prognosis. Here we present, a 50-year-old male who was transferred to our hospital for leukocytosis with blasts on peripheral blood. Patient has insomnia, chest tightness with intermittent dry coughs, fever and chills for the past 3 months. Peripheral blood flow cytometry showed markedly left-shifted granulocytosis with basophilia and circulating blasts (8%). Bone marrow biopsy showed markedly hypercellularity with left shifted myeloid hyperplasia and increased myeloid blasts (9-10%) and mild to moderate reticulin fibrosis, consistent with CML-AP. The presence of BCR/ABL1 (p210) fusion transcript was confirmed by RT-PCR. Chromosome analysis identified a translocation between chromosomes 9 and 22 in all the 20 metaphase cells examined, consistent with a clinical diagnosis of CML. In addition, ten of these cells (a subclone), exhibit secondary anomalies, including a reciprocal translocation between chromosomes 6 and 9 at breakpoints 6p22 and 9q34, resulting in the formation of a chimeric fusion gene between DEK (6p22) and NUP214/CAN (9q34), which was confirmed by concurrent FISH studies. A targeted NGS panel detected both fusions, BCR/ABL1 (VAF % 43.4), and DEK/NUP214 (VAF % 21.2), consistent with the cytogenetics/FISH results. Patient has been treated with dasatinib, cytoreduced with hydroxyurea two weeks prior to his transfer to our institution. Translocation (6;9) is extremely rare in CML patients, only one single CML patient has been reported to carry this t(6;9) translocation previously (PMID: 21156248). How this t(6;9) translocation affects the treatment and prognosis of CML is not clear given its association with a poor prognosis in patients with AML, however, more data is needed to verify this correlation.

Cancer Posters - Thursday

PB1212. Targeting mutant HLA-bound cancer peptide demonstrated in vivo efficacy in delaying tumour growth in a low mutational burden aggressive ovarian cancer

Authors:

T. Nguyen-Dumont¹, G. Ho¹, J. Chang¹, J. Wu¹, H. Barker², P. Eggenhuizen¹, J. Steen¹, J. Bedo², C. Vandenberg², P. Hertzog³, S. Grimmond⁴, T. Papenfuss², C. Scott², M. Southey¹, E. Segelov¹, P. Faridi¹, J. Ooi¹; ¹Monash Univ., Clayton, Australia, ²Walter Eliza Hall Inst., Melbourne, Australia, ³Hudson Inst., Clayton, Australia, ⁴The Univ. of Melbourne, Melbourne, Australia

Abstract Body:

Ovarian carcinosarcoma (OCS) are rare, aggressive cancers with poor prognosis due to limited effective treatments. OCS tumorigenesis is not driven by somatic mutation, but via reprogramming of gene expression. Therefore, the tumour mutation burden (TMB) is often low, and they are relatively non-responsive to single agent immunotherapy.

We have established a paired OCS patient-derived xenograft (PDX) model, SFRC01177 from baseline chemo-naïve and post-chemotherapy specimens in Nod-SCID-Gamma (NSG) major histocompatibility complex (MHC)^{null} mice. These mice are deficient in MHC class I/II expression to reduce acute graft-versus-host disease following injection of human peripheral mononuclear cells. The tumour is refractory to cisplatin and to 1 million human leukocyte antigens (HLA)-match CD8 T-cell injection consistent with expected tumour phenotype.

Whole genome/exome sequencing was performed on the baseline and recurrent tumours. Despite a low TMB of < 4 mutations/Mb, a total of 14,419 variants of mutant peptides ranging from 8 to 16 amino acid length were identified: 13,170 in the baseline and 10,998 in the recurrent tumour with 9,750 shared. Immunoprecipitation of the tumour HLA-complexes with their bound peptides followed by tandem mass spectrometry have identified over 2,300 HLA-bound peptides: 2,215 in the baseline and only 317 in the recurrent tumour with 239 shared. The analysis of the HLA-peptide motifs suggested that down-regulation of tumour cell MHC class I/II may account for this reduction. By combining the genomic and the immunopeptidomic data, a mutant HLA-bound neoepitope was identified. Neuropeptide W E100Q homologous to COSV61585942 (known immunogenic peptide; reported over 5 times in colon cancer) was associated with high areas under the curve in both the baseline and recurrent tumours. *Ex vivo* pulse neuropeptide W E100Q with HLA-matched dendritic and CD8 T-cells demonstrated delayed in OCS tumour *in vivo* growth and increased CD8 T-cells intra-tumour infiltration.

In summary, we can identify potential immunogenic mutant neoepitope using an integrated multi-omics approach in a TMB-low tumour. High affinity T-cell receptor discovery against this peptide is currently underway.

Cancer Posters - Thursday

PB1213. Telomeric nuclear organization of urinary circulating cells in bladder cancer (BC): A new method uncovers unreported markers.

Authors:

O. Doumbia¹, O. Samassekou¹, H. Berthé^{1,2}, M. Goita¹, S. Bamba³, M. Keita¹, I. Sissoko^{1,4}, G. Landoure^{1,5}, M. Traoré¹, A. Kassogue^{1,4}, A. Diarra^{1,6}, C. Traoré^{1,7}; ¹Univ. of Sci. Techniques and Technologies of Bamako (USTTB), Bamako, Mali, ²Urology Dept. of CHU Point G, Bamako, Mali, ³USTTB, Bamako, Mali, ⁴Urology Dept. of CHU Kati, Kati, Mali, ⁵Neurology Dept. of CHU Point G, Bamako, Mali, ⁶Urology Dept. of Hosp. " Mère-Enfant le Luxembourg", Bamako, Mali, ⁷Pathology Dept. Of CHU Point G, Bamako, Mali

Abstract Body:

Background: Bladder cancer is one of the most common cancers and the first cause of death by cancer in urology worldwide and the same epidemiological trend is observed in Mali. The poor prognosis of patients with bladder cancer could be due to the presence of aggressive forms associated with genomic instability which can be determined by the 3D telomeric nuclear organization of cancer cells studies. Biomarkers are lacking to predict therapeutic response.

Objectives: To assess the nuclear organization of telomeres in the urinary circulating cells of patients suffering from bladder cancer.

Methodology: We enrolled patients at diagnosis and recorded their clinical data. We carried out cytological preparation from their urine samples. Then, we did the 3-Dimensional (3D) fluorescence in situ hybridization by using the Cy3 labeled peptide nucleic acid probes specific for telomere. Thereafter, we captured 3D images of the cancer and the normal cells, and assessed the parameters (total number of telomeric signals, average intensity of telomeric signals, Total number of telomeric aggregates, nuclear distribution index, nuclear volume and telomeric distance from the center of the nucleus) defining the telomeric nuclear organization of cancer and normal cells by the 3D telomere analysis.

Results: We found that more than 90% of patients were diagnosed at late stage of the disease and the median survival rate was less than 180 days after the diagnosis, we observed a difference between all of the telomeric parameters of cancer and normal cells. For instance, cancer cells had more telomeres and telomeric aggregates, but had shorter telomeres than normal cells. The nuclei of tumor cells were twice as large as those of normal cells. The telomere nuclear distribution index, which corresponds to the cell proliferation index, was also twice higher in cancer cells compared to normal cells. Also, the telomeres of cancer cells had a more central nuclear location than those of normal cells. Furthermore, we found that cancer cells were more heterogeneous than normal cells. Finally, we found a correlation between the telomeric profile and patient prognosis.

Conclusion: We profiled, for the first time, the telomeric nuclear organization of bladder cancer cells and found that their alteration is associated with patient prognosis. The potentiality of finding a predictive biomarker for this disease might improve the abysmal survival rate of the patients in Mali.

Cancer Posters - Wednesday

PB1215. The effect of weight-loss on the colorectal transcriptome and its relation to colorectal cancer risk.

Authors:

E. Hazelwood^{1,2}, A. Nounu^{1,2}, V. Viallon³, R. J. Beeken⁴, N. Murphy³, J. C. Figueiredo^{5,6}, G. G. Giles^{7,8,9}, M. A. Jenkins⁸, S-S. Kweon^{10,11}, L. Le Marchand¹², L. Li¹³, R. K. Pai¹⁴, E. A. Platz¹⁵, F. J. B. Van Duijnhoven¹⁶, M. J. Gunter³, E. E. Vincent^{1,2,17}; ¹MRC Integrative Epidemiology Unit, Univ. of Bristol, Bristol, United Kingdom, ²Bristol Med. Sch., Univ. of Bristol, Bristol, United Kingdom, ³Nutrition and Metabolism Branch, Intl. Agency for Res. on Cancer, Lyon, France, ⁴Leeds Inst. of Hlth.Sci., Univ. of Leeds, Leeds, United Kingdom, ⁵Dept. of Med., Samuel Oschin Comprehensive Cancer Inst., Cedars-Sinai Med. Ctr., Los Angeles, CA, ⁶Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ⁷Cancer Epidemiology Div., Cancer Council Victoria, Melbourne, Victoria, Australia, ⁸Ctr. for Epidemiology and Biostatistics, Melbourne Sch. of Population and Global Hlth., The Univ. of Melbourne, Melbourne, Victoria, Australia, ⁹Precision Med., Sch. of Clinical Sci. at Monash Hlth., Monash Univ., Clayton, Victoria, Australia, ¹⁰Dept. of Preventive Med., Chonnam Natl. Univ. Med. Sch., Gwangju, Korea, Republic of, ¹¹Jeonnam Regional Cancer Ctr., Chonnam Natl. Univ. Hwasun Hosp., Hwasun, Korea, Republic of, ¹²Univ. of Hawaii Cancer Ctr., Honolulu, HI, ¹³Dept. of Family Med., Univ. of Virginia, Charlottesville, VA, ¹⁴Dept. of Lab. Med. and Pathology, Mayo Clinic Arizona, Scottsdale, AZ, ¹⁵Dept. of Epidemiology, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ¹⁶Div. of Human Nutrition and Hlth., Wageningen Univ. & Res., Wageningen, Netherlands, ¹⁷Sch. of Translational Hlth.Sci., Univ. of Bristol, Bristol, United Kingdom

Abstract Body:

Colorectal cancer is the second most common cause of cancer-related death. Obesity is a modifiable risk factor for this disease, although the mechanisms underpinning this effect remain largely unknown. Recent research has highlighted that weight-loss may lead to changes in tumourigenic biomarkers in normal colorectal tissue. However, further research is needed to characterise how adiposity influences colorectal tissue biology, and to investigate how these changes may in turn impact colorectal cancer susceptibility. As part of an exploratory single arm intervention study, we analysed gene expression microarray data from colorectal biopsies performed in eight individuals before and after an eight-week low-energy diet intervention. We used a permutation approach to identify colorectal tissue-specific gene expression changes resulting from the intervention. In addition, we performed ingenuity pathway analysis (IPA) to determine pathways which were up- or downregulated following weight-loss. We then performed Mendelian randomization (MR) to evaluate the potential causal effect of changes in the identified genes on site-specific colorectal cancer risk (GECCO/CCFR/CORECT consortium comprising 52,775 cases, 45,940 controls). We identified 499 genes which were differentially expressed following the weight-loss intervention (Perm-P < 0.05, fold change > 1.4). In the IPA, several pathways related to cancer development were identified as being differentially regulated following weight loss, including colony formation of colorectal cancer cell lines (5% false discovery rate (FDR)-adjusted P = 5.53 x 10⁻⁵), upper gastrointestinal tract tumour development (FDR-P = 7.09 x 10⁻⁵), senescence of epidermal cells (FDR-P = 3.86 x 10⁻⁴), BEX2 signalling pathway (FDR-P = 2.32 x 10⁻³), role of BRCA1 in DNA damage response (FDR-P = 1.95 x 10⁻²) and CDK5 signalling (FDR-P = 3.63 x 10⁻²). In MR analyses, there was evidence for a potential causal effect of expression of six of the differentially expressed genes on increased colorectal cancer risk (FDR-P < 0.05). Our analysis provides potential insight into the colorectal-tissue specific effects of weight-loss and identifies putative causal mechanisms linking adiposity and colorectal cancer.

Cancer Posters - Wednesday

PB1217. The role of senescent tumor cells in cancer progression

Authors:

T. Park, S. Park, Y. Choi; Ajou Univ. Sch. of Med., Suwon, Korea, Republic of

Abstract Body:

Senescent cells in cancer tissue, including senescent fibroblasts and macrophages, have been reported to increase the malignant potency of cancer cells by secreting senescence-associated secretory phenotype (SASP). Otherwise, Senescence of tumor cells has been believed to inhibit tumor growth by halting the massive proliferation and increasing the chances of immune clearance. In particular, senescent tumor cells (STCs) have been thought that they rarely exist in carcinomas because oncogene-induced senescence needs to be overcome for protumorigenic cells to become malignant. However, recent studies have revealed that a considerable number of STCs are present in cancer tissue, even in metastatic sites. In fact, STCs are widely involved in cancer progression by leading to collective invasion and building a cytokine barrier to protect nonsenescent tumor cells from immune attack. Furthermore, therapy-induced STCs can induce tumor progression and recurrence by increasing stemness. However, obscure causative factors and their heterogeneity in various cancers make it difficult to establish the physiological role of STCs. Here, we present the pathophysiology and role of STCs in cancer progression.

Cancer Posters - Thursday

PB1218. The Role of Tumor Microenvironment and Tumor-Active Pathways in the Progression of Glioma

Authors:

F. Bayram¹, P. Ata², V. Oğlin², M. Ziyal²; ¹Marmara Univ., Istanbul, Turkey, ²Marmara Univ. Sch. of Med., Istanbul, Turkey

Abstract Body:

Many pathways are effective in glioma formed because of differentiation of glial cells. Proliferation related Hippo pathway is one of them. Objective of our study is to express genes of *LATS2* (*Large Tumor Suppressor Kinase 2*), which effects metastasis, *WWTR1* (*WW Domain Containing Transcription Regulator*) that changes proliferation, and downstream *CYR61* (*Cysteine-rich angiogenic inducer 61*), which alters cell adhesion in HGG&LGG tissues and microenvironment via Hippo pathway. Five LGG (age: 47.2±17.9) and eight HGG (age: 60±10.06) patients who were operated in Marmara University Training and Research Hospital, Department of Neurosurgery were included. mRNA isolation and cDNA synthesis were performed. SYBR green with RT-qPCR method was used, expressions of genes were analyzed in BioRad CFX 96 device. *ACTB* (*B Actin*) gene was used as internal control. Fold increase of genes in HGGs compared to LGGs was analyzed with $2^{-\Delta\Delta C_t}$ method. When gene expressions of LGG&HGG samples were analyzed, expression values of *CYR61* were found 3.09 times higher in HGGs. Expressions of *LATS2* and *WWTR1* were not significant. When expressions of GBM tumor and tumor microenvironment were compared, it was found that expression of *LATS2* in tumor decreased, while *CYR61* and *WWTR1* increased. No statistically significant results were obtained. *CYR61* leads glioma by providing epithelial-mesenchymal transformation. This finding, which was found in our study, shows that tumor has capacity to transform into GBM. Gene expression analyzes in tumor microenvironment as well as in the tumor should be an important analysis tool that can provide information about clinical course

Cancer Posters - Wednesday

PB1219. The use of *TNF- α* polymorphic forms as a possible risk assessment tool for the management of breast cancer among Nigerian women.

Authors:

N. Alamukii^{1,2,3}, A. Odetunde⁴, C. P. Babalola^{5,6}, A. G. Falusi⁴, R. I. Nwuba¹; ¹Univ. of Med. Sci., Ondo, Nigeria, ²Dept. of Zoology, Faculty of Sci., Univ. of Ibadan, Nigeria, Ibadan, Nigeria, ³Consortium for Advanced Res. Training in Africa, Nairobi, Kenya, ⁴Genetics and Bioethics Res. Unit, Inst. for Advanced Med. Res. and Training (IAMRAT), Coll. of Med., Univ. of Ibadan, Nigeria, Ibadan, Nigeria, ⁵Dept. of Pharmaceutical Chemistry, Faculty of Pharmacy, Univ. of Ibadan, Nigeria, Ibadan, Nigeria, ⁶Chrisland Univ., Abeokuta, Nigeria

Abstract Body:

Introduction: Early diagnosis is a key factor in management, treatment and survival of breast cancer among Nigerian women. There is therefore a need for more research on suitable risk assessment tool for cancer prevention as well as early diagnosis. Tumour Necrosis factor alpha (*TNF- α*), an important cytokine in development of breast cancer is thus a likely marker for the detection of breast cancer. The single nucleotide polymorphisms (SNPs) of *TNF- α* were genotyped to determine their possible use as risk assessment tool for breast cancer among women. **Methodology:** A case-control study of 200 consenting women recruited from the University College Hospital, Ibadan was carried out. Blood samples were collected from participants for extraction of genomic DNA. Each allele of *TNF- α* SNPs were genotyped through allele specific polymerase chain reaction on extracted DNA of participants. The *TNF- α* (488 G/A, 238 G/A, 308 G/A, 380 G/A, 859 C/T and 1032 C/T) were genotyped. Sequencing of *TNF- α* isolates was done using a Nanopore sequencer. The *TNF- α* level was quantified by ELISA. **Results:** Of 12 alleles genotyped, 4 alleles showed significant association with reduced risk for breast cancer: *TNF- α 488G* (OR-0.24, CI-0.08-0.74), *TNF- α 380G* (OR-0.51, CI-0.51-0.93), *TNF- α 308A* (OR-0.33, CI-0.14-0.78), and *TNF- α 1032C* (OR-2.08, CI-1.18-3.65). The *TNF- α* level was significantly lower in cases compared to controls. *TNF- α 488G*, *308A* and *1032C* were associated with TNF α levels in cases while *TNF- α 488G*, *238A* and *1032T* were associated with *TNF- α* levels in controls. The observed results also showed non-linkage of *TNF- α* SNPs, indicating that each SNP is inherited independent of one another. Sequence analysis confirms the presence of SNPs in the expression of *TNF- α* . **Conclusion:** The *TNF- α 488G*, *308A*, *1032C* variants might be possible predictors for breast cancer among Nigerian women, hence, might be used as possible indicators for risk assessment of breast cancer among Nigerian women.

Cancer Posters - Thursday

PB1220. Transcriptome Wide Association Study identifies novel candidate risk genes for testicular germ cell tumors

Authors:

E. Ugalde Morales¹, R. Wilf², J. Pluta², B. Weathers², M. Damra², P. Kanetsky³, A. Monteiro³, K. Almstrup⁴, K. Nathanson^{5,6}, F. Wiklund¹, for the Testicular Cancer Consortium; ¹Department of Med. Epidemiology and Biostatistics, Karolinska Inst.t, Stockholm, Sweden, ²Div. of Translational Med. and Human Genetics, Dept. of Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ³Dept. of Cancer Epidemiology, Moffitt Cancer Ctr. and Res. Inst., Tampa, FL, ⁴Dept. of Growth and Reproduction, Copenhagen Univ. Hosp. - Rigshosp.et, Copenhagen, Denmark, ⁵Div. of Translational Med. and Human Genetics, Dept. of Med., Perelman Sch. of Med., Philadelphia, PA, ⁶Abramson Cancer Ctr., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

BACKGROUND

Testicular germ cell tumors (TGCT) are highly heritable. GWAS has been successful in identifying 78 independent susceptibility loci for testicular germ cell tumors. However, so far the GWAS data have not been integrated with gene expression data sets to identify causal genes, and novel genes associated with risk.

AIM

To identify candidate genes for which expression is associated with the genetic risk of TGCT.

METHODS

Study design: We conducted a transcriptome-wide association study (TWAS) using the FUSION method, which leverage GWAS summary data and heritable gene expression models trained on independent data sets. **Data:** We used summary statistics from the most recent TGCT GWAS meta-analysis and pre-computed expression prediction models from the Genotype-Tissue Expression (GTEx) version 8 and The Cancer Genome Atlas (TCGA) databases. **Prediction models:** We considered normal testis from GTEx as our primary target tissue (n=322; 12,669 genes), followed by TGCT from TCGA (n=149; 421 genes) and three principal multi-tissue features derived from 22 GTEx tissues, testis included, based on the FUSION sparse canonical correlation analysis (sCCA) summarizing shared regulatory features across tissues (7,128 genes). Models were built based on cis-SNPs from the 1000 genomes LD reference panel located in a 1 Mb window of each gene. Only genes with nominally significant SNP heritability were included in the TWAS analysis.

RESULTS

We tested for 20,218 gene-disease associations and found 168 genes associated at false-discovery rate less than 1%, 98 of which surpassed Bonferroni correction. The 168 genes were mapped to 59 non-overlapping genomic regions. In a joint-conditional analysis performed for each locus, 88 (52%) genes remained as independent hits. GWAS analyses conditioned on the predicted expression showed 48 (55%) genes located at 43 unique loci that could explain, to a large extent, cis-GWAS signals. Of the 48 plausible TGCT causal genes, 27 were novel, i.e., did not harbor any GWAS significant cis-SNP. Using a Bonferroni adjusted gene permutation test to adjust conservatively for enhanced GWAS signals, four novel genes (*BOD1L1*, *CRYL1*, *ARID3B*, *RP4-539M6.14*) and three genes at known TGCT risk loci (*UCK2*, *HEATR3*, *USP36*) were further prioritized.

CONCLUSIONS

We identified candidate TGCT predisposition genes at novel and known loci that are likely to be mediated by gene expression changes, mainly in the testis tissue. Functional analyses are planned to provide additional evidence of key molecular alterations potentially driving the disease onset.

Cancer Posters - Wednesday

PB1221. Transcriptome-wide association study identifies new breast cancer susceptibility genes in Latinas.

Authors:

P. Middha¹, L. Kachuri², A. C. Y. Mak¹, D. Hu¹, S. Huntsman¹, L. H. Kushi³, C. Haiman⁴, E. M. John², G. Torres-Mejia⁵, E. G. Burchard¹, S. L. Neuhausen⁶, L. Fejerman⁷, E. Ziv¹; ¹Univ. of California, San Francisco, San Francisco, CA, ²Stanford Univ., Stanford, CA, ³Kaiser Permanente Northern California, Oakland, CA, ⁴Univ. of Southern California, Los Angeles, CA, ⁵Natl. Inst. of Publ. Hlth., Cuernavaca, Mexico, ⁶City HOPE, Duarte, CO, ⁷Univ. of California, Davis, Davis, CA

Abstract Body:

Background: Genetic susceptibility to breast cancer has been studied extensively in European ancestry populations, but few studies have addressed genetic susceptibility in non-European women. Latinas are a genetically diverse population with contributions from European, African, and Indigenous American ancestries. Genome-wide association studies (GWAS) have identified unique variants in this population, particularly at the 6q25 locus. We conducted a transcriptome-wide association study (TWAS) to identify novel genes associated with risk of breast cancer in Latinas. Methods: We used individual level GWAS data from 2,396 Latina cases and 6,505 Latina controls from studies in Northern California (San Francisco Bay Area Breast Cancer Study, Northern California Breast Cancer Family Registry and Kaiser Permanente Genetic Epidemiology Research on Aging Cohort), Southern California (Multi-ethnic Cohort) and Mexico (CAMA study). We analyzed the association between genetically predicted whole blood (WB) gene expression and breast cancer risk using newly developed TWAS models based on 784 Mexican American individuals. We also conducted parallel analyses using breast mammary tissue (BT) TWAS models from GTEx v8. All analyses were adjusted for age, ancestry, and study. Associations with false discovery rate (FDR) probability <0.05 were considered statistically significant. Results: We tested 9,084 genes from models generated from WB from Latinos and 14,654 genes from models generated from GTEx BT. At FDR<0.05, we identified 10 genes from WB and 9 genes from BT models. Three of the genes were significantly associated with breast cancer risk in both the WB models and BT models. Lower expression of *MIB2* ($p^{\text{FDR}} = 4.6 \times 10^{-13}$ (WB) and 4.93×10^{-11} (BT)) and *SLC35E2B* ($p^{\text{FDR}} = 1.60 \times 10^{-5}$ (WB) and 2.48×10^{-5} (BT)) was associated with decreased risk of breast cancer. Higher expression of *PDGFA* was associated with decreased risk ($p^{\text{FDR}} = 2.86 \times 10^{-25}$) in GTEx BT reference models, but increased risk ($p^{\text{FDR}} = 1.74 \times 10^{-8}$) in the ancestry-specific WB model. Conclusion: This is the first TWAS investigating the relationship between genetically predicted gene expression and breast cancer risk in Latinas. By leveraging gene expression prediction models from whole blood from Latinos and combining these gene expression models that are tissue specific (breast), we have identified ten novel genes associated with breast cancer risk in Latinas. Of these, *MIB2* is a strong candidate for a mechanistic role in breast carcinogenesis. *MIB2* is involved in Notch signaling which plays an important role in breast carcinogenesis via its mismatched receptor-ligand interaction.

Cancer Posters - Thursday

PB1222*. Transcriptome-wide association study identifies novel susceptibility genes for childhood acute lymphoblastic leukemia in Latinos

Authors:

L. Kachuri¹, S. Jeon², A. Mak³, A. DeWan⁴, C. Metayer⁵, X. Ma⁴, E. Ziv³, E. G. Burchard⁶, N. Mancuso⁷, C. W. K. Chiang⁷, J. L. Wiemels⁷, A. J. de Smith⁷; ¹Stanford Univ., Stanford, CA, ²Keck Sch. of Med., USC, Pasadena, CA, ³UCSF, San Francisco, CA, ⁴Yale Sch. of Publ. Hlth., New Haven, CT, ⁵Univ. of California Berkeley, Berkeley, CA, ⁶Univ. of California San Francisco, San Francisco, CA, ⁷Univ. of Southern California, Los Angeles, CA

Abstract Body:

Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy. In the USA and globally children of Latin American origin have the highest risk of ALL, with incidence rates up to 40% higher than in predominantly European ancestry populations. There is accumulating evidence that genetic factors may underlie this striking disparity in risk.

To identify novel ALL susceptibility genes we conducted a transcriptome-wide association study (TWAS) in 1878 cases and 8441 controls from California, self-identified as Latino/Hispanic, using whole blood gene expression prediction models developed in Mexican American adolescents (n=784).

Comparative analyses were conducted using whole blood transcriptome models trained in predominantly European ancestry individuals in GTEx. Statistically significant (FDR<0.05) associations between genetically predicted transcript levels and ALL risk were detected at two novel loci: 19q13.33 (*FCGRT*, OR=1.61, p=2.0x10⁻⁵) and 20p11.23 (*CRNKLI*, OR=0.41, p=2.4x10⁻⁵). We also replicated established ALL risk genes in 7p12.2, 10p12.2, and 14q11.2. *CRNKLI* and *FCGRT* were missed using GTEx TWAS models due to the absence of ancestry-specific expression quantitative trait loci (eQTLs) - variants that are common in admixed populations with Indigenous American ancestry, but rare in other ancestry groups.

Next, we performed TWAS using a meta-analysis that incorporated additional non-Latino White participants (total n=3040 ALL cases, 65,872 controls). Using models trained in Mexican Americans yielded 16 genes with FDR<0.05, five of which were novel (*CRNKLI*, *FCGRT*, *REEP3*, *PGM2*, *CACNBI*). Applying TWAS models trained in a sample that included additional African American and Puerto Rican youths (n=2733) identified 19 genes, including two novel associations (*STOX1*, *INPP4B*). Only 9 known ALL risk genes achieved FDR<0.05 using GTEx, however, effect sizes were directionally concordant across populations for most significantly associated genes. Five novel genes (*FCGRT*, *CRNKLI*, *REEP3*, *PGM2*, *INPP4B*) colocalized with posterior probability (PP)>0.85 indicating strong evidence for shared genetic mechanisms of ALL susceptibility and regulation of gene expression. Given the proposed role of early-life immune dysregulation in ALL etiology, *FCGRT*, which encodes the Fc region of the immunoglobulin G receptor, is a strong mechanistic candidate due its role in transferring humoral immunity across the placental barrier.

In summary, our first TWAS of childhood ALL in Latino/Hispanic participants underscores the value of ancestrally diverse studies for unravelling mechanisms of disease susceptibility across populations.

Cancer Posters - Wednesday

PB1223*. Uncovering gene fusions with 3D genomics: from clinical validation to actionable insights for undiagnosable solid tumors

Authors:

A. Schmitt¹, K. Sikkink¹, K. Galbraith², M. Perez-Arreola¹, M. Movahed-Ezazi², G. Jour², M. Snuderl²; ¹Arima Genomics, San Diego, CA, ²New York Univ., New York, NY

Abstract Body:

Identifying gene fusions in tumor biopsies is critical for understanding disease etiology, however, clinical NGS panels often fail to yield clear genetic drivers. One challenge is that RNA-seq does not perform well in FFPE tissue blocks due to RNA degradation, low transcript abundance, and/or RNA panel design. To overcome this, we developed a novel DNA-based partner-agnostic approach for identifying fusions from FFPE tumors using 3D genomics based on Arima-HiC technology, in some cases with target enrichment (Capture-HiC), and NGS. Using this approach, we have profiled 108 FFPE tumors across tumor types. We first performed clinical validation of the Capture-HiC approach by re-analyzing 29 FFPE tumors comprising 34 actionable gene fusions detected by the RNA-based NYU FUSION SEQer CLIA assay. We observed a 91% concordance (31/34) between Capture-HiC and RNA panels, and are currently re-testing the 3 discordant fusions. We then analyzed 63 FFPE tumors using genome-wide HiC, including 36 CNS tumors, 10 gynecological sarcomas, and 12 solid hematological tumors (lymphoma / plasmacytoma). These tumors had no genetic drivers from prior CLIA-validated DNA and RNA panels. Amongst these, HiC analysis identified previously undetected fusions in 71% (45/63) of tumors. To attribute clinical significance to the fusions detected, we compared the genes implicated in our fusion calls with NCCN and WHO guidelines, and OncoKB, and assigned which tumors had a therapeutic level biomarker (e.g. *PD-L1*, *NTRK*, *RAD51*), or a diagnostic / prognostic biomarker (e.g. *MYBL1* in glioma). We found 39.7% (25/63) of tumors had fusions involving a therapeutic level biomarker and a further 12.7% (8/63) had fusions involving a diagnostic or prognostic biomarker, indicating an overall diagnostic yield of 52.4%. The remaining 19% (12/63) had fusions of potential clinical significance, according to OncoKB. To highlight examples, we identified a novel *PD-L1* rearrangement in a pediatric glioma that was not detected by DNA or RNA panels. Our finding was confirmed by PD-L1 IHC, and the patient was put on pembrolizumab off-label after tumor recurrence and has exhibited a complete response. We also identified *MYBL1* fusions in two glioma cases that were previously missed by RNA panels. In one case, our *MYBL1* fusion spared the patient unnecessary chemo post resection. Together, our findings demonstrate clinical validation, and highlights the utility for 3D genome profiling to increase diagnostic yield by finding clinically actionable fusions as a reflex test and as a frontline test in tumors without available NGS fusion assays (e.g. solid hematological tumors).

Cancer Posters - Thursday

PB1224*. UNISON: a software toolkit for streamlined detection of clonal hematopoiesis of indeterminate potential in large-scale sequencing studies

Authors:

S. Tian, G. Jenkinson, A. Ferrer, H. Yan, S. Baheti, T. Lasho, J. Morales Rosado, M. Patnaik, K. Lazaridis, E. Klee; Mayo Clinic, Rochester, MN

Abstract Body:

Clonal hematopoiesis of indeterminate potential (CHIP) is a premalignant state, in which leukemia associated genes acquire somatic mutations in peripheral blood, at a variant allele frequency (VAF) above 2%; yet the individual does not meet the diagnostic criteria for a hematologic neoplasm. CHIP represents a risk factor for various hematologic malignancies and cardiovascular diseases. The $\geq 2\%$ VAF cutoff was set arbitrarily, which considers the limitation of standard NGS in detecting small clones and the rarity of clinical consequences associated with mutations at lower VAFs. For example, individuals with CH at $\geq 1\%$ VAF also have a significantly increased risk of developing acute myeloid leukemia. Accurate detection of CHIP highly depends on sequencing platform and depth. At $\sim 100x$ coverage, typically for WES, most CHIPs with 2-5% VAFs can't be reliably detected. Deep targeted panel sequencing can reach a lower VAF of 0.5-1%, but it is costly to survey CHIP prevalence in large cohort studies. Currently, no bioinformatics pipeline has been developed for CHIP detection. Rather, a few generic callers have been used, such as GATK HaplotypeCaller, VarScan and MuTect(2). Popular variant calling algorithms often lose power on variants with low VAFs. Further, variant refinement is another key step to enhance CHIP discovery. Empirical filtering strategies are commonly applied, in which the thresholds are not necessarily well justified across all attributes. To systematically address these challenges, we present UNified SOmatic calling of Next-generation sequencing data, or UNISON for short, for streamlined CHIP discovery. UNISON takes raw sequencing reads or analysis-ready bam as input, and integrates variant calling and CHIP prediction in a seamless framework. It implements a meta-caller, which includes Mutect2 and Vardict, plus a hypersensitive in-house caller (VarTracker), and a pre-trained XGBoost classifier for predicting CHIP based on discriminative genomic features and other attributes. Analysis of feature of importance in XGBoost classifier revealed distinct features contributing to CHIP prediction through WES vs. WGS, and for SNVs and INDELS, respectively. UNISON was further assessed on 27 subjects that were sequenced by WES ($\sim 150X$) and by ultra-deep ($\sim 1,000X$) CHIP panel for orthogonal validation. UNISON with WES reaches the lowest VAF of 1.27%, agreeing with ultra-deep CHIP panel in 79.5% of the calls. Finally, applying UNISON to Mayo Biobank WGS data has demonstrated its robustness in discovering the most frequent CHIP genes and in surveying CHIP prevalence, mutation pattern and phenotypic associations at the population scale.

Cancer Posters - Wednesday

PB1225. Unraveling the mechanisms behind the malignant behavior of mast cell tumors using laser capture microdissection mediated RNA sequencing on canine models.

Authors:

A-S. Vander Plaetsen¹, K. Deserranno¹, S. De Vos², W. De Spiegelaere², H. de Rooster², D. Deforce¹, F. Van Nieuwerburgh¹; ¹Ghent Univ. - Lab. of Pharmaceutical Biotechnology, Ghent, Belgium, ²Ghent Univ. - Faculty of Vet. Med., Ghent, Belgium

Abstract Body:

Mast cell tumors (MCTs) are the most frequently diagnosed form of canine skin cancer. They are a diverse class of mast cell malignancies, which are also present in human patients with benign or poor prognosis. Therefore, based on the remarkable resemblances in terms of both biological and clinical features, dogs may serve as an attractive model to study MCTs in humans. Currently, in dogs, MCTs are resected with wide margins, histopathologically graded post-resection, and, depending on post-resection histopathology results, the need for adjuvant therapy is determined. This a posteriori grading requires surgeons to apply large surgical margins, resulting in increased morbidity and risk of complications, even in less aggressive MCTs and delays optimal adjuvant therapy in aggressive MCTs. Therefore, a molecular biomarker that can be used to classify tumors up-front would be of value to personalize patient care. Unfortunately, comprehensive basic and translational knowledge on the molecular characteristics of the MCTs and their correlation with the current canine grading systems and possible metastatic disease is lacking.

The goal of this research is to study the transcriptome across all MCT grades. Furthermore, transcriptome analysis of specific cells in the tumor microenvironment (TME) and in the draining lymph nodes will be performed to elucidate the underlying molecular pathways. Laser capture microdissection sequencing (LCM-seq) is an ideal method to isolate these specific cells from tissue slides based on histological features. Subsequently, cell type specific RNA sequencing can be performed to evaluate the transcriptome. A major advantage of this LCM-seq technology, in contrast to for example flow cytometry-based methods, is the fact that tissue sections can be kept intact and information about the spatial position of each captured cell in the TME is preserved, allowing both archival formalin-fixed paraffin-embedded tissue blocks and cryosections to be evaluated. A cost-effective miniaturized full-length RNA sequencing strategy capable of handling a limited input number of cells will be established based on the recent Smart-seq3xpress protocol. The full-length transcript information will be used to assess differential gene expression and isoform variation in mast cells and cancer-associated fibroblasts (CAFs) in the TME. In addition, further optimization of our RNA sequencing strategy would also allow for long non-coding RNA (lncRNA) biomarker screening.

Cancer Posters - Thursday

PB1226. Unravelling RecQ helicase function in genome stability using Strand-seq

Authors:

Z. Hamadeh¹, V. Hanlon¹, P. Lansdorp²; ¹BC Cancer Res. Ctr., Vancouver, BC, Canada, ²BC Cancer Agency, Vancouver, BC, Canada

Abstract Body:

Helicases are a highly conserved family of motor proteins responsible for interacting with and unwinding canonical and non-canonical DNA and RNA structures. The RecQ class of helicases, known to suppress illegitimate recombination, are implicated in aging and cancer with three of the five human RecQ helicases directly linked to premature aging syndromes characterized by strong cancer predisposition. While no human disease has been associated with the RECQL5 helicase, loss of this gene in cells is known to result in elevated double strand breaks (DSBs) and sister chromatid exchange events (SCEs), a phenotype of genome instability similar to what is observed in RecQ helicase-linked diseases of strong cancer predisposition. Until recently, studying SCEs has been limited to cytogenetic assays that map at megabase resolution. We used single cell template strand sequencing (Strand-seq) to map SCEs as changes in template strand orientation before and after loss of RECQL5 at kilobase resolution. We generated over 20 single and double knockout models for RECQL5 as well as BLM, WRN and RECQL1 helicases using CRISPR-Cas9 in the human haploid cell line, KBM7, and mapped SCEs to the genome using custom bioinformatic approaches to improve resolution and accuracy of SCE detection. We performed enrichment analysis to show SCEs are frequently occurring near actively transcribed genes with guanine quadruplexes and common fragile sites further supporting the role of these helicase genes in suppressing inappropriate recombination at specific genomic elements. We also developed novel bioinformatic approaches to generate genotype-specific call sets for copy number alterations (CNAs), inversions, and translocations. Uncovering the role of DNA helicases in DNA repair and replication pathways is critical for understanding their significance in cancer and aging. Strand-seq offers a unique method to study helicases by mapping the location of SCEs arising in their absence.

Cancer Posters - Wednesday

PB1227. Using human genetics to evaluate the causal role of circulating inflammatory markers in risk of adult cancer.

Authors:

J. Yarmolinsky¹, J. Robinson¹, K. Tsilidis², A. Dehghan², M. Johansson³, D. Mariosa³, M. Gunter³, B. Kiemeny⁴, G. Davey Smith¹, R. Martin¹; ¹Univ. of Bristol, Bristol, United Kingdom, ²Imperial Coll. London, London, United Kingdom, ³Intl. Agency for Res. on Cancer, Lyon, France, ⁴Radboud Univ. Med. Ctr., Nijmegen, Netherlands

Abstract Body:

Tumour-promoting inflammation is a “hallmark” of cancer and laboratory and epidemiological studies have reported links between various inflammatory markers and cancer risk. The causal nature of these relationships and, thus, the suitability of these markers as intervention targets for cancer prevention is unclear. We meta-analysed 6 genome-wide association studies of circulating inflammatory markers comprising 59,969 participants of European ancestry. We then used combined *cis*-Mendelian randomization and colocalisation to evaluate the causal role of 75 circulating inflammatory markers in risk of 32 adult cancers in 347,300 cancer cases and up to 1,015,204 controls. Genetic instruments for inflammatory markers were constructed using genome-wide significant ($P < 5.0 \times 10^{-8}$) *cis*-acting SNPs (i.e. ± 250 KB from the gene encoding the relevant protein) in weak linkage disequilibrium (LD, $r^2 < 0.10$). Effect estimates were generated using inverse-variance weighted random-effects models and standard errors were inflated to account for weak LD between variants with reference to the 1000G Phase 3 European panel. We find strong evidence for effects of genetically-proxied circulating adrenomedullin on breast cancer risk (OR:1.19, 95%CI:1.10-1.29, $P=2.02 \times 10^{-5}$, $H_4=84.3\%$), interleukin-23 receptor on pancreatic cancer risk (OR:1.42, 95%CI:1.20-1.69, $P=6.72 \times 10^{-5}$, $H_4=73.9\%$), antithrombin on triple-negative breast cancer risk (OR:3.62, 95%CI:1.70-7.70, $P=8.30 \times 10^{-4}$, $H_4=72.7\%$), and macrophage migration inhibitory factor on bladder cancer risk (OR:1.14, 95%CI:1.05-1.23, $P=1.43 \times 10^{-3}$, $H_4=76.1\%$), among other findings. Our comprehensive analyses represent the largest human genetics evaluation of circulating inflammatory markers in risk of adult cancers to date. Our findings highlight various novel inflammatory markers implicated in cancer development and suggest pharmacological targeting of these markers as a potential strategy for primary cancer prevention.

Cancer Posters - Thursday

PB1228. Utility of hereditary cancer panel testing in individuals with a known familial variant.

Authors:

E. Weltmer¹, K. O'Brien¹, R. Johnson¹, C. Terhaar¹, Y. Fesko²; ¹Quest Diagnostics, San Juan Capistrano, CA, ²Quest Diagnostics, Secaucus, NJ

Abstract Body:

Accurate reporting of a familial pathogenic or likely pathogenic (P/LP) variant allows an individual to receive targeted testing for the familial variant. However, families and individuals with P/LP variants in 2 different genes have also been reported. The incidence of such findings has not been well-described, particularly in those who already have a known familial P/LP variant. This study aims to describe the percentage of individuals positive for at least one P/LP variant on hereditary cancer panel testing whose clinical history mentioned a P/LP variant in a different gene than was identified through panel testing. This information would add to the understanding of the utility of panel testing for individuals with a known familial P/LP variant. We performed a retrospective analysis of results from hereditary cancer panel tests that were positive for a P/LP variant between June 2016 and March 2022. Our analysis focused on panels representing broad cancer phenotypes such as breast, colon, and multi-cancer phenotypes. Panels targeting a specific syndrome (eg, Lynch syndrome and hereditary breast/ovarian cancer syndrome) were excluded. Cases were reviewed for clinical history of a familial P/LP variant that was either confirmed (genetic test result or genetics family letter specifying gene and variant provided) or suspected (gene provided via verbal report only). Results from 8 hereditary cancer panel tests were included in our analysis: 2 panels of genes associated with breast cancer, 2 associated with colorectal cancer, 1 associated with endocrine cancers, and 3 associated with multiple cancer types. A total of 1,233 of these tests were positive for a P/LP variant, and 151 (12.2%) of these had a familial variant in the clinical history. Among the 151 cases where a familial P/LP variant was mentioned in the clinical history, 8% (12/151) were negative for the familial variant but positive for a P/LP variant in a different gene, 2% (3/151) were double heterozygotes for both the familial variant and a P/LP variant in a different gene, and 90% were positive for only the familial variant. Although targeted testing would have been sufficient for 90% (136/151) of the individuals in our cohort, it would have missed a clinically significant variant in 10% (15/151). These data demonstrate that testing for genes in addition to those specifically noted in an individual's family history may provide opportunities to identify additional at-risk individuals and more appropriately manage those who have additional risks due to a P/LP variant in another gene.

Cancer Posters - Wednesday

PB1229*. Utilizing Electronic Health Records (EHR) and Tumor Panel Sequencing to Demystify Prognosis of Cancer of Unknown Primary (CUP) patients

Authors:

I. Moon^{1,2}, J. LoPiccolo², S. Baca², K. Kehl², A. Gusev²; ¹Massachusetts Inst. of Technology, Cambridge, MA, ²Dana-Farber Cancer Inst., Boston, MA

Abstract Body:

Background:When a standardized diagnostic test fails to locate the primary site of a metastatic cancer, it is diagnosed as a cancer of unknown primary (CUP). This type of cancer represents about 3-5% of all cancers. Due to the hidden nature of primary sites for CUP tumors, oncologists often resort to empiric treatments and patients with CUPs typically have very poor outcomes. Therefore, there is great need for an accurate method to identify the primary site of CUP and empower more informed clinical decision making.

Methods:We used the next generation sequencing (NGS) data collected in routine clinical care as part of the AACR project GENIE (n = 34,567). We trained a gradient boosted decision tree algorithm we call “Boosted Cancer Type” or “BCT” to classify 22 primary cancer types using molecular features abstracted from the NGS data. We validated the model using cross-validation on primary and metastatic cancers with known primary (CKP) as well as germline Polygenic Risk Scores (PRS) among a cohort of 833 patients with CUP. Finally, among 159 patients with CUP who received palliative treatment as their first therapy and had clinical outcomes available, we evaluated whether concordance between BCT predicted sites and first palliative treatment was predictive of improved survival using a multivariable Cox Proportional Hazard regression with the clinical covariates.

Result:BCT achieved high performance on held-out test data consisting of 7,289 primary and metastatic tumor samples from 22 known cancer types (weighted F1 : 0.789). Applying BCT to 838 patients with CUP, we showed that patients with a BCT predicted site have a higher germline risk for the corresponding cancer, providing evidence that CUPs are genetically correlated with known cancers. Furthermore, BCT identified subtypes with significant prognostic differences, which were significantly correlated with relative median survival in their corresponding CKPs (Spearman’s rho 0.810, p-value : 0.015). Importantly, patients with CUP who received first palliative treatments concordant with their BCT predicted sites showed better 5-year survival than those assigned to discordant treatments (H.R. 0.32, 95% C.I. 0.17 - 0.62, p-value : < 0.001).

Conclusion:We demonstrated accurate primary site classification from routinely collected, multi-center NGS panels. BCT stratified patients with CUP into significantly different survival groups, and first palliative treatment concordant with the BCT predictions were predictive of improved survival. BCT thus offers the potential for meaningful clinical decision support for managing patients with CUP.

Cancer Posters - Thursday

PB1230. Variants in *FANCM* confer risk to estrogen receptor-negative breast cancer among Latinas

Authors:

J. Nierenberg¹, A. Adamson², L. Fejerman³, S. B. Gruber², C. A. Haiman⁴, D. Hu⁵, S. Huntsman¹, E. M. John⁶, L. H. Kushi⁷, R. Lee¹, G. Torres-Mejia⁸, C. Patrick⁹, C. Ricker¹⁰, Y. Shieh¹¹, L. Steele², J. N. Weitzel¹², E. Ziv¹, S. L. Neuhausen²; ¹Univ. of California San Francisco, San Francisco, CA, ²Beckman Res. Inst. of City of Hope, Duarte, CA, ³Univ. California Davis, Davis, CA, ⁴Univ. of Southern California, Los Angeles, CA, ⁵Univ California San Francisco, San Francisco, CA, ⁶Stanford Univ. Sch. of Med., Stanford, CA, ⁷Kaiser Permanente Northern California, Oakland, CA, ⁸Inst. Natl. de Salud Pública, Cuernavaca, Mexico, ⁹City of Hope, Duarte, CA, ¹⁰USC Norris Comprehensive Cancer Ctr., Los Angeles, CA, ¹¹Weill Cornell Med. Coll., New York, NY, ¹²Latin American Sch. of Oncology, Tuxla Gutierrez, Mexico

Abstract Body:

Introduction: Genetic testing for pathogenic variants in breast cancer susceptibility genes is currently used to identify women at high risk of breast cancer, who may benefit from intensive screening and preventive interventions, for cascade testing in families, and to inform the use of targeted treatments. However, most evidence supporting this testing is from studies of European ancestry participants. We investigated the impact of known and suspected breast cancer susceptibility genes among over 8,500 Latinas from California and Mexico.

Methods: We conducted a pooled case-control analysis of breast cancer in Latinas from the San Francisco Bay Area, Los Angeles, and Mexico (4,178 cases and 4,344 controls). This included cases from high-risk studies (age below 50 years, family history, or bilateral breast cancer) and from general population studies. Genetic assays included whole exome sequencing on 1,043 cases and 1,188 controls and a targeted panel of 800 genes on the remainder of the samples. We tested the association of loss of function (LoF) variants (frameshift, truncating, or splice site) in each gene with overall, estrogen receptor (ER) positive (+), and ER negative (-) breast cancer, using ancestry-adjusted SKAT-O analyses. We calculated odds ratios for breast cancer with the presence of LoF variants in each gene, using ancestry-adjusted logistic regression models. Our alpha threshold for novel genes was $0.05/20,000=2.5 \cdot 10^{-6}$.

Results: We saw a strong association of LoF variants in *FANCM* with ER- disease ($p=3.7 \cdot 10^{-7}$, OR [CI]: 6.7 [5.8-7.5]) and a nominal association with overall breast cancer risk. Among known susceptibility genes, three were strongly associated with breast cancer, *BRCA1* ($p=5.9 \cdot 10^{-7}$, OR [CI]: 12.3 [11.1-13.4]), *BRCA2* ($p=9.2 \cdot 10^{-11}$, OR [CI]: 7.1 [6.4-7.7]), and *PALB2* ($p=1.2 \cdot 10^{-8}$, OR [CI]: 6.5 [5.8-7.2]). We also found nominally significant associations with other genes, including *CHEK2*, *RAD51D*, and *TP53*.

Discussion: In addition to confirming the role of several accepted breast cancer susceptibility genes in Latinas, we found that LoF variants in *FANCM* were strongly associated with ER- breast cancer risk among Latinas from California and Mexico. *FANCM* is involved in homologous recombination repair like other breast cancer susceptibility genes and has previously been proposed as a possible susceptibility gene for ER- disease, but is not routinely tested in clinical practice. Our results suggest that *FANCM* should be added to breast cancer gene panels, especially among Latinas.

Cancer Posters - Wednesday

PB1231. Whole genome CpG-resolution DNA methylation profiling of HNSCC reveals previously ignored heterogeneity among HPV(+)tumors and clinical implications

Authors:

T. Qin¹, S. Li¹, L. Henry¹, E. Chou¹, L. Rozek², M. Sartor^{1,3}; ¹Dept. of Computational Med. and Bioinformatics, Univ. of Michigan Med. Sch., Ann Arbor, MI, ²Dept. of Environmental Hlth.Sci., Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, ³Dept. of Biostatistics, Univ. of Michigan, Sch. of Publ. Hlth., Ann Arbor, MI

Abstract Body:

Background: DNA methylation is an important epigenetic modification and has been widely studied as a source of biomarkers for cancers. Methylation alterations have been reported in head and neck squamous cell carcinoma (HNSCC), however, most of the findings were based on array data with limited coverage and resolution, and the methylation differences have mainly been explored by HPV status ignoring the high heterogeneity of this disease. *Methods:* In this study we employed whole genome bisulfite sequencing (WGBS) to assay the methylation landscape of a well-studied HNSCC cohort (n=36), already characterized by gene expression, copy number alterations (CNAs), and mutations. We investigated the methylation changes between fine-scaled HNSCC subtypes, including two HPV(+) subtypes. To interpret and validate the findings, we performed integrative pathway analysis to identify the gene sets with both differential methylation and expression, and applied Methylation-based Inference of Regulatory Activity (MIRA) analysis to examine differences in regulatory potential of relevant DNA-binding proteins between the subtypes. *Results:* The previous observation of genome-wide hypermethylation in the HPV(+) patients as compared to HPV(-) was found to be dominantly present in the immune-strong (IMU) subtype of HPV(+) tumors, but not in the keratinized (KRT) subtype (99% vs. 1% of significant regions). The genes formerly reported to be associated with hypermethylation in HPV(+) patients were also found to be more hypermethylated and/or carry more hypermethylated regions among IMU subtype. Notably, the methylation levels of repetitive elements were significantly higher in IMU than in KRT and HPV(-) patients, indicating that the IMU subtype is more genomically stable, in line with the fact that IMU subtype has better prognosis. This was confirmed by the significantly negative correlation between patient-wise methylation values and genomic instability scores based on CNAs. Differential methylation pathway analysis recapitulates most gene expression differences between the IMU and KRT subtypes. Among the KRT and HPV(-) patients, the MIRA analysis identified higher regulatory potential of keratinocyte CTCF-binding that positively correlated with an expression-based keratinization score. *Conclusion:* Hypermethylation in the IMU subtype may explain the higher chance of HPV integration and less favorable prognosis in the KRT subtype. Together, these findings suggest the importance of fine-scale subtype analysis in such a heterogenous disease, and reveal previously unidentified epigenomic alterations and clinical implications in subtypes of HPV(+) HNSCC.

Complex Traits Posters - Thursday

PB1232. A catalog of genetic predictors impacting lifelong medication use patterns in hyperlipidemia, hypertension and type 2 diabetes

Authors:

T. Kiiskinen^{1,2}, P. Helkkula¹, K. Krebs³, J. Karjalainen^{1,2}, E. Saarentaus¹, N. Mars¹, A. Lehisto¹, W. Zhou⁴, M. Cordioli¹, S. Jukarainen¹, J. Rämö⁵, J. Mehtonen¹, K. Veerapen⁶, M. Räsänen⁷, S. Ruotsalainen¹, FinnGen, T. Niiranen^{8,9,10}, T. Tuomi^{1,11,12,13}, V. Salomaa⁸, M. Kurki¹⁴, M. Pirinen¹, A. Palotie^{1,15,16}, M. Daly^{1,15,16}, A. Ganna¹, A. Havulinna¹, L. Milani³, S. Ripatti^{1,2,17}; ¹Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, ²Broad Inst. of the Massachusetts Inst. of Technology and Harvard Univ., Cambridge, MA, ³Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, ⁴Massachusetts Gen. Hosp., Boston, MA, ⁵Broad Inst. of MIT and Harvard, Cambridge, MA, ⁶MGH / Broad Inst., Brookline, MA, ⁷Stanford Univ., Stanford, CA, ⁸Finnish Inst. for Hlth. and Welfare, Helsinki, Finland, ⁹Dept. of Internal Med., Univ. of Turku, Turku, Finland, ¹⁰Div. of Med., Turku Univ. Hosp., Turku, Finland, ¹¹Folkhälsan Res. Ctr., Helsinki, Finland, ¹²Dept. of Endocrinology, Helsinki Univ. Hosp., Helsinki, Finland, ¹³Lund Univ. Diabetes Ctr., Malmö, Sweden, ¹⁴Broad Inst. of MIT and Harvard, Boston, MA, ¹⁵Program in Med. and Population Genetics and Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA, ¹⁶Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, ¹⁷Dept. of Publ. Hlth., Clinicum, Faculty of Med., Univ. of Helsinki, Helsinki, Finland

Abstract Body:

Little is known about the genetic determinants of medication use in the prevention of cardiometabolic diseases. Using the nationwide Finnish drug purchase registry with follow-up since 1995 we performed genome-wide association analyses of longitudinal patterns of medication use in hyperlipidemia, hypertension, and type 2 diabetes (T2D) in up to 193,933 individuals in the FinnGen. In meta analyses of up to 567,671 individuals combining FinnGen together with drug purchasing data in the Estonian Biobank and prescription data in the UK Biobank (UKBB), we discovered 333 independent genetic loci ($p < 5 \times 10^{-9}$): 74 with leading associations for hyperlipidemia, 181 for hypertension, and 78 for T2D medication patterns. Fine-mapping revealed 494 95% credible sets (CS) associated with the total number of medication purchases, longitudinal changes in medication combinations, or discontinuation of medication, including 46 CS in 40 medication use associated loci that have not been previously associated with the underlying treatment targets. Of the 333 independent medication loci, 158 were also associated with coronary artery disease (CAD) with shared effect directions in a novel CAD meta-analysis combining FinnGen, UKBB, and the CARDIoGRAMplusC4D consortium ($n=811,555$). Combining the genome-wide associations on CAD and the medication patterns to a novel multi-trait polygenic risk score (PRS) improved the prediction of CAD beyond a state-of-the-art CAD PRS (21% higher R^2) in an independent sample (UKBB, $n=343,676$). We also show that the PRSs for low-density lipoprotein, systolic blood pressure, and T2D are strongly associated with the medication use behavior. In summary, we demonstrate novel medication pattern-based strategies for identifying cardiometabolic risk loci and provide new genome-wide tools for medication-based prevention of cardiovascular diseases.

Complex Traits Posters - Wednesday

PB1233. A comparative PRS-PheWA of 14 autoimmune disorders, and their genetic relationship.

Authors:

A. Topaloudi, P. Jain, M. Martinez, P. Paschou; Purdue Univ., West Lafayette, IN

Abstract Body:

Autoimmune diseases (ADs) are a group of more than 80 heterogeneous disorders that occur when there is a failure in the self-tolerance mechanisms triggering self-attacking autoantibodies. Most of the autoimmune disorders are polygenic and associated with genes in the human leukocyte antigen (HLA) region. However, additional non-HLA genes are also found to be associated with different ADs, and many times they are implicated in more than one disorder. Previous studies have observed associations between various health-related and lifestyle phenotypes and ADs. A way to calculate an individual's genetic liability to a phenotype is through Polygenic risk scores (PRS), estimated as the sum of the risk alleles weighted by their effect sizes in a genome-wide association study (GWAS). Here, for the first time, we conducted a comparative PRS-PheWAS analysis for 14 different ADs (Celiac Disease, Crohn's Disease, Inflammatory Bowel Disease, Juvenile Idiopathic Arthritis, Multiple Sclerosis, Myasthenia Gravis, Primary Sclerosing Cholangitis, Psoriasis, Rheumatoid Arthritis, Systemic Lupus Erythematosus, Type 1 Diabetes, Ulcerative Colitis, Vitiligo Early Onset, Vitiligo Late Onset) and 3,281 outcomes available in the UK Biobank that cover a wide range of lifestyle, socio-demographic and health-related phenotypes. Additionally, we further explored the genetic relationships of the selected ADs by estimating the genetic correlation and performing cross-disorder GWAS meta-analyses for the identified AD sub-groups. In total, we observed 554 outcomes significantly associated with at least one disorder PRS, and 300 outcomes were significant after variants in the HLA region were excluded from the PRS calculations. From the genetic correlation followed by the genetic factor analysis we observed five genetic factors and the cross-disorder meta-analyses in each factor revealed genome-wide significant pleiotropic loci. Overall, our analyses indicate potential factors associated with genetic risk for ADs, some of which have been reported previously, and also novel observations that need to be further explored.

Complex Traits Posters - Thursday

PB1234. A comprehensive coronary artery disease risk prediction framework incorporating genetic and nongenetic risk factors.

Authors:

S-F. Chen^{1,2}, H. J. Sadaei¹, L. Sang-Eun¹, N. Wineinger^{1,2}, A. Torkamani^{1,2}; ¹Scripps Res. Translational Inst., La Jolla, CA, ²Scripps Res., La Jolla, CA

Abstract Body:

Coronary artery disease (CAD) is the leading cause of mortality and morbidity worldwide. With the polygenic inheritance of CAD, the findings from large-scale genome-wide association studies led to a proliferation of polygenic risk scores (PRSs), which variably demonstrated the potential implications of genetic studies to aid in clinical prevention and treatment development. However, the current utility of current PRS models is confined to the identification of at-risk populations in the top percentiles of genetic risk. Despite the known heterogeneous risk factors mediating CAD risk from a broad spectrum of genetic and environmental risk factors, there have been limited efforts to blend these factors into a comprehensive risk prediction model. A major goal of CAD risk prediction should be to develop schemes integrating genetic risk factors and be able to recommend appropriate medical interventions with higher therapeutic responses to reduce the burden of disease. Herein, we performed deep phenotyping on biobank data to ascertain diseases and curated all accessible nongenetic risk factors from individualized profiling of laboratory tests, diagnosis, interview, and lifestyle from self-reported or electronic health records. All accessible PRSs for expected medical measurements and predisposition of comorbidity were also calculated from whole-genome genotyping data. To improve generalizability and reduce bias from a single contributing model, we aggregated the performance by leveraging the learnability of numerous state-of-art machine-learning (ML) algorithms. Multiple risk prediction models for contributing risk factors were combined into a meta-estimator predicting the 10-year CAD cumulative incidence. To capture gene-by-environment interactions, the ensemble framework was built upon decision-tree-based algorithms capable of discovering the non-linear interplay between predictors. We also interpreted the feature importance with Shapely additive explanations value, which can isolate important contributing risk factors at an individual level and imply preventative interventions. Our preliminary results outperformed contemporary risk calculators not only with higher predictive accuracy but also delivering the risk assessment in an actionable text. With increased adoption and connectivity of personal digital health data sources, this framework provided an opportunity to promote personal or clinical utility. Altogether, the final model will be able to suggest and prioritize conforming consequential decisions to facilitate substantial personalized care and precision medicine.

Complex Traits Posters - Wednesday

PB1235*. A comprehensive genetic map of cytokine responses to pathogens in 1063 Lyme patients reveals novel regulatory mechanisms underlying diseases.

Authors:

J. Botey-Bataller^{1,2}, H. D. Vrijmoeth^{1,3}, J. Ursinus^{4,3}, B-J. Kullberg¹, C. van den Wijngaard³, H. t. Hofstede¹, C-J. Xu^{1,2,5,6}, M. G. Netea^{1,7}, J. W. R. Hovius⁴, L. A. B. Joosten^{1,8}, Y. Li^{1,2,5}; ¹Dept. of Internal Med. and Radboud Ctr. for Infectious Diseases, Radboud Univ. Med. center, Nijmegen, Netherlands, ²Dept. of Computational Biology for Individualised Infection Med., Ctr. for Individualised Infection Med., Helmholtz Ctr. for Infection Res., Hannover Med. Sch., Hannover, Germany, ³Natl. Inst. for Publ. Hlth. and Environment (RIVM), Ctr. for Infectious Disease Control, Bilthoven, Netherlands, ⁴Dept. of Internal Med., Div. of Infectious Diseases & Ctr. for Experimental and Molecular Med., Amsterdam UMC, Univ. of Amsterdam, Amsterdam, Netherlands, ⁵TWINCORE, Ctr. for Experimental and Clinical Infection Res., a joint venture between the Hannover Med. Sch. and the Helmholtz Ctr. for Infection Res., Hannover, Germany, ⁶Dept. of Gastroenterology, Hepatology and Endocrinology, Hannover Med. Sch., Hannover, Germany, ⁷Dept. for Genomics and Immunoregulation, Life and Med. Sci. Inst. (LIMES), Univ. of Bonn, Bonn, Germany, ⁸Dept. of Med. Genetics, Iuliu Hatieganu Univ. of Med. and Pharmacy, Cluj-Napoca, Romania

Abstract Body:

Inter-individual variability of immune responses determines the susceptibility to certain infectious diseases or chronic conditions. To tackle this, we established a cohort of 1063 Lyme Borreliosis patients and measured the *ex vivo* cytokine responses to various stimuli before and after antibiotic treatment. We characterized the effect of the disease course on the responses and found that IL-10 responses were increased after antibiotic treatment and negatively correlated with antibody levels against *Borrelia*. This confirms the detrimental role of IL-10 in mounting an adaptive response against *Borrelia*. We characterized the effect of genetic variation through cytokine quantitative trait loci (cQTL) mapping. 45 independent loci showed a genome-wide significant association ($P < 5 \times 10^{-8}$), of which 44 are novel loci. The identified cQTL showed consistency between before and after treatment in terms of allelic effect size and direction. Among the newly identified cQTL, the *CFH* locus is associated with cytokine responses to *Candida albicans* and *Borrelia*. We captured the cis-regulating cQTL at the *IL10* and *IL6* loci. The *KL* locus, which codes for the anti-aging protein Klotho, was found significant for IL-6 and IL-1 β responses. Of note, the strongest cQTL signal was found to be the *TLRI-6-10* locus, which was reported in our early study in Health. Interestingly, we found that it regulates IL1-Ra production in response to *Borrelia* but not to Pam3Cys, which was validated in Pam3Cys stimulation RNA-seq data of 95 samples in an independent cohort. Adding to this, we assessed the causal variants by fine-mapping. This yielded a single non-synonymous variant, rs5743618, as the only significant in stimulations before treatment and multiple signals as causal for stimulations after treatment. Genes identified from cQTL were significantly enriched (GSEA, $P < 2 \times 10^{-3}$) in response to *Borrelia* stimulation based on RNA-seq data. Furthermore, cQTL were functionally enriched in open chromatin peaks of immune-cell populations based on ENCODE and Roadmap epigenomics data ($P < 1 \times 10^{-4}$). In total, four cQTL were colocalized and/or causally linked to the genetic regulation of immune-mediated diseases, adding new perspectives to their aetiology. In summary, we provide an expanded list of genetic determinants associated with variability in immune responses. In the clinical context, knowing the genetic predisposition towards different responses will help in better patient stratification and more tailored treatments.

Complex Traits Posters - Thursday

PB1236. A *Drosophila* platform to validate and assess pathogenicity of LMNA variants identified from dilated cardiomyopathy patients.

Authors:

Z. Han, H. Liu, Y. Fu, J. van de Leemput; Univ. of Maryland Sch. of Med., Baltimore, MD

Abstract Body:

LMNA is one of the most sequenced genes in humans - its variants associated with at least 18 different diseases. Yet, hundreds of *LMNA* variants remain classified as "variant with conflicting interpretations" or "variant with uncertain significance" in ClinVar, making it one of the top priority genes in need of a functional validation system to unravel its variant-to-clinical significance classifications. Here, we developed a highly efficient *Drosophila* platform to provide *in vivo* functional analysis for *LMNA* variants associated with dilated cardiomyopathy (DCM). We found that cardiac-specific silencing of *LamC*, the fly homolog of human *LMNA*, leads to a typical DCM phenotype (dilated heart tube, reduced myofibrils, reduced contraction, abnormal function), which can be completely rescued by expressing wildtype human *LMNA*, but not by mutant *LMNA* carrying the established pathogenic DCM variants, in these flies, *i.e.* our "gene replacement" approach. Using this platform, we screened 13 *LMNA* variants: 6 pathogenic variants without *in vivo* evidence, 4 variants with conflicting interpretations, and 3 variants of uncertain significance. The fly heart data validated and (re-)classified the variants as 8 pathogenic and 5 benign variants, demonstrating the effectiveness of this *Drosophila* platform in providing the much-needed "variant-level" cardiac functional validation, even for genes with complex interpretations like *LMNA*.

Complex Traits Posters - Wednesday

PB1237. A genome wide association study of chronic spontaneous urticaria risk and heterogeneity.

Authors:

D. Chang, C. Hammer, C. Holweg, S. Selveraj, N. Rathore, M. I. McCarthy, B. L. Yaspan, D. F. Choy; Genentech, South San Francisco, CA

Abstract Body:

Chronic spontaneous urticaria (CSU) is a dermatologic condition characterized by spontaneous, pruritic hives and/or angioedema that persist for six weeks or longer with no identifiable trigger. Mast cell and basophil degranulation are key factors in CSU etiology. One mediator of this degranulation is autoantibody associated FcεRI/IgE crosslinking, though additional mechanisms are likely involved. Existing therapeutic options for CSU patients such as antihistamines and omalizumab (an anti-IgE monoclonal antibody) relieve symptoms in some patients, while others remain symptomatic, suggesting additional mediators can be targeted therapeutically. Investigating the genetics of CSU may provide clues to other CSU-associated pathways, but to date, only candidate gene studies in small cohorts have been carried out with limited replication. The aim of this study was to carry out a genome-wide association study (GWAS) of CSU risk and heterogeneity. In our CSU risk GWAS of 679 CSU patients and 4,446 controls, we identified two significantly associated loci. The strongest association maps to position 56 of HLA-DQα ($P=1.69 \times 10^{-9}$), where the arginine residue was associated with increased CSU risk ($OR=1.64$). The second association signal ($P=1.57 \times 10^{-8}$, $OR_{C-allele}=1.44$) colocalized with expression-quantitative trait loci in whole blood for *ITPKB* ($P=1.90 \times 10^{-8}$, $\beta_{C-allele}=0.095$), a secondary messenger in calcium signaling ($\text{probability}_{\text{colocalization}} > 0.90$). We also investigated the genetics of CSU heterogeneity as measured by the Chronic Urticaria (CU) index - an assay that tests for basophil histamine releasing factors e.g. serum IgG autoantibodies. Comparing 455 CU-index low to 187 CU-index high patients, we found that the CSU risk associated arginine residue at position 56 of HLA-DQα was also associated with high CU-index ($P=6.15 \times 10^{-5}$, $OR=1.86$), while the *ITPKB* association was not associated with CU-index ($P=0.64$). Further analyses revealed three autoimmune polygenic scores (hypothyroidism, type-1 diabetes, and vitiligo) are associated with CSU risk and CU-index ($P < 2.34 \times 10^{-3}$, $OR > 1.72$). In summary, this GWAS identifies a role for calcium signaling in CSU pathology and provides a genetic basis for autoimmune mechanisms in risk and clinical presentation of CSU.

Complex Traits Posters - Thursday

PB1238. A genome-wide association study in Peruvians suggests new risk loci for Alzheimer disease

Authors:

F. Rajabli¹, M. Cornejo-Olivas², N. Custodio³, M. Soto-Añari⁴, R. Montesinos⁵, D. Reyes-Dumeyer⁶, J. Cedenó⁶, J. Rivero¹, P. R. Mena¹, L. D. Adams¹, P. G. Whitehead¹, K. L. Hamilton-Nelson¹, J. Rios-Pinto⁷, A. Medina-Colque⁷, C. Dalgard⁸, R. Isasi¹, I. Cornejo-Herrera⁷, M. Illanes-Manrique², E. Ochoa-Valle⁷, E. Sarapura-Castro⁷, K. Mejía⁹, K. Milla-Neyra², S. Castro-Suarez⁷, E. R. Martin¹, A. Griswold¹, M. L. Cuccaro¹, J. M. Vance¹, G. W. Beecham¹, M. A. Pericak-Vance¹, G. Tosto⁶; ¹John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, ²Neurogenetics Res. Ctr., Inst. Natl. de Ciencias Neurológicas, Lima, Peru, ³Unidad de diagnóstico de deterioro cognitivo y prevención de demencia, Inst. Peruano de Neurociencias, Lima, Peru, ⁴Univ. Católica San Pablo, Arequipa, Peru, ⁵Unidad de diagnóstico de deterioro cognitivo y prevención de demencia, Lima, Peru, ⁶Taub Inst. for Res. on Alzheimer's Disease and the Aging Brain, Coll. of Physicians and Surgeons, Columbia Univ., New York, NY, ⁷Neurogenetics Res. Ctr., Inst. Natl. de Ciencias Neurológicas, Lima, Peru, ⁸Dept. of Anatomy, Physiology & Genetics, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, ⁹Daniel Alcides Carrion Natl. Hosp., Callao, Peru

Abstract Body:

Background Increasing racial/ethnic diversity in genetic studies is critical for defining the genetic architecture of Alzheimer disease (AD). Native American (NA) populations are substantially underrepresented in AD genetic studies. The Peruvian (PE) population, with up to ~80% of NA ancestry, provides a unique opportunity to assess the role of NA ancestry in AD. We performed the first genome-wide association study (GWAS) in the PE population to identify novel AD susceptibility loci and characterize the known AD genetic risk loci in the PE population. **Methods** The PE dataset includes array-genotype and phenotype data from 630 individuals (180 cases; 450 controls), all imputed to the NHLBI TOPMed v5 haplotype reference panel. We performed genome-wide association analyses with a generalized linear mixed-model using the SAIGE software. The model included genotype, sex, age, and principal components (population substructure) as fixed effects and genetic relationship matrix as a random effect. **Results** We identified the *APOE* gene as a genome-wide significant with an effect size comparable to the effect size found in non-Hispanic white (NHW) populations ($OR=3.35$ (2.31-4.87), $P=4.6 \times 10^{-10}$) (GIF=1.04). Variants at four additional loci reached suggestive significance ($P < 10^{-5}$): *PPCS* ($P=4.6 \times 10^{-6}$), *NFASC* ($P=1.6 \times 10^{-7}$) on chromosome 1, *DGKB* ($P=2.6 \times 10^{-7}$) on chromosome 7, and *KDM4C* ($P=1.4 \times 10^{-6}$) on chromosome 9. Follow-up ancestry-aware analysis at the *DGKB* locus showed increased NA ancestry among the risk allele carriers. The high frequency of the risk allele in NA ancestry increased the power to detect association and suggests that the signal at the *DGKB* locus was driven by the NA ancestry. In addition to the *APOE* locus, two more known AD loci, *BINI* ($P=0.03$) and *ECHDC3* ($P=0.04$), showed nominal significance. **Conclusions** Our PE GWAS generalized the association was seen in NHW individuals of *APOE*, *BINI*, and *ECHDC3* with AD risk and identified four novel suggestive susceptibility loci. *DGKB* (Diacilglicerol kinase beta) gene is a promising candidate that plays a key role in cellular processes and has been implicated in AD risk in a Caribbean Hispanic GWAS study.

Complex Traits Posters - Wednesday

PB1239. A genome-wide association study of lifetime estrogen exposure in Korean postmenopausal women

Authors:

M. Yuk, Y. Kwak, S. Yang, J. Lee, J. Youn, N. Song; Coll. of Pharmacy, Chungbuk Natl. Univ., Cheongju, Korea, Republic of

Abstract Body:

BACKGROUND: Estrogen exposure is a principal index for women's health. Abnormal estrogen exposure can induce several age-related diseases such as cardiovascular disease, osteoporosis, endometrial cancer, and ovarian cancer. There are plenty of studies about the age at menarche or menopause but not enough for estimated lifetime estrogen exposure (ELEE) even though epidemiological studies have reported that ELEE is associated with women's health. To identify novel genetic loci for ELEE and to replicate previously identified loci by genome-wide association studies (GWAS), we conducted GWAS in three cohorts of Korean postmenopausal women and meta-analyses.

METHODS: Among study participants of the Korean Genome and Epidemiology Study including the Health EXAMinee (HEXA, N=38,401) study, the CARDioVascular disease Association Study (CAVAS, N=5,046), and the Ansan and Ansong study (N=2,877) with genome-wide single-nucleotide polymorphism (SNP) data, we included women who experienced natural menopause for this study (N=18,535 in HEXA, N=2,827 in CAVAS, and N=1,434 in Ansan and Ansong study). ELEE was calculated by subtracting the age at menarche from the age at menopause. To identify SNPs associated with ELEE, multivariable linear regression model was used adjusted for age and age at menarche and the results of three GWAS were meta-analyzed. We selected 40 menopause, menarche, and sex hormone susceptibility SNPs ($P < 5 \times 10^{-8}$) in East Asian population from the GWAS catalog for replication study.

RESULTS: Among 37 SNPs satisfying nominal genome-wide significance level ($P < 5 \times 10^{-8}$) in HEXA study, 12 SNPs were replicated in CAVAS or Ansan and Ansong study ($P < 0.05$). In meta-analysis, we found 26 novel SNPs ($P < 5 \times 10^{-8}$). The strongest association with ELEE was observed in the SNP rs13438 (*HI-10*, Beta (SE)=-0.4705 (0.0560), $P=4.63 \times 10^{-17}$) and rs60527165 (*HMCES*, Beta (SE)=-0.4896 (0.0586), $P=7.31 \times 10^{-17}$). The most frequent genetic association on ELEE was mapped to two genes, *SLC44A4* (rs660594, rs644774, rs494620, rs2242665, rs2242664, and rs3130482) and *PRRC2A* (rs3763295, rs2857693, rs2736171, rs2242657, and rs1046089). In replication study, 8 out of 40 previous GWAS-identified SNPs ($P < 5 \times 10^{-8}$; rs13438, rs365132, rs353493, rs2269475, rs28797500, rs3814226, rs2277339, and rs28403619) were replicated to be associated with ELEE.

CONCLUSION: We replicated 8 GWAS-identified susceptibility SNPs for ELEE and identified 26 novel SNPs for ELEE in the current GWAS. The new and replicated genetic variants for ELEE could facilitate identification of population with the longer duration of lifetime estrogen exposure for women's health.

Complex Traits Posters - Thursday

PB1240. A genome-wide association study of mammographic density phenotypes among pre-menopausal women of European ancestry.

Authors:

T. Harrison¹, H. Chen¹, C. Turman², C. Scott³, I. Andrulis⁴, K. Aronson⁵, M. Bolla⁶, G. Chenevix-Trench⁷, F. Couch⁸, J. Dennis⁶, D. Easton⁶, D. G. Evans⁹, P. A. Fasching¹⁰, M. Gago^{11,12}, M. Garcia-Closas¹³, C. A. Haiman¹⁴, P. Hall¹⁵, J. L. Hopper¹⁶, R. S. Houlston¹⁷, E. M. John¹⁸, K. Michailidou^{19,6}, R. L. Milne¹⁶, J. Stone²⁰, Q. Wang⁶, C. M. Vachon²¹, P. Kraft^{2,22}, S. Lindstrom^{1,23}, R. M. Tamimi²⁴, N. F. Boyd²⁵; ¹Univ. of Washington, Seattle, WA, ²Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ³Div. of Clinical Trials and Biostatistics, Dept. of Quantitative Hlth.Sci., Mayo Clinic, Rochester, MN, ⁴Lunenfeld-Tanenbaum Res. Inst., Mount Sinai Hosp., Toronto, ON, Canada, ⁵Dept. of Publ. Hlth.Sci., Sch. of Med., Queen's Univ., Kingston, ON, Canada, ⁶Ctr. for Cancer Genetic Epidemiology, Dept. of Publ. Hlth.and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, ⁷QIMR Berghofer Med. Res. Inst., Brisbane, Australia, ⁸Dept. of Hlth.Sci. Res., Mayo Clinic, Rochester, MN, ⁹Div. of Evolution and Genomic Med., Sch. of Biological Sci., Univ. of Manchester, Manchester, United Kingdom, ¹⁰Dept. of Gynecology and Obstetrics, Comprehensive Cancer Ctr. ER-EMN, Erlangen, Germany, ¹¹Galician Publ. Fndn. of Genomic Med. (FPGMX), Santiago de Compostela, Spain, ¹²Moore's Cancer Ctr., Univ. of California, La Jolla, CA, ¹³Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, ¹⁴Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ¹⁵Dept. of Med. Epidemiology and Biostatistics, Karolinska Inst.t, Stockholm, Sweden, ¹⁶Ctr. for Epidemiology and Biostatistics, Sch. of Population and Global Hlth., Univ. of Melbourne, Melbourne, Australia, ¹⁷Div. of Genetics and Epidemiology, The Inst. of Cancer Res., London, United Kingdom, ¹⁸Dept. of Epidemiology and Population Hlth., Stanford Univ. Sch. of Med., Stanford, CA, ¹⁹Postgraduate Sch. of The Cyprus Inst. of Neurology and Genetics, Nicosia, Cyprus, ²⁰Sch. of Population and Global Hlth., Univ. of Western Australia, Perth, Australia, ²¹Div. of Epidemiology, Dept. of Quantitative Hlth.Sci., Mayo Clinic, Rochester, MN, ²²Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ²³Publ. Hlth.Sci. Div., Fred Hutchinson Cancer Res. Ctr., Seattle, WA, ²⁴Div. of Epidemiology, Population Hlth.Sci., Weill Cornell Med., New York, NY, ²⁵Princess Margaret Comprehensive Cancer Ctr., Toronto, ON, Canada

Abstract Body:

Background: Several genetic variants have been found to be associated with mammographic density, a strong risk factor for breast cancer. However, most prior studies consisted of post-menopausal women, and genetic factors associated with mammographic density may differ across the lifespan. In particular, our understanding of the genetic influence of breast composition during adolescence, pregnancy, and breastfeeding is very limited. Methods: We utilized existing genotype and breast composition data from four studies, including a study of young women with magnetic resonance measurements of breast water (N=851), and studies of pre-menopausal women with digitized mammograms from the Breast Cancer Association Consortium (BCAC, N=6,322). Measurements of breast density within each study were transformed and standardized to improve comparability across studies. We performed a GWAS meta-analysis in up to 7,173 women for three mammographic density phenotypes: percent density (PD), dense area (DA), non-dense area (NDA), and adjusting each study for age at breast composition measurement, BMI, and principal components. Results: We identified one region, 8p11.23, associated with NDA at a genome-wide significant level (p-value = 1.01×10^{-10} for lead SNP, rs16885613). This variant has been identified in previous mammographic density and breast cancer GWAS with the same directions of effect. We did not identify any statistically significant findings for the DA and PD GWAS. However, we observed suggestive significance for associations with MD phenotypes, including 19q13.31 with NDA (p-

ASHG 2022 Annual Meeting Poster Abstracts

value = 1.59×10^{-07} for lead SNP, rs56681946) and 4q13.3 with DA (p-value = 1.10×10^{-07} for lead SNP, rs62314947). When restricting the analysis to the young women only, we observed no genome-wide significant associations. Conclusion: Despite the high heritability of mammographic density (approximately 0.60-0.70), very little is known about the genetic variants and biological mechanisms influencing breast tissue composition in early life. Additional analyses in larger and more diverse sample sizes of young and/or pre-menopausal women are warranted to better elucidate genetic factors associated with breast tissue composition, including mammographic density, over the life course.

Complex Traits Posters - Wednesday

PB1241. A Haptoglobin (HP) Exon Deletion Polymorphism Alters the Effect of *APOE* Alleles on Alzheimer's Disease in Individuals of European Ancestry with *APOEε4*

Authors:

H. Bai^{1,2}, A. Naj³, Y. Leung⁴, L-S. Wang⁴, G. Schellenberg⁴, R. Mayeux⁵, M. Pericak-Vance⁶, M. Pericak-Vance⁷, L. Farrer⁸, L. Dumitrescu⁹, T. Hohman⁹, B. Kunkle⁷, E. Martin⁷, J. Haines^{1,10}, W. Bush^{1,10}; ¹Cleveland Inst. for Computational Biology, Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, ²System Biology and Bioinformatics, Dept. of Nutrition, Case Western Reserve Univ. Sch. of Med., Cleveland, OH, ³Dept. of Pathology and Lab. Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, ⁴Penn Neurodegeneration Genomics Ctr., Dept. of Pathology and Lab. Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, ⁵Dept. of Neurology, Columbia Univ., New York, NY, ^{6,7}John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL, ⁸Dept.s of Med. (BioMed. Genetics), Neurology, Ophthalmology, Biostatistics, and Epidemiology, Boston Univ. Sch.s of Med. and Publ. Hlth., Boston, MA, ⁹Vanderbilt Memory & Alzheimer's Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, ¹⁰Dept. of Population & Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Abstract Body:

Haptoglobin (HP) is an antioxidant of apolipoprotein E (APOE) — the strongest risk gene for sporadic Alzheimer's disease (AD). The *HP* gene has two functional alleles, *HP2* and *HP1*, which contains a two-exon deletion that changes its protein structure conformation. We hypothesize that this structural variant is associated with AD. To investigate this, we imputed the *HP* genotypes for 12,418 cases and 11,684 controls from the Alzheimer's Disease Genetics Consortium (ADGC), respectively within each cohort of 29 European ancestry cohorts, using a custom reference panel. We then evaluated the effects of *HP* genotype and *APOE* genotype on AD status, both individually and their interaction, using a logistic regression model, adjusting for sex, age, and the first 3 ancestry-based principal components. In this model, we included the *APOE ε2*, *ε3*, and *ε4* alleles by encoding them additively (i.e. 0, 1, and 2) as a continuous variable — *APOEε2-3-4* — to capture both the protective effect of *APOEε2* and the detrimental effect of *APOEε4*. We found *HP* influences AD risk through interactions with *APOE*. Among *APOEε4* carriers (i.e. individuals with *APOEε24*, *ε34*, or *ε44*, n=9,949), *HP2* additively decreased AD risk (OR=0.83 per allele, p=0.031). Furthermore, the interaction of *HP* and *APOEε2-3-4* among *APOEε4* carriers also showed that each copy of the *HP2* allele leads to an increase in AD risk from *APOEε2-3-4* (OR=1.18 per allele, p =0.025). In contrast, among *APOEε2* carriers (n=2,515) the *HP2* allele increased the AD risk (OR=1.40 per allele, p=0.071), while the *HP2* allele led to a decreased change of the risk in *APOEε2-3-4* (OR=0.75 per allele, p=0.035). Analyses of *APOEε3* carriers did not show any significant effects of the *HP* alleles or any significant *HP/APOE* interactions. We also analyzed age at AD onset with a Cox proportional hazard regression model. Significant effects were found for *APOEε2-3-4*, and the *HP2* effect was marginally significant (p=0.07). The directions of these effects (positive/negative) are congruent to our findings in logistic regression models. *HP* impacts AD risk and modifies the effect of *APOE*. These findings suggest the *HP* genotypes should be considered where *APOE* adjustment and stratification is needed for AD risk determination. Though the precise mechanism still awaits investigation, HP is known to bind directly to APOE and our results may suggest a possible differential anti-oxidative ability of HP proteins for distinct APOE proteins that impact their function.

Complex Traits Posters - Thursday

PB1242. A large-scale genome-wide association meta-analysis of polycystic ovary syndrome in over 540,000 women provides insights into biological mechanisms

Authors:

K. Actkins¹, L. M. E. Moolhuijsen², F. Day³, T. Karaderi⁴, B. H. Mullin⁵, J. Zhu⁶, N. Pujol-Gualdo⁷, C. K. Welt⁸, L. K. Davis⁹, International PCOS Consortium; ¹NIH, Research Triangle Park, NC, ²Univ. Med. Ctr. Rotterdam, Rotterdam, Netherlands, ³Cambridge Univ., Cambridge, United Kingdom, ⁴Univ. of Cambridge Sch. of Clinical Med., Cambridge, United Kingdom, ⁵Univ. of Western Australia, Perth, Australia, ⁶Boston Children's Hosp., Boston, MA, ⁷Univ. of Tartu, Tartu, Estonia, ⁸Univ. of Utah Sch. of Med., Salt Lake City, UT, ⁹Vanderbilt Univ Med Ctr., Nashville, TN

Abstract Body:

INTRODUCTION: Women with polycystic ovary syndrome (PCOS) often exhibit adverse cardiometabolic profiles in addition to suffering from poor reproductive outcomes and increased testosterone levels. Symptom variation exhibited by patients has been attributed to the complex genetic architecture of PCOS, where symptoms can present as reproductive or metabolic features. To identify novel genetic associations involved in the etiology of the syndrome, we performed the largest genome-wide association (GWA) meta-analysis of PCOS to date in women of diverse ancestral origins. **MATERIAL AND METHODS:** We conducted a fixed-effect, inverse-variance weighted meta-analysis in 11,653 PCOS cases and 423,614 controls from 13 cohorts of European (86.5%), East Asian (10%), African (2.2%), and Hispanic (1.3%) ancestry. (1) Age-adjusted, and (2) age and body mass index (BMI)-adjusted models were used in all-ancestries combined and European only meta-analyses. Results were then meta-analyzed with a previously published European GWA meta-analysis, bringing the total to 21,570 PCOS cases and 523,971 controls. Secondary analysis included annotation of genome-wide identified variants using Summary-data-based Mendelian Randomization (SMR) in relevant tissues from the Genotype-Tissue Expression project and a genetic risk score (GRS) analysis within the UK Biobank using PRS-CS. **RESULTS:** In total, 21 loci were identified in the all-ancestries meta-analysis, of which 8 loci were novel. For the first time, a genetic association in the *FTO* region was identified, supporting reported epidemiological associations between PCOS and BMI. Using SMR, we identified several potential effector genes acting through reproductive tissues, including the DNA-repair *NEIL2* gene located at 8p23.1 in ovarian and pituitary tissues. Similar to the previous GWA meta-analysis, risk alleles were linked to genes involved in regulating reproductive hormonal pathways, including *FSHB*, *SHBG*, *INHBB* and *TEX41*. PCOS GRS was significantly associated with cardiovascular, metabolic, and mental health outcomes, such as coronary artery disease and depression, in both men and women. **CONCLUSION:** The current study identified novel PCOS-associated genetic variants, providing new insights into biological mechanisms involved in PCOS etiology, and highlighting disease implications for women as well as men within the genetic risk spectrum of PCOS.

Complex Traits Posters - Wednesday

PB1243. A Mendelian randomization study of genetic liability to Post-traumatic stress disorder and risk of Ischemic Stroke.

Authors:

O. Soremekun¹, C. Musanabaganwa², A. Maddalena³, T. Chikowore⁴, M. Uddin⁵, S. Fatumo⁶; ¹MRC/UVRI & LSHTM, Entebbe, Uganda, ²Med. Res. Ctr., Rwanda Biochemical Ctr., Rwanda, Rwanda, ³Dept. of Epidemiology and Biostatistics, Imperial Coll., London, United Kingdom, ⁴16MRC/Wits Dev.al Pathways for Hlth.Res. Unit, Dept. of Pediatrics, Faculty of Hlth.Sci., Univ. of the Witwatersrand,, Johannesburg, South Africa, ⁵Coll. of Publ. Hlth., Univ. of South Florida, University of South Florida, FL, ⁶UVRI P.O.Box 49, Entebbe-Uganda., Entebbe, Uganda

Abstract Body:

Background:Observational studies have shown an association between post-traumatic stress disorder (PTSD) and ischemic stroke (IS) but given the susceptibility to confounding it is unclear if these associations represent causal effects. Mendelian randomization (MR) facilitates causal inference that is robust to the influence of confounding.**Methods:** Using two sample MR, we investigated the causal effect of genetic liability to PTSD on IS risk. Ancestry specific genetic instruments of PTSD and four quantitative sub-phenotypes of PTSD, including hyperarousal, avoidance, re-experiencing, and total symptom severity score (PCL-Total) were obtained from the Million Veteran Programme (MVP) using a threshold P-Value (P) of $< 5 \times 10^{-7}$, clumping distance of 1 Megabase (Mb) and $r^2 < 0.01$. Genetic association estimates for IS were obtained from the MEGASTROKE consortium (Ncases = 34,217, Ncontrols = 406,111) for European ancestry individuals and from the Consortium of Minority Population Genome-Wide Association Studies of Stroke (COMPASS) (Ncases = 3,734, Ncontrols = 18,317) for African ancestry individuals. We used the inverse-variance weighted (IVW) approach in our primary analysis and performed MR-Egger and the weighted median methods as pleiotropy-robust sensitivity analyses.**Results:** In European ancestry individuals, we found evidence of an association between genetic liability to hyperarousal, avoidance, and PCL-Total and IS risk (odds ratio (OR)1.06, 95% Confidence Interval (CI) 1.000 to 1.125, P = 0.031 for hyperarousal; OR 1.05, 95% CI 1.016 to 1.085, P = 0.004 for avoidance; OR 1.03, 95% CI 1.010 to 1.041, P = 0.001 for PCL-Total). In African ancestry individuals, we found evidence of an association between genetically liability to PCL-Total and IS risk (OR 0.95 (95% CI 0.913 to 0.988, P = 0.01) but no association was observed for PTSD case control, hyperarousal, avoidance, or re-experiencing. Similar estimates were obtained with MR sensitivity analyses. **Conclusion:** This study provides genetic support for a casual effect of some PTSD traits on IS risk in European ancestry individuals. Except in PCL-Total, no MR evidence was found in some PTSD traits and risk of IS in African ancestry individuals. These findings warrant further investigations in larger genetic cohorts especially in African ancestry individuals.

Complex Traits Posters - Thursday

PB1244. A multi-ancestry genome-wide association study identifies two novel loci associated with increased risk of diabetic macular edema.

Authors:

A. Stockwell¹, M. Chang¹, W. Forrest¹, L. Orozco¹, S. Selvaraj¹, N. Creps¹, A. Mahajan¹, L. Macri², A. N. Neerajan², M. McCarthy¹, M. van der Brug², S. Scales¹, B. Yaspan¹; ¹Genentech, South San Francisco, CA, ²Character BioSci.s, San Francisco, CA

Abstract Body:

Diabetic retinopathy (DR) is a common complication of diabetes. Approximately 20% of DR patients have diabetic macular edema (DME), characterized by fluid leakage into the retina. While there is an accepted genetic component to DR and DME risk, currently there are few replicable loci. Focusing on the DME subtype could increase power to find novel loci associated with DME risk.

We conducted a multi-ancestry, common variant (MAF $\geq 1\%$) genome-wide association study to assess DME risk in 1,195 DME patients and 4,703 controls using an additive model, adjusted for genetic ancestry, age, and sex. Ancestry specific cohorts were analyzed separately. An additional cohort of 307 DME cases and 900 controls was used to replicate genome-wide significant results using METAL. Patient samples were obtained from Roche/Genentech clinical trials and Character Biosciences.

We identified two novel loci reaching the pre-specified GWAS significance threshold of $p \leq 5E-8$. The strongest association in our discovery dataset was rs2239785 (OR=1.71; $P=5.05E-23$), a missense variant (K150E) in *APOLI* previously associated with chronic kidney disease (CKD) in African Americans. The frequency of this SNP was highest in African American DME patients (EAF=0.76), more than twice of European DME patients (EAF=0.31). The population attributable risk percent in the African American cohort is 5.16 for the *APOLI* locus, a substantially increased compared to those of European descent (1.85). This is due to the G1 and G2 haplotypes at *APOLI*, unique to individuals of African descent, and also well-established to associate with CKD. The second genome-wide significant peak was rs10402468 (OR=1.58; $P=4.6E-13$), intronic within *BABAMI* and a cis-eQTL with *PLVAP* and *ANKLE1* in vascular / endothelium tissues. We replicated both rs2239785 (OR=1.41; $P=0.003$) and rs10402468 (OR=1.31; $P=0.03$) in an independent cohort.

In our multi-ethnic GWAS, we report two novel loci for risk of DME, one of which, *APOLI*, is a major risk factor in African Americans with diabetes for developing DME complications. Our lead SNP at *APOLI*, rs2239785, is polymorphic in multiple ancestries; however, unique variation of African descent on the same haplotype identified in CKD patients substantially increases its overall risk on a population level.

Complex Traits Posters - Wednesday

PB1245. A multi-ethnic genome-wide association study in type 1 diabetes.

Authors:

D. A. Michalek¹, C. Tern¹, W-M. Chen^{1,2}, W. Zhou^{3,4,5}, T1DGC, S. Onengut-Gumuscu^{1,2}, S. S. Rich^{1,2}; ¹Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, ²Dept. of Publ. Hlth.Sci., Univ. of Virginia, Charlottesville, VA, ³Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, ⁴Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, ⁵Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA

Abstract Body:

Background: Type 1 diabetes (T1D) is an autoimmune disorder in which pancreas produces little or no insulin from the destruction of pancreatic β -cells. To date, GWAS meta-analyses and fine mapping studies have identified over 80 loci associated with risk of T1D. **Objective:** Utilize a multi-ethnic GWAS approach in T1D cases, controls and families for discovery of novel genetic variants associated with the risk and age-at-onset of T1D. **Methods:** The T1DGC data included 3,222 families (11,476 individuals, majority of European ancestry), 891 unrelated African ancestry (AFR, 409 T1D cases, 482 controls) and 308 unrelated Admixed (AMR, 153 T1D cases, 155 controls) individuals. All samples were genotyped on the Illumina CoreExome BeadChip array. A total of 430,930 SNPs passed quality control metrics and were used for imputation (TOPMed), and HLA region gene and amino acid alleles were imputed on the four-digit multi-ethnic HLA reference panel. 13,965,323 variants (MAF > 0.01) were obtained from imputation, including 20,425 HLA region variants (MAF > 0.005). HLA region variants were analyzed using HLA-TAPAS. GWAS for the T1D risk was conducted based on a logistic mixed model using SAIGE and GWAS for the age-at-onset of T1D was performed based on a survival model using GATE. **Results:** Five loci were significantly associated with T1D risk and/or age-at-onset, including the MHC (T1D risk: rs9273364, $P = 4.73 \times 10^{-322}$; age-at-onset: rs9273368, $P = 1.13 \times 10^{-290}$; both in *HLA-DQB1*), *INS* (rs3842753, T1D risk: $P = 1.06 \times 10^{-29}$; age-at-onset: $P = 3.01 \times 10^{-26}$), *PTPN22* (T1D risk: rs6679677, $P = 6.22 \times 10^{-15}$; age-at-onset, $P = 1.38 \times 10^{-14}$), *SH2B3* (T1D risk: rs10774624, $P = 1.40 \times 10^{-11}$; age-at-onset, $P = 2.22 \times 10^{-10}$), *ERBB3* (T1D risk: rs11171747, $P = 2.40 \times 10^{-8}$) and *RAB5B* (age-at-onset: rs773107, $P = 2.15 \times 10^{-8}$). A new locus, *NRPI* (rs722988, $P = 2.18 \times 10^{-8}$) on 10p11.22 was identified and reached the genome-wide significance in AFR, AMR only meta-analysis. Fine mapping across ancestry groups in the HLA region revealed *HLA-DQA1*03:01* as the most significant HLA allele in AFR and AMR ($P > 10^{-8}$) and *HLA-DRB1*03:01* in EUR ($P > 10^{-8}$). **Conclusions:** Inclusion of subjects with diverse genetic ancestry allowed us to identify new genome regions associated with risk and progression of T1D. Age-at-onset analyses pinpointed five regions harboring variants that may accelerate or delay onset of T1D. Further fine mapping studies can help us identify new targets that can be exploited to delay T1D progression and improve patient outcome.

Complex Traits Posters - Wednesday

PB1246. A nationwide approach to understand the role of health, socioeconomic and genetic information in predicting COVID-19 vaccination uptake

Authors:

A. Ganna¹, B. Jermy², T. Hartonen¹, P. Vartiainen¹, M. Daly³, H. Ollila², S. Ripatti⁴, M. Perola⁵; ¹Inst. for Molecular Med. Finland, Helsinki, Finland, ²Univ. of Helsinki, Helsinki, Finland, ³Massachusetts Gen. Hosp., Boston, MA, ⁴Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, ⁵Finnish Inst Hlth & Welfare, Helsinki, Finland

Abstract Body:

The high prevalence of people unvaccinated for COVID-19 is a key issue in public health. Identifying individuals likely to decline vaccinations can help to reduce the disease burden. Here, we aim to identify potential predictors of COVID-19 vaccination uptake by combining nationwide registries covering disease history, drug purchases, environmental, socioeconomic and demographic information for the entire Finnish population (5.5 million individuals). We considered 2,994 predictors measured before 31.12.2019 and vaccination uptake between 27.12.2020 and 31.10.2021. To identify genetic predictors, we performed a GWAS of vaccination uptake in a subset of 247,600 individuals (8% unvaccinated) part of the FinnGen study.

Among the strongest phenotypic predictors were speaking a mother tongue other than Finnish or Swedish (OR \approx 1.27, P-value $<$ 2.2e-16), having ever received social assistance (OR \approx 1.08, P-value \approx 4.0e-4), having ever been in an assisted living facility (OR \approx 1.11, P-value \approx 1.0e-3) and having had an Hepatitis C infection (OR \approx 1.23, P-value \approx 2.5e-3). We combined individual predictors within consistent phenotypic categories and used LASSO to predict vaccination uptake. We showed that income in 2019, job occupation, drug purchase history, and place of residence were the most predictive categories. For example, a model trained using income information from year 2019 obtained an AUC of 0.687 (95% CI: 0.684-0.689) which is approximately 16% higher than a baseline model (age and sex). Each of the individual categories were outperformed by a model combining all predictors, obtaining an AUC of 0.774 (95% CI: 0.772-0.776).

We observe a strong genetic correlation with a GWAS of participation in optional components of UK Biobank ($r_g = -0.35$, $p_{val} = 6 \times 10^{-7}$) indicating shared underlying effects between participation in scientific studies and vaccination uptake. Higher genetic risk for COVID-19 is correlated with not taking the vaccine ($r_g = 0.22$, $p_{val} = 0.0378$). A polygenic risk score predicting vaccination status obtained an AUC of 0.578 (95% CI: 0.567-0.589). Replication in Estonia Biobank is ongoing.

We provide a comparative, nationwide framework to identify the most important predictors of COVID-19 vaccination. Socio-economic information such as job occupation were the strongest predictors of vaccination uptake, more than health and genetic information. Genetic signal was low ($h^2_{SNP} \sim 1\%$), yet genetic correlations could be used to identify unmeasured behavioral correlates of vaccination uptake. Data-driven identification of subpopulations to target with information campaigns can help in resource allocation and in reducing the COVID-19 burden.

Complex Traits Posters - Thursday

PB1247. A novel missense variant in melanopsin associated with delayed sleep phenotype: A whole genome sequencing study.

Authors:

J. Brzezynski, A. Kaden, C. Johnson, S. Smieszek, C. Polymeropoulos, G. Birznieks, M. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC

Abstract Body:

Melanopsin (*OPN4*) is a blue light-sensitive opsin-type G-protein coupled receptor (GPCR). It is highly expressed in photosensitive retinal ganglion cells which mediate responses to light, including regulation of sleep, circadian photoentrainment, and pupillary light response. Consequential variants in *OPN4* were previously associated with increased risk of developing seasonal affective disorder.

We have conducted a rigorous observational research study in suspected delayed sleep-wake phase disorder (DSWPD) patients with the aim of detecting consequential variants that may be associated with the delayed sleep phenotype. Altogether, 119 participants were consented and 76 participants provided samples for whole genome sequencing (WGS).

We report a carrier of rs143641898 *OPN4*:NM_033282:exon4:c.C502T:p.R168C, a rare missense variant (gnomAD MAF 0.0002, predicted damaging). R168C is a highly conserved variant and is part of the E/DRY motif found in nearly all GPCRs (positive charge to polar, increased hydrophobicity).

Interestingly, this variant was tested using an *in vitro* expression system as part of a study aiming to determine the functional phenotypes of missense human *OPN4* variants. This variant was shown to be incapable of binding retinal chromophore, suggesting that it renders the *OPN4* protein non-functional. Additionally, introduction of R168C in *OPN4* abolished responses to light. This carrier is a 61-year-old White male manifesting delayed sleep symptoms since high school, meeting the DSWPD diagnostic criteria. Their Dim Light Melatonin Onset (DLMO) was 23:25 (based on a salivary threshold of 3 pg/mL), which also supports the delayed phenotype. We do not detect this variant in our healthy super control set of WGS samples (n = 300), nor in our control set of WGS samples (n = 1900). In this WGS study of suspected DSWPD patients, we report one other rare stop gain variant within the *OPN4* gene (*OPN4*:NM_033282:exon6:c.847_848insGGTAGC;p.W283delinsWX).

The identification of consequential *OPN4* variants such as R168C that lead to the disruption of protein function and melanopsin-based light perception will help identify patients with an increased risk of sleep disturbances and circadian dysfunction.

Complex Traits Posters - Wednesday

PB1248. A pan cancer immunogenomic atlas and its applications for immune checkpoint blockade immunotherapy

Authors:

J. Yang, Q. Liu, Y. Shyr; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

A low response rate is the major challenge for existing immune checkpoint blockade (ICB)-based cancer immunotherapies. A number of biomarkers have been reported to be associated with the likelihood of patient responses to ICB, including the expression of target ligands such as *PD-L1*, co-inhibitory receptors such as *PDCDI*, tumor mutational burden (TMB), tumor infiltration lymphocytes, T cell repertoire. However, these biomarkers showed variable performance in different cohorts. And most published signatures were only evaluated on limited cohorts and showed ungeneralizable performance across different cohorts. In addition, establishing a cohort with adequate sample size requires multiple years of multi-center efforts. Therefore, it is still a lack and a challenge for evaluation of efficacy of known biomarkers and the discovery of new signature(s) in a large-scale study. Here we developed Cancer-Immu (<http://bioinfo.vanderbilt.edu/database/Cancer-Immu/>) to prioritize genomic features associated with ICB response, containing 3,652 samples for 16 cancer types. Cancer-Immu provides two analysis strategies, meta-analysis and pan-cancer analysis. Meta-analysis reveals consistent signatures across multiple tumor/study cohorts, while pan-cancer analysis enhances our ability to detect and analyze rare features by aggregating samples across cohorts/tumor types. Furthermore, Cancer-Immu enables linking dynamic features with immunotherapy response. Cancer-Immu also allows users to upload and analyze their own data independently or to co-analyze with existing data simultaneously. Based on these data and methods, we not only identified known features such as TMB, *IFNG*, *PDCDI* and *LAG3* expression, but also novel biomarkers such as proportions of M1 macrophage, expression of *IL21*, *ZBED2*, *CCL19*, CTL pathway and CXCR3 binding pathway. Multiple genes and pathways related to ICB response are involved in macrophage activity, T cell cytotoxicity and recruitment, suggesting the potential role of interaction between T cells and macrophages in cancer immunotherapy. Single-cell RNA sequencing confirmed that *IFNG* from T cells induced *IRF1* expression of M1 macrophage, TAM, and tumor, promoting T cell chemo-attractants CXCL10 and/or CXCL11 secretion in those cells. The secretion of CXCL10/11 further recruited T cells and enhanced anti-tumor immunity through CXCR3-CXCL10/11 axis. Overall, Cancer-Immu provides a comprehensive resource for identifying predictive genomic features of immunotherapy response, and presents an easy way for known biomarker assessment, novel signature discovery, and signature-of-interest validation.

Complex Traits Posters - Thursday

PB1249. A polygenic score-based approach improves discrimination between familial, idiopathic, and pathologic short stature in a cohort of pediatric patients presenting for evaluation

Authors:

J. Shelley¹, M. Shi², B. Schuler¹, S. Van Driest¹, J. H. Simmons¹, J. Mosley³; ¹Dept. of Pediatrics, Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Dept. of BioMed. Informatics, Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Introduction: Short stature in children may be an early manifestation of systemic disease or related to a benign, genetic predisposition to shorter height. A child's anticipated height is estimated by mid-parental height (MPH), an average of the parental heights. Children with short stature who have a low MPH are classified as having "familial short stature" (FSS). Children shorter than their MPH without underlying disease are classified as having "idiopathic short stature" (ISS), while those found to have underlying disease have a pathologic form of short stature (PSS). Height is a highly polygenic trait. We tested whether a polygenic score (PGS) for height could discriminate among children classified as FSS, ISS, and PSS.

Methods: We used summary statistics from the GIANT-UKB meta-analysis of adult height to create a PGS comprising 4,803 independent SNPs (r -square=0.01, MAF>1%) associated ($p < 5 \times 10^{-5}$) with height. We curated a cohort of pediatric patients referred for an evaluation of short stature from a clinical DNA biobank linked to an electronic health record (BioVU) to identify children with ISS, FSS and PSS. Wilcoxon signed-rank tests were used to test for differences in the PGS values across diagnoses and c -statistics were used to measure discrimination.

Results: From the cohort of 535 children, 73 were diagnosed with ISS, 44 with FSS, and 39 with PSS. 37.2% were female and median age was 9.5 (IQR: 6.6, 12.5) years. PSS was most commonly due to autoimmune disease, genetic syndrome, or hypopituitarism. The respective median (IQR) z -scores for MPH and PGS were (-0.74 [-1.50, -0.45], -0.92 [-1.49, -0.40]) for the FSS group, (0.06 [-0.68, 0.62], -0.90 [-1.61, -0.33]) for ISS and (0.39 [0.17, 0.73], 0.06 [-0.53, 0.78]) for PSS. There was a significant difference in the PGS values between PSS and both FSS (difference=0.98 standard deviations, $p = 2.7 \times 10^{-5}$), and ISS (difference=0.96, $p = 3.6 \times 10^{-6}$). There was no difference in the PGS values between FSS and ISS ($p = 0.73$). The PGS significantly discriminated both FSS from PSS (c -statistic: 0.76 [95% CI: 0.65, 0.87]) and ISS from PSS (0.77 [0.67, 0.86]), but not FSS and ISS (0.52 [0.42, 0.57]).

Conclusion: A PGS for height differed significantly among these short stature phenotypes. Consistent with classification based on MPH, children with pathological short stature were not genetically predisposed to lower height. However, children diagnosed with idiopathic short stature were frequently predisposed to shorter height based on the PGS, inconsistent with their MPH estimates. This suggests that some of these children likely have familial short stature. A PGS could improve classification of some short stature subtypes.

Complex Traits Posters - Wednesday

PB1250. A second wave SARS-CoV-2 variant in Quebec is associated with persistent infection.

Authors:

D. Fournelle^{1,2}, E. Brunet-Ratnasingham^{1,3}, F. Mostefai^{1,2}, R. Poujol², J-C. Grenier², N. Chomont^{1,3}, S. Moreira⁴, D. E. Kaufmann^{1,3}, M. Craig^{1,5}, J. Hussin^{1,2}; ¹Univ. de Montreal, Montreal, QC, Canada, ²Montreal Heart Inst., Montreal, QC, Canada, ³Ctr. de recherche du CHUM, Montreal, QC, Canada, ⁴Laboratoire de santé publique du Québec, Montreal, QC, Canada, ⁵Ctr. de recherche du CHU Sainte-Justine, Montreal, QC, Canada

Abstract Body:

We are still living in the SARS-CoV-2 pandemic over two years after its start. In the Canadian province of Québec, the second wave of infections occurred between the end of August 2020 and the end of March 2021. This was an interesting period of viral diversity owing to imposed travel restrictions that created competition between multiple pockets of local strains prior to the introduction of international variants of concern such as Alpha, Delta and Omicron. The height of the second wave in Quebec was characterized by two now defunct variants (B.1.160 and B.1.1.176) that accounted for just under 50% of sequenced cases in the province. The B.1.160 variant has a mutation to the S protein (S477N) that is known to strengthen binding to receptor ACE2. In Québec, a substrain of B.1.160 also developed a T20I mutation on the S protein. Case studies from patients infected with the Quebec specific B.1.160, showed that this specific variant had the potential to cause infections that persisted for several months, including periods during which the virus was not detectable by PCR tests. We analyzed 4 time points and found that the second and third time points, taken a little before 5 months of infection, have several mutations that completely disappear by the fourth time point, taken over 6 months after the first sample was collected. The patient also acquires 2 different mutations on codon 484 of the S protein, both associated with immune escape and present in VOCs Beta, Gama and Omicron. This longitudinal analysis shows that mutational patterns of intra-host mutations in our patient are in line with the presence of a viral reservoir. This raises the question of whether some cases of long COVID were due to lingering infections that could not be detected by PCR tests.

Complex Traits Posters - Thursday

PB1251. A single nucleus transcriptomics study of alcohol use disorder in postmortem brain tissue

Authors:

S. Clark¹, L. Y. Xie², K. A. Aberg², E. J. C. Van den Oord²; ¹Texas A&M Univ., College Station, TX, ²Virginia Commonwealth Univ., Richmond, VA

Abstract Body:

Alcohol use disorder (AUD) is a chronic and debilitating brain disorder. Gene expression studies offer promising opportunities to better understand the underlying pathogenic processes. As cells differ in their function, gene expression will typically vary across cell types. When studying bulk tissue, failure to account for this cellular diversity has a detrimental impact on the ability to detect disease associations. We therefore assayed the transcriptomes of 32,531 individual nuclei extracted from the nucleus accumbens (NAc) of 9 donors with AUD and 9 controls. We obtained 241,497 high quality reads per nucleus. Our study identified 17 clearly delineated cell-types. We detected 26 transcriptome wide significant association signals (q value < 0.1) that mainly involved a group of medium spiny neurons with both D1 type and D2 type dopamine receptors, microglia and oligodendrocytes. The alcohol related pathways detected for each cell type were consistent with the functions of those cells reported in the literature. Thus, for neurons we observed pathways for alcohol related neurodegeneration, disruption of circadian rhythms, alterations in glucose metabolism, and changes in synaptic plasticity. For microglia we found neuroinflammation and immune related pathways and for oligodendrocytes disruptions in pathways related to myelination. This identification of the specific cells from which the association signals originate is key for designing proper followup experiments and, eventually, for developing new and targeted clinical interventions.

Complex Traits Posters - Wednesday

PB1252. A Systematic Multi-omics Approach Identifies miRNA and Proteomic Dysregulation in Human Post-traumatic Stress Disorder

Authors:

J. Wang, R. Wilson, H. Li, T. Lam, A. Nairn, Traumatic Stress Brain Research Group, J. Krystal, R. S. Duman, H. Zhao, M. Girgenti; Yale Univ., New Haven, CT

Abstract Body:

Post-traumatic stress disorder (PTSD) is a severe psychiatric disease with a prevalence of ~7% in the general population. However, despite its prevalence, our understanding of the molecular determinants is limited and examination of postmortem brain tissue from donors with PTSD across genomic layers is necessary to better understand the etiopathology of this disorder. Here we present findings from the first genome-wide microRNA (miRNA) expression profile followed by proteomic analysis in postmortem PTSD brain. We identified 124 differentially expressed miRNAs that organized into PTSD-associated co-expression modules. We confirmed regulatory relationships between several differentially expressed miRNAs and their putative target proteins and protein co-expression modules. Integrative analysis identified miRNA hsa-mir-589 as a core regulating miRNA in the translational process of disease-associated protein modules for PTSD. In addition, we identified significant enrichment of genetic risk genes for other neurodegenerative and psychiatric disorders within these PTSD co-expression modules, suggesting a shared molecular pathology. Our findings support a role for miRNA dysfunction in the pathological alterations of PTSD and provide a novel framework for future studies integrating small RNA transcriptomics with proteomic profiling.

Complex Traits Posters - Thursday

PB1253. A trans ancestry genomics based approach to study the interplay between the immune system, infectious type, and HLA type, ancestry and sepsis outcome

Authors:

K. Chhugani, T. Ramesh, Y. Chang, B. Nadel, Y. Patel, A. Wong-Beringer, S. Mangul; Univ. of Southern California, Sch. of Pharmacy, Los Angeles, CA

Abstract Body:

Sepsis is a life threatening, dysregulated host response to infection that is a major cause of mortality worldwide. Sepsis has complex pathobiology which relates to the infectious agent and host immune response. The outcome of sepsis is the result of a complex interaction between pathogens and host immune defense during the course of infection. Current research approaches usually over simplify the interactions between pathogens and innate and adaptive immune systems. We developed and benchmarked transcriptomic based bioinformatics methods to accurately infer the features of the innate and adaptive immune system and virulence of infectious agents. We performed extensive benchmarking of the developed method based on the most comprehensive gold standard dataset available. We combined publicly available transcriptomics data into the largest trans-ancestry retrospective sepsis cohort with a rich set of clinical phenotypes. Our cohort will include over 3705 individuals diagnosed with sepsis across 33 individual studies. We focused on the identification of immune features like cell type composition and sepsis type respectively, associated with the severity of the infection and poor outcomes in sepsis and studying the complex interplay between the immune system and the infectious type (viral or bacterial), facilitating insights on clinical severity scores and survival status of patients. To achieve this goal, we assembled a large scale multicentre and trans-ancestry retrospective sepsis (MCMERS) cohort, including individuals diagnosed with sepsis from publicly available datasets. We compared the cell type composition of samples derived from survivors and non-survivors groups using our recently developed tool (GEDIT). Our results suggest that the relative abundance of various immune cells (e.g. Mast cells, Neutrophils) were significantly different across survivors and non-survivors (p -value $<10^{-4}$). The advanced set of genomic and immunological features which was inferred across a large scale heterogeneous population with available information about genetic ancestry and HLA type makes this study uniquely positioned to provide new insights into the complex immune response to sepsis. Results from our study will improve our understanding of the relationship between the immune system and infectious type (viral or bacterial) across individuals of diverse ancestry backgrounds.

Complex Traits Posters - Thursday

PB1254. Accurate identification of causal variants of intracellular LDL uptake via Bayesian modeling of base editor reporter screens

Authors:

J. Ryu^{1,2}, S. Barkal^{3,1}, T. Yu⁴, L. Pinello^{2,1,5}, R. Sherwood^{1,3}; ¹Harvard Med. Sch., Boston, MA, ²Massachusetts Gen. Hosp., Charlestown, MA, ³Brigham and Women's Hosp., Boston, MA, ⁴Brigham and Women's Hosp., Boston, MA, ⁵Broad Inst., Cambridge, MA

Abstract Body:

Since its first application as a genome editing tool, CRISPR-Cas9 has been harnessed to precisely alter the genome and epigenome. CRISPR base editors that induce single-nucleotide transitions have emerged as powerful tools to test the phenotypic impacts of disease-associated variants in the human population. However, base editing often results in incomplete efficiency at the target base as well as bystander edits within a window of around 5 bp. This variability in editing efficiency and genotypic outcomes hinders the identification of causal variants in high-throughput base editor screens. We, along with other studies, have recently developed base editor reporter techniques to address this problem by incorporating a target site surrogate (reporter) sequence with the same spacer and PAM sequence as the endogenous target site into the guide RNA (gRNA) construct. As the gRNA introduces base edits in both the endogenous target site and the reporter sequence, the overall editing pattern in the endogenous target site can be inferred. To utilize the reporter editing pattern to improve the resolution of base editor screens, we have built a novel computational method that explicitly models the data generation process using the observed editing rate and pattern of the reporter sequence. We have employed this model to analyze base editor reporter screen data from two screens measuring the effects of over 10,000 gRNAs on uptake of fluorescent LDL-cholesterol. These screens include a saturation tiling screen of LDLR and a survey of LDL-cholesterol GWAS-associated variants. Our computational method shows better performance in inferring the correct effect sizes of coding and noncoding variants compared to previous approaches. Our approach has four major advantages over previous screen analysis methods based solely on gRNA counts: 1) flexible normalization to account for the different editing activity of each gRNA; 2) accurate identification of causal edits at individual nucleotides based on the reporter editing patterns; 3) interpretable and statistically robust model estimates of the posterior distribution of phenotype change induced by each variant rather than the log fold change across conditions; 4) flexibility to adapt to various experimental setups such as cell sorting methods as the model directly reflects the experimental procedure that generated the read counts. Moreover, the model improves sensitivity in screens without reporter data by accurately accounting for the editing rate of each guide. This work provides a novel and widely applicable approach to improving the power of base editor screens.

Complex Traits Posters - Wednesday

PB1255. Activity-dependent transcriptional program in glutamatergic neurons enriched for genetic risk for schizophrenia.

Authors:

Y. Ma¹, J. Bendl¹, B. J. Hartley¹, J. F. Fullard¹, R. Abdelaal¹, S-M. Ho¹, R. Kosoy¹, P. Gochman², J. Rapoport², G. E. Hoffman¹, K. J. Brennand³, P. Roussos¹; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²NIMH, Bethesda, MD, ³Yale Sch. of Med., New Haven, CT

Abstract Body:

Converging evidence from large-scale genetic and post-mortem studies highlight the role of aberrant control of synaptic plasticity in schizophrenia. To better understand the molecular mechanisms of activity-dependent regulome contributing to risk for psychiatric disorders, we profiled the epigenomic and transcriptomic changes following membrane depolarization in human induced pluripotent stem cell (hiPSC)-derived glutamatergic neurons from individuals with children-onset schizophrenia patients and healthy controls. Multi-omic data integration associated global patterns of chromatin accessibility with gene expression and reconstructed the maps of enhancer-promoter interactions in glutamatergic neurons. Within one hour of KCl-induced depolarization, glutamatergic neurons displayed substantial changes in the expression of genes regulating synaptic function. Consistently with increased density of schizophrenia risk variants in activity-dependent enhancers, dysfunction of many of those genes was previously likewise linked to neurodevelopmental disorders. Furthermore, network analysis revealed multiple co-expressed modules consisting of both schizophrenia- and activity-associated genes whose key drivers might represent novel targets for advanced diagnosis and therapy. Overall, our framework offers an alternative approach to studying the schizophrenia-related mechanisms but also adds to mounting evidence that sequence variation within activation-dependent regulatory elements contribute to the genetic risk for schizophrenia.

Complex Traits Posters - Thursday

PB1256. Advancing rare disease research through web-based recruitment: the 23andMe Systemic Sclerosis Research Study

Authors:

K. Kukar¹, C. Wong¹, M. Moreno¹, J. Matthews¹, L. Babalola¹, W. Fahy², M. Ehm², J. Reid², S. Shringarpure¹; ¹23andMe, Inc, Sunnyvale, CA, ²GlaxoSmithKline, Brentford, United Kingdom

Abstract Body:

Rare diseases affect fewer than 200,000 individuals in the United States. Due to the low prevalence, it is challenging for scientists to find enough research participants to study rare disease genetics. Here, we demonstrate how rare disease study cohorts can be built through web-based recruitment using various patient-engagement approaches, with the example of the 23andMe Systemic Sclerosis Research Study. Systemic sclerosis (SSc) is a rare autoimmune disease characterized by vascular abnormalities, fibrosis of the skin and internal organs and the production of autoantibodies, for which there is no cure. In 2020, 23andMe Inc. launched the 23andMe Systemic Sclerosis Research Study, aimed at enrolling 1000 individuals who self-reported a diagnosis of systemic sclerosis¹. The goals were to learn more about the genetics of people living with the disease to help lead to the development of better treatments. Consenting participants were asked to provide a saliva sample for genetic analysis and complete surveys about their disease over a three year period. We used various methods to engage participants, including online advertisements, blog posts, and a collaboration with the Scleroderma Research Foundation. Within 18 months, 1594 eligible individuals enrolled in the study. Of the enrolled participants, 94% were female. The self-reported ethnicities of the enrolled participants are: 75% White or European, 5% American Indian or Alaska Native, 4% Black or African, 3% Asian, less than 1% Middle Eastern and Native Hawaiian or Pacific Islander, and 12% other. 95% of enrolled participants reported being diagnosed by a rheumatologist, and 94% reported undergoing an autoantibody blood test before diagnosis. We found this cohort to have many clinical features similar to other SSc cohorts. The median age-of-onset was 35 years for Raynaud's phenomenon, and 38 years for non-Raynaud's symptoms. Limited cutaneous SSc was the most commonly reported subtype (43%), with diffuse cutaneous SSc reported by 27%, CREST syndrome by 6%, and the remaining participants (24%) reported being unsure of their disease subtype. Among symptoms reported by participants, digital ulcers (48%) were most common, followed by interstitial lung disease (30%), pulmonary arterial hypertension (14%), and renal crisis (3%). The results from this study show that participant-powered research can help build cohorts to advance the study of rare diseases, and that web-based research models can replicate findings from other clinical cohorts. 23andMe will share data from this study through dbGaP to enable further SSc research.¹ Individuals with morphea or linear and guttate sub-types were ineligible.

Complex Traits Posters - Wednesday

PB1257. Age differences in preadipocyte proportions and expression profiles link to obesity and insulin resistance.

Authors:

A. Kar¹, M. Alvarez¹, H. Huang^{1,2}, S. T. Lee¹, M. Deal¹, K. L. Mohlke³, J. S. Sinsheimer^{1,2,4}, K. H. Pietiläinen^{5,6}, M. Laakso⁷, P. Pajukanta^{1,2,8}; ¹Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ²Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, ³Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁴Dept. of Computational Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ⁵Obesity Res. Unit, Res. Program for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, ⁶Obesity Ctr., Endocrinology, Abdominal Ctr., Helsinki Univ. Central Hosp. and Univ. of Helsinki, Helsinki, Finland, ⁷Dept. of Med., Univ. of Eastern Finland and Kuopio Univ. Hosp., Kuopio, Finland, ⁸Inst. for Precision Hlth. at UCLA, Los Angeles, CA

Abstract Body:

Obesity and related cardiometabolic diseases (CMD) display well-established age effects in epidemiological studies. However, mechanisms underlying age effects have remained largely elusive. We hypothesized that adipose tissue contributes to these CMD-related age effects in a cell-type-specific way. To this end, we first estimated the main adipose cell-type proportions in bulk adipose expression data using Bisque from 53 monozygotic (MZ) BMI-discordant twin pairs utilizing single nucleus RNA-seq (snRNA-seq) data as a reference from 6 of the 106 twins. Next, we divided each BMI discordant MZ twin pair to lower and higher BMI groups and compared cell-type proportions by age in each non-related group. In the lower BMI individuals, the younger individuals (mean age=30.4, SD=4.8) had significantly higher preadipocyte proportions ($p=0.00252$) than the older ones (mean age=64.8, SD=4.0). Noteworthy, as preadipocytes generate adipocytes, their proportions are essential for metabolically healthy adipose tissue functions. No such age-dependent preadipocyte proportion difference was seen in the higher BMI individuals. We replicated these findings in the adipose bulk RNA-seq data from the METSIM cohort ($n=335$, mean age=54.1, SD=4.9), where we similarly estimated bulk preadipocyte proportions using snRNA-seq data from 84 METSIM men. As in the twins, among the normal weight men (BMI<25), the men in the youngest age quartile had higher preadipocyte proportions than the men in the oldest age quartile ($p=0.0435$). No such age-dependent difference in preadipocyte proportions was observed in the overweight or obese METSIM men. Next, we found that 50 of the 151 preadipocyte marker genes are DE by age using snRNA-seq data. The 50 genes are significantly enriched (FDR<0.05) for multiple tissue and organ developmental pathways using WebGestalt. Of the 50 genes, 12 are correlated with adipocyte cell volume ($\text{adj}P<0.05$) and 10 with adipocyte cell weight ($\text{adj}P<0.05$) in the twins, and 21 were significantly correlated with insulin sensitivity using the Matsuda index in METSIM (Q-value<0.05). Our results show that preadipocyte proportions differ by age in normal weight individuals and that obesity reduces this metabolically favorable age-dependency in younger individuals' preadipocyte proportions. We also identify 50 preadipocyte marker genes DE by age and enriched for developmental pathways, of which 42% are correlated with insulin sensitivity. Our results suggest that adipose cell-type specific factors likely contribute to age differences in obesity-related CMDs.

Complex Traits Posters - Thursday

PB1258. Altered expression levels of *TASIR1*, *TASIR3* genes among SARS-CoV-2 variants

Authors:

M. Ergoren¹, Q. Hoti¹, G. Akan², G. Tuncel², T. Sanlidag²; ¹Sch. of Med., Near East Univ., Nicosia, Cyprus, ²DESAM Res. Inst., Near East Univ., Nicosia, Cyprus

Abstract Body:

Background: The most common symptoms of COVID-19 are fever, cough, shortness of breath, headache, ache of joints and etc. But latest scientific and medical records showed that patients with proven COVID-19 have developed a loss of smell (anosmia), and loss of taste (hypogeusia and ageusia). These symptoms are usually temporary and might manifest in absence of any other clinical symptoms but a chronic loss of smell and taste is occurring in many patients. Also, prior studies suggested that smell and taste receptors were associated with pathogenic detection and immunity. We aim to evaluate the expression profile of genes that are related to taste, smell and, appetite receptors in COVID-19 patients and their correlation with SARS-CoV-19 Delta and Omicron BA.1 variants of concerns (VOC). **Method:** Expressions of *TASIR1*, *TASIR3*, *TAS2R38*, *OR51E1*, *LEPR*, *GHRL* were detected in 100 SARS-CoV-2 RT-qPCR test result's positive patients (50 SARS-CoV-2 Delta, 50 Omicron BA.1 variants) and 100 SARS-CoV-2 RT-qPCR test result's negative samples as a control group. **Results:** The expression levels of *TASIR2* and *TASIR3*, genes were significantly decreased in COVID-19 patients who were infected with Delta variants compared to control group as well as in COVID-19 patients who were infected with Omicron BA.1 variants ($p < 0.05$). Also, *TAS2R38* gene expression level significantly lower when compared to control group ($p < 0.05$). Moreover, we found that *TASIR2* gene expression was positively correlated with *TASIR3*, and *TAS2R38* ($r = 0.655$, $p = 0.001$, $r = 0.301$ $p = 0.025$, respectively). **Conclusion:** We observed that *TASIR2*, *TASIR3* and *TAS2R38* gene expression were decreased in Delta variant which is the most disruptive and symptomatic variant caused of hospitalizations, and deaths compared to other variants. The results of our study provides a clue for loss of taste in especially in Delta variant may be cause of *ACE2* is expressed in the taste buds then the replication of SARS-CoV-2 in infected gustatory cells in the taste bud can generate inflammation and eventually destroy the cells. This direct cell damage may cause malfunction of the gustatory system.

Complex Traits Posters - Wednesday

PB1259. Alzheimer Disease plasma biomarker pTau-181 in individuals of diverse admixed ancestral backgrounds

Authors:

A. Griswold^{1,2}, F. RAJABLI¹, T. Gu¹, C. Garcia-Serje¹, J. Arvizu¹, C. Golightly¹, P. Whitehead¹, K. Hamilton-Nelson¹, L. Adams¹, M. Contreras¹, J. Sanchez¹, S. Tejada¹, P. Mena¹, T. Starks³, M. Cornejo-Olivas⁴, M. Illanes-Manrique⁴, C. Silva⁵, W. Bush^{6,7}, M. Cuccaro^{1,2}, J. Vance^{1,2}, B. Feliciano-Astacio⁵, G. Byrd³, G. Beecham^{1,2}, J. Haines^{6,7}, M. Pericak-Vance^{1,2}; ¹John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL, ²Dr. John T Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami, Miami, FL, ³Maya Angelou Ctr. for Hlth.Equity, Wake Forest Univ., Winston-Salem, NC, ⁴Neurogenetics Res. Ctr., Inst. Natl. de Ciencias Neurologicas, Lima, Peru, ⁵Dept. of Internal Med., Univ. Central Del Caribe, Bayamon, PR, ⁶Dept. of Population & Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, ⁷Cleveland Inst. for Computational Biology, Cleveland, OH

Abstract Body:

Background. Plasma proteins, including phosphorylated threonine-181 of Tau (pTau181) can be used as readily available and inexpensive biomarkers for differential diagnosis and preclinical detection of Alzheimer disease (AD). However, most measurements of these analytes are from individuals of non-Hispanic, European ancestry. Given differences in AD risk depending on both genetics and environmental factors, generalizability of these previous findings is not assured in diverse ancestry individuals. This study evaluates the utility of plasma pTau181 in discriminating clinically diagnosed AD from cognitively intact, age-matched controls in genetically admixed cohorts. **Methods.** We measured plasma pTau181 with Simoa chemistry with the pTau181 AdvantageV2 assay on the Quanterix HD-X. Our cohorts consisted of 642 African Americans (162 AD and 480 cognitively intact controls), 906 Puerto Ricans (385 AD and 521 controls), 149 Peruvians (49 AD and 100 controls), 60 Cubans (26 AD and 34 controls), 246 individuals of non-Hispanic, European ancestry (22 AD and 224 controls), and 58 autopsy confirmed AD cases of European ancestry with plasma isolated from EDTA blood tubes. Samples were randomized, measurements performed in duplicate, and non-parametric Kruskal-Wallis tests used to detect differences in biomarker concentrations between cases and controls overall and in each cohort. Receiver operating characteristic (ROC) area under the curve (AUC) were generated to test the predictive ability of pTau181 levels in each cohort. **Results.** Overall pTau levels in AD cases was higher than controls ($2.51 \pm 1.84 \text{ pg/mL}$ vs $1.12 \pm 3.24 \text{ pg/mL}$, $p_{\text{corr}} = 1.59 \times 10^{-76}$) and this was observed in all cohorts: African Americans ($2.30 \pm 1.14 \text{ pg/mL}$ vs $1.15 \pm 2.99 \text{ pg/mL}$, $p_{\text{corr}} = 2.0 \times 10^{-27}$); Puerto Ricans ($2.33 \pm 1.82 \text{ pg/mL}$ vs $1.44 \pm 1.21 \text{ pg/mL}$, $p_{\text{corr}} = 8.2 \times 10^{-32}$); Peruvians ($2.63 \pm 1.64 \text{ pg/mL}$ vs $2.13 \pm 1.42 \text{ pg/mL}$, $p_{\text{corr}} = 0.02$); Cubans ($2.09 \pm 1.16 \text{ pg/mL}$ vs $1.35 \pm 0.67 \text{ pg/mL}$, $p_{\text{corr}} = 0.02$); and European ancestry ($2.40 \text{ pg/mL} \pm 0.78 \text{ pg/mL}$ vs $1.54 \text{ pg/mL} \pm 1.44 \text{ pg/mL}$, $p_{\text{corr}} = 0.02$). ROC analyses showed that European and African Americans had higher predictive value (AUC = 0.78) than in Hispanic cohorts (AUC = 0.73 in Puerto Ricans, 0.70 in Cubans, and 0.62 in Peruvians). **Conclusion.** This study suggests pTau181 as an AD biomarker is generalizable across ancestries, though the predictive value may differ depending on background. Ultimately, combining genomic and biomarker data, including pTau181 and other AD related plasma biomarkers such as A β 40 and A β 42, from diverse individuals will increase understanding of genetic risk and refine clinical diagnoses in individuals of diverse ancestries.

Complex Traits Posters - Thursday

PB1260. Alzheimer's disease genetic risk and plasma biomarker profiles in the Midwestern Amish

Authors:

D. Dorfsman^{1,2}, M. B. Prough¹, L. J. Caywood¹, J. E. Clouse¹, S. D. Herington¹, S. H. Slifer¹, L. D. Adams¹, R. A. Laux^{3,4}, Y. E. Song^{3,4}, A. Lynn^{3,4}, M. Fuzzell^{3,4}, S. L. Fuzzell^{3,4}, S. D. Hochstetler^{3,4}, K. Miskimen^{3,4}, L. R. Main^{4,5}, M. D. Osterman^{3,4}, P. Ogrocki⁶, A. J. Lerner⁵, J. M. Vance², M. L. Cuccaro², J. L. Haines^{3,4}, M. A. Pericak-Vance^{1,2}, W. K. Scott^{1,2}; ¹John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, ²Dr. John T. Macdonald Fndn. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, ³Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, ⁴Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, ⁵Case Western Reserve Univ. Sch. of Med., Cleveland, OH, ⁶Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH

Abstract Body:

At least two-thirds of dementia cases are attributable to Alzheimer's disease (AD), a neurodegenerative disorder characterized by progressive loss of brain integrity. AD is complex, meaning the risk of occurrence depends on both inherited and environmental exposures. AD research has predominantly focused on identifying genetic variation that elevates risk. The discovery of risk-decreasing variants suggests that AD variability is also explained by genetic factors that reduce risk, highlighting an alternate strategy that prioritizes the discovery of protective variation. Studies of the aged and unimpaired have found genetic + AD pathology signatures to be diverse. The aim of this study is to explore the combined utility of whole-genome-sequence (WGS) derived AD genetic risk scores (GRS) and plasma-biomarker measures of AD pathology, to better understand how they differ within the cognitively impaired (CI) and unimpaired (CU). The study sample is drawn from the Midwestern Amish population, an endogamous population isolate with a shared lifestyle, resulting in increased genetic homogeneity with few environmental confounders. Variant calls from 1,055 WGS were obtained via the Illumina sequencing platform using GATK best practices. 312 clinically diagnosed (140 CI, 172 CU) subjects aged ≥ 75 have plasma measurements of pTau181 and total Tau. GRS were calculated using genome-wide effect sizes, including APOE, reported in Kunkle et al. (2019). By multivariate logistic regression, the best-fitting model using significant univariate results confirmed that increased GRS ($p = 0.0015$, OR = 1.83), pTau181 ($p = 0.003$, OR = 1.51), and age ($p = 0.018$, OR = 1.06) were significantly associated with cognitive impairment (CI). Further examining the CI group using multiple linear regression we find that the increasing GRS is associated with increased pTau181 ($p = 0.045$, $\beta = 0.22$), adjusting for age. Among the unimpaired, we did not observe an association between GRS and pTau181 ($p = 0.53$, $\beta = 0.015$), adjusting for age. These results indicate that there might be distinct genetic + biomarker signatures among CI and CU groups. While elevated genetic risk is associated with increased AD biomarker in the CI, we do not see the same trend in the CU. This may be explained by differences in the GRS composition that promote pathology, or by the presence of other genetic factors that actively mitigate or resist pathology. These findings will inform our efforts to characterize subgroups defined by unique genetic risk + AD pathological signatures to aid the discovery of genetic variation associated with preserved cognitive function.

Complex Traits Posters - Wednesday

PB1261*. Alzheimer's disease pathology drives distinct homeostatic and active microglia phenotypes revealing clues to early pathogenic transcriptional switches

Authors:

S. Jayadev¹, K. Prater², K. Green³, W. Sun⁴, S. Mamde¹, A. Sotelo¹, C. Smith³, K. Chiou⁵, I. Heath⁶, S. Rose¹, C. D. Keene¹, R. Kwon¹, N. Snyder-Mackler⁵, E. Blue¹, B. Logsdon⁷, J. Young¹, G. A. Garden⁸; ¹Univ. of Washington, Seattle, WA, ²Univ. of Washington Neurology, Seattle, WA, ³Univ. of Washington Neurology, Seattle, WA, ⁴Fred Hutchinson Cancer Res. Ctr., Seattle, WA, ⁵Arizona State Univ., Phoenix, AZ, ⁶SAGE, Seattle, WA, ⁷Cajal NeuroSci., Seattle, WA, ⁸Univ. of North Carolina, Raleigh, NC

Abstract Body:

Microglia contribute to Alzheimer Disease (AD) progression and mediate a significant portion of AD genetic risk. Single cell transcriptomics have demonstrated that microglia adopt a spectrum of functional states. However, identifying human microglia AD molecular profiles to elucidate disease mechanisms has been limited by small numbers of microglia relative to other central nervous system cells. We report a novel approach that vastly enriched the population of microglia that can be studied for single-nucleus RNAseq (snRNAseq). Using fluorescence activated nuclei sorting followed by snRNAseq we analyzed more than 127,000 microglia from 12 AD and 10 control autopsy brain cortex resulting on average 5790 microglia per individual and 10 microglia subtypes. Beyond established subtypes such as “inflammatory” or “senescent”, the data revealed three distinct subpopulations characterized by enrichment for endolysosomal pathways with unique immunometabolic profiles. One of these three is previously unrecognized and is increased in AD cases. This endolysosomal subtype is unique in the increased expression of cytosolic DNA/RNA recognition molecules, interferon response factors (IRFs) and AD GWAS genes. We provide immunohistochemical evidence for microglia with this phenotype including accumulation of cytoplasmic DNA and aberrant lysosomal vesicles. Single-Cell Regulatory Network Inference and Clustering analysis demonstrated cluster enrichment of specific transcriptional networks within microglia subtypes and further identified apical transcription factor networks that could act as molecular switches that specify functional heterogeneity. To study the dynamics of microglia transcriptomic transitions between states, we employed the Monocle trajectory inference method. We map the progression of “homeostatic” microglia through several phenotypes and end-states. We analyzed the homeostatic microglia population and find a uniquely AD subtype differentially enriched for cellular motility and calcium regulation. Taken together our data nominate candidate genes and pathways that signify early drivers of microglia reactivity in AD. We report that the interferon signature associated with AD may relate to microglial accumulation of cytoplasmic nucleic acids and endolysosomal dysfunction. Our study demonstrates the value of deeply profiling microglia populations in each sample to allow for identification of novel and AD specific microglial subtypes including those that likely reflect early disease microglial changes.

Complex Traits Posters - Thursday

PB1262. An Emirati cohort of autism spectrum disorders.

Authors:

N. Nassir¹, A. Albanna^{1,2}, A. Abou Tayoun^{2,1}, A. Ahmed¹, M. Verano¹, N. Karuvantevida³, C. Walsh⁴, M. Uddin⁵; ¹Mohammed bin Rashid Univ. of medicine and Hlth. Sci.s, Dubai, United Arab Emirates, ²Al Jalila Children's Specialty Hosp., Dubai, United Arab Emirates, ³MBR Univ., DHCC, United Arab Emirates, ⁴Children's Hosp Boston, Boston, MA, ⁵Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates

Abstract Body:

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder that encompasses clinical (characterized by social impairment and repetitive behavior) and genetic heterogeneity, including rare *de novo* or inherited mutations or chromosomal rearrangements. Whole exome sequencing has emerged as a fundamental tool in the detection of susceptibility variants and causative genes underlying genetic causes of diseases, but identification of all causative variants remains elusive. We aim to identify clinically relevant variants in the ASD cohort from an ethnically homogenous population of the United Arab Emirates. We have conducted short read based whole-exome sequencing in 29 parent-offspring trio samples from the Emirati population. The analysis pipeline consists of genome alignment using BWA-MEM, Genome Analysis Toolkit GlobalHaplotyper (GATK), and annotation using Annovar. The variants were classified by the GenomeArc Analytics platform following the ACMG guidelines. We identified 121 *de novo* variants and observed that VOUS were frequently observed, in about 96% of our cases. We detected 4 likely pathogenic rare *de novo* variants impacting genes, *GLI2*, *LZTR1*, *DMWD*, and *CAPZA2*, which were validated using Sanger sequencing. We also identified novel *de novo* variants in *BRSK1* and *MAST3* genes. *BRSK1* is a part of brain-specific kinase family of genes, of which *BRSK2* is previously associated with neurodevelopmental disorders. *MAST3* is a part of microtubule associated serine-threonine kinase family of genes, whereas *MAST1* is conclusively associated with neurological diseases. The analysis of *de novo* variant containing genes showed enrichment in pathways such as the development of spinal cord and neuronal differentiation, suggesting an important role in ASD pathogenesis. Expressed early in brain development, both *BRSK1* and *MAST3* are expressed in neurons and oligodendrocytes, whereas *GLI2* is expressed in astrocytes. Our study supports the utility of whole-exome sequencing in the identification of disease-causing variants and novel ASD candidate genes, which aid in facilitating precision genetic diagnosis and its integration into novel therapeutic approaches and management of ASD patients.

Complex Traits Posters - Wednesday

PB1263. An integrated genomic and phenotypic analysis of sleep traits in autism spectrum disorder

Authors:

J. Weissenkampen¹, A. Ghorai¹, J. Zhang¹, D. Rader², R. Sebro³, T. Brodtkin¹, L. Almasy⁴, M. Bucan¹; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Univ of Pennsylvania, Philadelphia, PA, ³Mayo Clinic Florida, Jacksonville Beach, FL, ⁴Children`s Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Autism spectrum disorders (ASD) are highly heritable, yet quite heterogeneous, neurodevelopmental disorders characterized by impaired social interactions and stereotyped behavior. Variability is typical within cognitive, executive, emotional, and sleep-related phenotypes. Sleep disturbances are common in individuals with ASD and may share an underlying neurological basis with ASD. We recruited ASD individuals and their family members (370 participants) and assessed: quantitative ASD-related traits (using questionnaires), three consecutive weeks of actimetry data, and whole genome sequence (WGS) data. Forty-eight sleep, wake, and activity measures were derived from actimetry data using two different algorithms. Five-fold cross validation least absolute shrinkage and selection operator (LASSO) identified measures associated with ASD status in the training set (area under the curve (AUC): 0.84), which were then further validated in the test set (AUC: 0.78). In addition to sleep traits, LASSO identified several physical activity traits associated with ASD. The link between physical activity and ASD traits was further confirmed through polygenic risk score (PRS) in the Simons Foundation Powering Autism Research (SPARK) collection (~3,200 participants). Conditional logistic regressions stratified by family show strong correlations between ASD and walking PRS ($p = 0.022$) and most active 10 hour period activity level PRS ($p = 0.020$). PRS deciles for walking were also significantly correlated with reported sleep problems ($p < 0.002$). Significant genetic correlation was observed between walking and ASD ($p = 0.04$). An unbiased, machine learning approach identified not only sleep traits, but physical activity traits associated with ASD status in our sample and the SPARK collection. These traits may be studied further to investigate common neurological correlates between them and ASD. Furthermore, interventions targeting these traits may also potentially alleviate some sleep disturbances in autistic individuals.

Complex Traits Posters - Thursday

PB1264. An unprecedented level of complexity in the schizophrenia-associated 3q29 region of the human genome with unique segments that increase the risk for non-allelic homologous recombination.

Authors:

F. Yilmaz¹, U. GURUSAMY², T. Mosley³, Y. Mostovoy⁴, R. Purcell³, T. Shaikh⁵, M. E. Zwick⁶, P-Y. Kwok⁴, C. Lee¹, J. G. Mulle⁷; ¹The Jackson Lab. for Genomic Med., Farmington, CT, ²Univ. of California, San Francisco, San Francisco, CA, ³Emory Univ., Atlanta, GA, ⁴Cardiovascular Res. Inst. and Inst. for Human Genetics, Univ. of California San Francisco Sch. of Med., San Francisco, CA, ⁵Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ⁶Dept. of Genetics, Rutgers Univ. – New Brunswick, New Brunswick, NJ, ⁷Dept. of Psychiatry, Robert Wood Johnson Med. Sch., Rutgers BioMed. and Hlth.Sci., Rutgers Univ., New Brunswick, NJ

Abstract Body:

Structural variations (SVs) contribute to genomic variation within the general population, but can also lead to genomic disorders (e.g., 3q29 deletion syndrome) by disrupting and changing the copy number of dosage-sensitive genes. Non-allelic homologous recombination (NAHR) between highly identical paralogous copies of segmental duplications (SDs) in the 3q29 region of the human genome can cause a ~1.6 million base pairs (Mbp) deletion or duplication and results in 3q29 deletion or duplication syndrome. Patients with the 3q29 deletion or duplication syndrome have neurodevelopmental and psychiatric phenotypes. However, the risk factors and underlying structures that contribute to NAHR at this locus are not well understood. To characterize the complex 3q29 region, we used long-read sequencing and optical mapping techniques. We resolved the haplotypes for the 3q29 region in 16 probands with 3q29 deletion syndrome, two probands with 3q29 duplication syndrome, and 161 individuals from 26 populations, including African, American, European, East Asian, and South Asian, from the 1000 Genomes Project. Within these sample cohorts, we identified a total of thirty-six distinct 3q29 haplotypes. Our results show that 3q29 haplotype prevalence varies significantly between populations. For instance, the H8 haplotype is observed only in individuals of African origin, while the H5 haplotype is enriched in East Asians (p-value = 5.7397e-06). Strikingly, the difference between the largest (859 kbp) and the shortest 3q29 haplotype (287 kbp) is ~580 kbp. Optical mapping data analysis of the 3q29 Project patients enabled us to resolve breakpoints in 89% of probands with 3q29 deletion or duplication syndrome, localizing these for the first time to a 5 kbp segment within complex SDs, which may indicate increased risk for NAHR, and long-read sequencing data helped to characterize breakpoints at the sequence level and refine the breakpoint junction to a 373 bp interval. Furthermore, breakpoint analysis allowed us to categorize breakpoints into three deletion and two duplication classes. We also identified large inversions ranging in size from 2.03 Mbp to 2.13 Mbp in three individuals, but contrary to previous studies of other genomic disorder loci, we did not observe these in the parental chromosome of origin for the 3q29 deletion. The results demonstrate extraordinary variation in the human 3q29 region among individuals from diverse populations. Pinpointing the breakpoints to a 5 kbp putative risk segment within SDs informs future in vitro functional studies, and helps to elucidate the underlying structures and variations that may play a crucial role in disease etiology.

Complex Traits Posters - Wednesday

PB1265. Analysis of 3,273 undiagnosed neurodevelopmental disorder trios reveals over-transmission of polygenic risk and a female protective effect

Authors:

E. Wigdor¹, Q. Huang¹, V. Chundru¹, K. E. Samocha^{2,3}, M. E. Hurles⁴, H. C. Martin¹; ¹Wellcome Sanger Inst., Hinxton, United Kingdom, ²Massachusetts Gen. Hosp., Boston, MA, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Wellcome Sanger Inst., Cambridge, United Kingdom

Abstract Body:

Work in the Deciphering Developmental Disorders (DDD) study revealed an excess of rare, inherited, likely damaging coding variants (RVs) among probands without a genetic diagnosis compared to controls. Moreover, these variants are over-transmitted to probands from unaffected parents, suggesting they are pathogenic but incompletely penetrant. Using the two largest datasets of undiagnosed neurodevelopmental disorders (NDDs), we tested whether polygenic scores (PGS) modify penetrance of RVs. Further, we explored the relationship between sex and diagnostic status with NDD-associated PGS within families.

In a meta-analysis of 3,273 European-ancestry trios from DDD and Genomics England (GEL), we tested whether unaffected parents over-transmitted NDD-associated PGS to undiagnosed NDD probands. We found nominally significant over-transmission of the DDD-derived NDD PGS in 1,573 GEL probands (SD=0.05, $p=0.02$). In the meta-analysis, probands over-inherited PGS for schizophrenia (SCZ; SD=0.07, $p=2.9 \times 10^{-3}$), but transmission of educational attainment (EA) and cognitive performance (CP) PGS did not differ from expectation. However, unaffected parents (N=6,546) had lower PGS for EA (OR=0.73, $p=1.2 \times 10^{-128}$) and CP (OR=0.80, $p=5.2 \times 10^{-60}$) than controls (N=28,214), and higher PGS for NDDs (OR=1.1, $p=6.8 \times 10^{-8}$) and SCZ (OR=1.1, $p=1.2 \times 10^{-3}$). As expected, these differences were also seen between probands and controls ($p<2.2 \times 10^{-4}$), confirming common variants contributing to EA and CP play a role in NDD risk, but suggesting probands' and unaffected parents PGS do not necessarily differ for these traits.

To explore whether PGS modify the penetrance of RVs, we compared PGS between i) unaffected parents with and without RVs, and ii) RV-transmitting parents and their offspring. We did not find any significant differences.

We then investigated the relationship of sex and diagnostic status with PGS within families. Undiagnosed probands (N=1,203) had lower PGS for EA than diagnosed probands (N=3,273, OR=0.89, $p=8.5 \times 10^{-4}$), consistent with a liability threshold model. Undiagnosed male probands had nominally significantly lower PGS for NDDs than undiagnosed female probands (OR=0.87, $p=0.02$). Similarly, unaffected fathers had lower PGS for SCZ (OR=0.93, $p=0.01$) and NDDs (OR=0.91, $p=0.01$) than unaffected mothers. Both findings are consistent with a female protective effect for NDDs.

Our results confirm the role of polygenic background in NDDs but find no evidence it modifies penetrance of RVs in these disorders. Larger sample sizes, more powerful PGS and better delineation of which RVs confer risk is needed to better understand the genetic architecture of NDDs.

Complex Traits Posters - Thursday

PB1266. Analysis of adipose and liver expression profiles from dual-tissue transcriptomic cohort discovers 10 serum biomarker candidates for NAFLD.

Authors:

N. Darci-Maher¹, M. Alvarez¹, U. Arasu², S. Lee¹, D. Z. Pan¹, Z. Miao¹, S. Das¹, D. Kaminska^{1,3,4}, T. Örd², J. Yang⁵, Z. Li⁵, B. Han⁶, M. P. Mellody^{6,7}, R. Damoiseaux^{6,7,8,9}, J. N. Benhammou¹⁰, J. R. Pisegna¹¹, V. Männistö¹², J. S. Sinsheimer^{1,13,14}, M. U. Kaikkonen², J. Pihlajamäki^{3,15}, P. Pajukanta^{1,16,17}; ¹Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ²A. I. Virtanen Inst. for Molecular Sci., Univ. of Eastern Finland, Kuopio, Finland, ³Inst. of Publ. Hlth. and Clinical Nutrition, Univ. of Eastern Finland, Kuopio, Finland, ⁴Turku PET Ctr., Turku Univ. Hosp., Turku, Finland, ⁵Dept. of Med., Ctr. for Human Nutrition, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ⁶California NanoSystems Inst., Univ. of California, Los Angeles, Los Angeles, CA, ⁷Dept. of Bioengineering, Univ. of California, Los Angeles, Los Angeles, CA, ⁸Dept. of Molecular and Med. Pharmacology, Univ. of California, Los Angeles, Los Angeles, CA, ⁹Jonsson Comprehensive Cancer Ctr., UCLA, Los Angeles, CA, ¹⁰Vatche and Tamar Manoukian Div. of Digestive Diseases and Gastroenterology, Hepatology and Parenteral Nutrition, David Geffen Sch. of Med. at UCLA and VA Greater Los Angeles HCS, Los Angeles, CA, ¹¹Dept. of Med. and Human Genetics, Div. of Gastroenterology, Hepatology and Parenteral Nutrition, David Geffen Sch. of Med. at UCLA and VA Greater Los Angeles HCS, Los Angeles, CA, ¹²Dept. of Med., Univ. of Eastern Finland and Kuopio Univ. Hosp., Kuopio, Finland, ¹³Dept. of Biostatistics, UCLA Fielding Sch. of Publ. Hlth., Los Angeles, CA, ¹⁴Dept. of Computational Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ¹⁵Dept. of Med., Endocrinology and Clinical Nutrition, Kuopio Univ. Hosp., Kuopio, Finland, ¹⁶Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, ¹⁷Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract Body:

Non-alcoholic fatty liver disease (NAFLD) is a global health problem that remains grossly underdiagnosed due to the lack of proper diagnostic tools in primary care. NAFLD comprises a heterogeneous spectrum, from simple steatosis to severe endpoints such as liver fibrosis and non-alcoholic steatohepatitis (NASH). To address the current challenges in NAFLD diagnosis, we leverage a dual-tissue transcriptomic cohort of morbidly obese Finnish individuals with RNA-seq data available from both subcutaneous adipose tissue (n=262) and liver (n=267), coupled with gold-standard liver histology-based diagnosis of steatosis, fibrosis, and NASH. We aimed to identify serum biomarker candidates (SBCs) for NAFLD secreted from the adipose tissue because we hypothesize that obesity-induced inflammatory dysfunction of adipose tissue hampers the efficient storage of fat in subcutaneous fat depots, which initiates ectopic fat deposition into the liver in metabolically unhealthy obese individuals. We discover 947 genes that are differentially expressed (DE) in adipose tissue for steatosis, fibrosis, or NASH, and designate 645 of these as adipose aware DE genes by removing genes also DE in the liver. From these, we then select genes that 1) encode for proteins secreted to serum, 2) are highly expressed in adipose tissue, and 3) are expressed >10x higher in adipose tissue than in liver, resulting in 10 SBCs for NAFLD. Using best subsets regression, we find that the adipose expression of the key SBCs, coiled-coil domain containing 80 (*CCDC80*) and superoxide dismutase 3 (*SOD3*), explains a significant proportion of variance in fibrosis ($p=1.97 \times 10^{-3}$) and NASH ($p=3.50 \times 10^{-4}$) compared to all pairs of genes in genome-wide permutations. Noteworthy, *SOD3* has been implicated for NAFLD in mice. Next, we investigate the cellular functions of *CCDC80* and *SOD3* via siRNA knockdown in human preadipocytes differentiated to adipocytes, followed by RNA-seq at four adipogenesis timepoints. We find that *CCDC80* knockdown increases the expression of the master regulator of fatty acid biosynthesis,

ASHG 2022 Annual Meeting Poster Abstracts

sterol regulatory element binding transcription factor 1 (*SREBF1*), as well as the key triglyceride hydrolysis enzyme, lipoprotein lipase (*LPL*) (adjP<0.05). We also observe that *SOD3* knockdown decreases the expression of satiety adipokine, leptin (*LEP*) (adjP<0.05). Thus, *CCDC80* and *SOD3* may contribute to NAFLD pathogenesis by modulating key functions of healthy adipose tissue. Overall, our findings have a great translational potential to improve patient health, by discovering 10 new adipose aware SBCs for NAFLD that can be further tested in independent NAFLD cohorts and ultimately in primary care.

Complex Traits Posters - Wednesday

PB1267. Analysis of both shared and specific associations among various autoimmune disorders greatly helps the identification of the causal gene(s) and the functional mechanisms

Authors:

W. Yang, X. Dang, Y. Wang; Univ. of Hong Kong, Hong Kong, Hong Kong

Abstract Body:

GWAS has identified hundreds of associated loci for autoimmune diseases such as systemic lupus erythematosus (SLE). However, in many cases the causal gene(s) in an associated locus is unclear. Functional annotations and gene proximity information often fall short in defining the causal gene(s). In this study, we analyzed the detailed shared and specific association signals extracted from GWAS Catalog for various autoimmune diseases, including SLE, rheumatoid arthritis (RA), multiple sclerosis (MS), Sjogren's disease (SjS), primary biliary cirrhosis (PBC) and type I diabetes (T1D). We observed widespread independent association signals for both SLE and other autoimmune disorders in more than half of the loci known to have associations for at least one of the autoimmune disorders. We conclude that this is on the one hand not surprising, suggesting a variety of mechanisms in gene expression regulation, which in many cases are likely cell type-specific and context-dependent. On the other hand, this poses challenges to functional characterizations of associated loci. We observed both shared association signals among the autoimmune diseases and specific signals to one disease or a subset of diseases, sometimes within a single genomic region. We believe that further studies of the detailed mechanisms of the associations and the shared and specific signals among the autoimmune disorders will greatly facilitate our understanding of these diseases. We argue that these valuable details have been under-explored, and in many cases, they were merely used to emphasize the validity of an association. Using examples in loci such as ETS1, CD44, TRAF3, and IRF4, we demonstrate that the comprehensive analysis of association signals from various autoimmune diseases greatly facilitate the identification of the causal gene(s) and offer a venue to better understand the mechanisms of the associations and pathogenesis of the autoimmune diseases.

Complex Traits Posters - Thursday

PB1268. Analysis of MRI-derived spleen iron in the UK Biobank identifies genetic variation linked to iron homeostasis and hemolysis

Authors:

E. Sorokin¹, N. Basty², B. Whitcher², Y. Liu¹, J. D. Bell², R. L. Cohen¹, M. Cule¹, E. L. Thomas²; ¹Calico Life Sci. LLC, South San Francisco, CA, ²Univ. of Westminster, Res. Ctr. for Optimal Hlth., London, United Kingdom

Abstract Body:

The spleen plays a key role in iron homeostasis. It is the largest filter of the blood and performs iron reuptake from old or damaged red blood cells. Despite this role, spleen iron concentration has not been measured in a large, population-based cohort. In this study, we quantify spleen iron concentration in 41,764 participants of the UK Biobank using magnetic resonance imaging, and provide the first reference range for spleen iron in an unselected population. Spleen iron is only partially correlated with other measures of iron stores including serum and liver iron. Thus, spleen iron is a novel and highly tractable trait for genetic analysis. Through genome-wide association study of spleen iron, we identify associations between spleen iron and regulatory variation at two genes, *ANK1* and *SPTA1*, encoding structural components of red blood cells. Loss in function in either of these genes is known to result in a Mendelian red blood cell disorder, hereditary spherocytosis; by contrast, the common alleles found through GWAS are associated with increased expression of *SPTA1* and *ANK1*. Opposite to rare disease-causing alleles, the common alleles are also associated with protective effects red blood cell parameters including reticulocyte volume. As genetic modifiers, these common alleles may help explain the variable expressivity and penetrance of hereditary spherocytosis that has been observed clinically. After fine-mapping, our genetic study also identifies a novel association which co-localizes with a splicing quantitative trait locus for *MS4A7*, and we show this gene is abundantly expressed in the spleen and in macrophages. In summary, we combine deep learning and efficient image processing to enable the first large-scale measurement of spleen iron, and we characterize genetic factors linked to the lytic phase of the red blood cell life cycle, as well as iron reuptake by the spleen. Our work advances understanding of iron homeostasis in the human body and illustrates the power of modern imaging-based techniques to deeply phenotype non-invasively and thereby enable discovery research in large human cohorts.

Complex Traits Posters - Wednesday

PB1269. Analysis of nascent RNA reveals distinct gene signatures for acute inflammation in cytokine stimulated keratinocytes.

Authors:

M. Patrick, M. K. Sarkar, M. T. Paulsen, I. V. Narayanan, M. Ljungman, J. E. Gudjonsson, L. C. Tsoi; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Acute inflammation (i.e. the immediate immune response to injury or infection) is challenging to investigate, due to the complex, dynamic and time-sensitive processes involved. Conventional RNA sequencing only profiles total RNA at the steady state, which is subject to varying RNA synthesis and turnover rates, thereby insufficiently capturing the mechanisms involved in early proinflammatory response. To address these issues, we applied Bru-seq (an approach that directly measures nascent RNA transcription by bromouridine labeling) to stimulated (IL-17A+TNF or IL-13) and unstimulated immortalized keratinocytes (N/TERTs). We repeated our experiments using 30 and 120 minutes of cytokine stimulation, with three samples for each condition. 378 and 652 genes were differentially expressed ($FDR \leq 0.05$, $FC \geq 1.5$ or $FC \leq 2/3$) for IL-17A+TNF after 30 and 120 minutes of stimulation, respectively, while 15 and 131 genes were differentially expressed for IL-13, compared with unstimulated controls. Interestingly, although pathways such as TNF signaling (under IL-17A+TNF stimulation) were consistently enriched across each time period ($p = 1.4 \times 10^{-19}$, $OR = 10.28$ at 30 minutes and $p = 3.2 \times 10^{-31}$, $OR = 10.93$ at 120 minutes), only 79 of the 378 (21%) IL-17A+TNF genes and 2 of the 15 (13%) IL-13 genes from 30 minutes of stimulation were also differentially expressed at 120 minutes. Indeed, different genes from the TNF signaling pathway were upregulated at each time period (e.g. *IRF1* and *NFKB1* at 30 minutes, *ICAM1* and *TNFAIP2* at 120 minutes IL-17A+TNF stimulation), suggesting different components of the pathway may be subject to dynamic regulation. Specific pathways enriched at 30 minutes but not 120 minutes include IFN-g and IL-5 signaling for IL-13 stimulation, with REL/RELB protein-protein interaction and the KLF6 transcription factor for IL-17A+TNF. We also compared the upregulated genes from our Bru-seq experiments with those upregulated in acute atopic dermatitis from our previous steady state study and found substantial overlap, including *CISH* and *NOD2* for IL-13 stimulation, *IRF1* and *IL32* for IL-17A+TNF. Our results can provide insight into the specific mechanisms involved in the early inflammatory responses in keratinocytes, which will allow us to better understand the disease onset and acute inflammation of different inflammatory skin conditions.

Complex Traits Posters - Thursday

PB1270. Analysis of poison exons caused by splice donor variants in genes associated with developmental brain disorders

Authors:

A. Hare-Harris, K. Kelchner, A. Hunter; Bloomsburg Univ., Bloomsburg, PA

Abstract Body:

The role of nonsense mediated decay (NMD) facilitated by loss of function (LOF) genomic variants in developmental brain disorders (DBD) has been well established. Whole exome sequencing (WES) studies have helped to identify pathogenic LOF variants within the coding regions of >500 genes among individuals with DBD phenotypes; however, the role of NMD facilitated by intronic variants has been largely unexplored. Genomic variants that disrupt the conserved splice donor motif can lead to NMD due to the inclusion of an intronic sequence that introduces a premature stop codon to the resulting transcript. Variants that create these 'poison exons' (PEX) have been previously identified in SCN1A in individuals with Dravet Syndrome. However, a broad survey of poison exons in other DBD genes has yet to be conducted. This study identified pathogenic splice donor variants in DBD cases that do not appear in control individuals in gnomAD v3 using publically available data from ClinVar. We developed an automated pathogenicity prediction algorithm (PEX-DETEX) to identify splice donor variants that facilitate PEX from VCF datafiles. Variants were curated using RefSeq Select and MANE transcripts (hg38). To date, approximately 5000 splice donor variants across 512 DBD genes were analyzed with PEX appearing in 62% of these genes in ClinVar and 49% of these genes in gnomAD. These results show that ClinVar is significantly enriched for PEX compared to gnomAD (# PEX in ClinVar=1894; # PEX in gnomAD=968; $p < 0.00001$) and 58 genes were found to have PEX in ClinVar cases only. Of the 166 variants reported in ClinVar cases only, 15 variants were originally classified as variants of unknown significance (VOUS). This study provides increased evidence of pathogenicity for these variants. This is the first study of its kind to develop and utilize an automated algorithm to assess the clinical implications of poison exon variants in DBD phenotypes. The inclusion of this algorithm in clinical variant curation pipelines can help to elucidate the clinical manifestation of intronic VOUS and can be extended to the analysis of other genetic causes of poison exons.

Complex Traits Posters - Wednesday

PB1271. Analysis of Relationship between Metabolic Syndrome and Genome-Wide Association Study (GWAS) of Vitamin D in Koreans: Community-Based Cohort.

Authors:

H. Kim¹, Y. Lee², J. Lee¹, J. Seo²; ¹Dept. of Family Med., Veterans Med. Ctr., Korea (ROK), Seoul, Korea, Republic of, ²Veterans Med. Res. Inst., Korea (ROK), Seoul, Korea, Republic of

Abstract Body:

Purpose It is well known that vitamin D levels are reduced in people with metabolic syndrome. However, it remains unclear whether there is causal relationship or not. The purpose of this study was to evaluate the causal relationship between vitamin D levels and metabolic syndrome risk in Korean community-based population by assessing the links between vitamin D SNPs and occurrence of metabolic syndrome. **Methods** Study subjects were selected from the Korean Genome and Epidemiology Study, KoGES, which is a large prospective cohort study project to identify the interactions between chronic diseases and genetic/environmental factors in Korea. A linear regression analysis was initially performed to step forward to verifying the association vitamin D and metabolic syndrome in Korean community-based population. Then, vitamin D related SNPs from Genome Wide Association Study catalogue, GWAS, were extracted and selected overlapped vitamin D SNPs from Korean Chip. Then, Mendelian Randomization analysis, MR that has been widely applied to assess the causal relationship between exposure and outcome was applied to see the causal influence of vitamin D in people with metabolic syndrome. Subjects with or without metabolic syndrome from community-based population were designated for the outcome dataset, and SNPs associated with vitamin D were selected for the exposure dataset. **Results** In observational study, we also confirmed that vitamin D level was significantly lower about 2 times in population with metabolic syndrome than those without it. The 580 validate vitamin D SNPs were extracted from GWAS, and then 270 SNPs were selected to perform MR study. MR analysis was performed using inverse variance weighted method with vitamin D SNPs variants. MR analysis supported that decreased vitamin D was significantly associated with increased risk of metabolic syndrome. In conclusion, we showed a significant relationship between vitamin D and metabolic syndrome in the Korean population through MR analysis, which may provide a basis for future study of mechanism related metabolic syndrome.

Complex Traits Posters - Thursday

PB1272. *APOE* genotype and rare variants of APP processing complex in autopsy-confirmed rapidly progressive Alzheimer disease.

Authors:

P. Wang¹, A. Lynn¹, K. Miskimen¹, Y. E. Song¹, C. Kim^{2,3}, M. Cohen^{2,4}, B. S. Appleby^{2,3,4,5}, J. G. Safar^{2,3,6}, J. L. Haines^{1,7}; ¹Dept. of Population and Quantitative Hlth.Sci., Sch. of Med., Case Western Reserve Univ., Cleveland, OH, ²Dept. of Pathology, Case Western Reserve Univ., Cleveland, OH, ³Dept. of Neurology, Case Western Reserve Univ., Cleveland, OH, ⁴Natl. Prion Disease Pathology Surveillance Ctr., Case Western Reserve Univ., Cleveland, OH, ⁵Dept. of Psychiatry, Case Western Reserve Univ., Cleveland, OH, ⁶Dept. of NeuroSci.s, Case Western Reserve Univ., Cleveland, OH, ⁷Cleveland Inst. for Computational Biology, Cleveland, OH

Abstract Body:

Prion disease and Alzheimer disease (AD) are distinct disorders; however, recent data suggests that amyloid beta and misfolded tau protein aggregates have self-propagating ability within brain tissue, and such distinct prion-like strains are associated with accelerated disease progression in AD. The role of genetic factors in different strains of pathogenic proteins and progression rates of AD is largely unknown. We established a rapidly progressive AD (rpAD) cohort from individuals referred to the National Prion Disease Pathology Surveillance Center with rapidly progressive dementia mimicking prion disease but subsequently confirmed as AD after neuropathological examination. Medical records were reviewed for clinical data, coronal sections of human brain tissue were obtained at autopsy, and DNA was extracted from brain tissue. Coding regions of AD risk genes, including *APOE*, *PSEN1*, *PSEN2*, and *APP* were analyzed using an Illumina Global Screening Array. Among 190 individuals with autopsy-confirmed rpAD, 53.4% were males and 80% were of non-Hispanic European (NHE) ancestry. The average age of onset was 70.2 ± 11.5 years, and the median survival was 6 months (IQR: 2-20 months). In 153 cases with *APOE* genotypes, the allele frequency of *APOE* $\epsilon 2$ was 4.25%, $\epsilon 3$ was 63.40%, and $\epsilon 4$ was 32.35%. The frequency of *APOE4* is somewhat lower than in a typical NHE AD cohort, which is consistent with findings from prion centers in Japan and Europe. In the 158 sequenced rpAD cases, we found two rare missense variants (*G206A* and *E318G*) in the *PSEN1* gene, one rare synonymous variant (*p.Ala252=*) in the *PSEN2* gene, and five rare intronic variants (*rs62222412*, *rs147490386*, *rs71317448*, *rs7283060*, *rs117873188*) in the *APP* gene. In 53 rpAD cases presenting with an early onset of disease (≤ 65 years old), 18.9% (n=10) carry a rare variant. Overall, 26 individuals had rare variants in *PSEN1* (n=9), *PSEN2* (n=2), and *APP* (n=14), and one case had variants in both *PSEN1* and *APP*. Compared to individuals with no rare variants, rare variant carriers had a younger age of onset (66.7 ± 11.8 vs 70.9 ± 11.4 , $p = 0.09$), and a longer and more variable median survival (10 months [IQR: 4-42] vs 6 months [IQR: 2-15]). We found a higher frequency of rare variants in known AD risk genes. Both rare variants in *PSEN1* are associated with a younger age of AD onset, abnormally elevated total tau and phosphorylated-tau, and decreased $A\beta$ levels. There is not yet strong evidence for the pathogenicity of the rare variants in *PSEN2* and *APP*. Our preliminary findings start to shed light on possible genetic risks associated with variable rates of progression in AD and an interplay with different strains of misfolded amyloid beta and tau proteins.

Complex Traits Posters - Wednesday

PB1273*. Assessing T-Cell Receptor Diversity and Genetic Variation in Complex Phenotypes.

Authors:

D. Ercelen, N. A. Zaitlen, M. J. Thompson; UCLA, Los Angeles, CA

Abstract Body:

Hypervariable complementary-determining regions (CDRs) in the genome determine the structure of immunoglobulins and T-cell receptors (TCRs). The genetic diversity of TCRs is critical for maintaining and developing adaptive immunity to successfully fight off infections. Nonetheless, the extent to which an individual's immune repertoire is predetermined remains contested, largely due to limitations of available profiling tools. To gain more insight on the genetic components of immune diversity, we utilized IMREP, a software tool designed to detect CDR regions with off-target reads from RNA-seq. The novelty we introduce in our study is the usage of whole exome sequencing (WES) data from the UKBiobank consortium. We focus on CDR regions as well as low coverage areas in the entire genome to maximize potential targets. Confirming the validity of our approach, we observed relatively high concentrations of alpha and beta chains, in comparison to delta or gamma chains, of TCRs. We further analyze the phenotypes generated by IMREP by running genome-wide association studies (GWAS) to identify potential genetic markers of immune diversity. We examine effects of TCR diversity on other complex phenotypes such as cancer, immune disorders, diet, location, and profession through linear and logistic regressions. GWAS on our IMREP generated TCR phenotypes found two statistically significant loci inside the intronic regions of the genes *MEOX1* (rs1724438; $p=4.63 \times 10^{-26}$) and *PLEKHB2* (rs145650663; $p=6.22 \times 10^{-44}$). These two genes have previously been shown to be related to regulatory roles in T-cells and influenza vaccine efficacy, respectively. In our regressions with complex phenotypes, we found significant correlations between the number and alpha diversity of TCR alpha chains (TCRA) and complex phenotypes such as the usage of the drug Omeprazole ($p=5.83 \times 10^{-4}$). Following these results we aim to further extract mitochondrial and ribosomal DNA from our data set, as well as microbiome information, and run similar analyses. Our results not only validate the use of IMREP on large whole exome datasets, but also suggest that immune signatures, in addition to gene environment interactions, may be partly predetermined by genetics.

Complex Traits Posters - Thursday

PB1274. Association analysis of *LHCGR* gene polymorphisms with Polycystic Ovary Syndrome in Punjabi population.

Authors:

S. Singh¹, M. Kaur¹, R. Kaur², A. Kaur³; ¹Guru Nanak Dev Univ., Amritsar, India, ²Dept. of Human Genetics, Guru Nanak Dev Univ., Amritsar, India, ³Guru Nanak Dev Univ., AMRITSAR, India

Abstract Body:

Background: Polycystic ovary syndrome (PCOS) is one of the most common endocrinopathies characterized by multiple hormonal imbalances dominated by hyperandrogenism, menstrual irregularities and/or small cysts in one or both ovaries. It is a heterogeneous disorder with the involvement of multiple gene and environmental interactions. Some polymorphisms of luteinizing hormone choriogonadotrophin receptor (*LHCGR*) gene may predispose women to develop PCOS. This study was designed to investigate the association of *LHCGR* polymorphisms with PCOS in Punjabi population since there are very few studies that have tried to understand the underlying genetic mechanisms in PCOS in North Indian population. **Methodology:** A total of 719 women comprising of 419 PCOS cases and 300 regularly menstruating women participated in the present study. Genotyping of *LHCGR* polymorphisms (rs2293275 and rs4539842) was done using polymerase chain reaction-restriction fragment length polymorphism. Restriction digestion was carried out using *RsaI* and *PvuII* enzymes for rs2293275 and rs4539842, respectively. Statistical analysis was performed using SPSS (version21, IBM SPSS, NY, USA). **Results:** The *LHCGR* variants demonstrated no significant association with PCOS cases. The genotypic frequency of SNPs rs2293275 and rs4539842 were not significantly different between PCOS cases and controls ($p=0.389$ and $p=0.805$, respectively). BMI was significantly different in both groups ($p<0.00001$). There was a significant elevation in the levels of cholesterol ($p=0.00002$) and triglycerides ($p<0.00001$) in women with PCOS as compared to controls whereas HDL levels were significantly lower in PCOS subjects ($p=0.0001$). **Conclusion:** These results suggest that none of these two polymorphisms of *LHCGR* gene was significantly associated with risk of PCOS in Punjabi population and that might be due to racial background, PCOS phenotype and the diagnostic criteria used.

Complex Traits Posters - Wednesday

PB1275. Association analysis of structural variants in whole-genome sequencing of 150K UK Biobank participants

Authors:

J. Liu¹, V. Ashton¹, J. Davitte¹, P. Gormley², R. Scott³, A. Cortes³, Y. Lo¹; ¹GSK, Collegeville, PA, ²GSK, Cambridge, MA, ³GSK, Stevenage, United Kingdom

Abstract Body:

Structural variants (SVs, insertions, deletions, and duplications over 50bp) are known to play an important role in complex traits, though are not well captured by targeted sequencing and array-based genotyping methods. The UK Biobank whole-genome sequencing (WGS) data includes over 637,000 SVs identified in ~150,000 participants using pangenome graphs implemented in GraphTyper. Using these data, we will present results from genome-wide association analysis of SVs across over 300 manually derived disease endpoints and ~500 quantitative biomarker, anthropometric and imaging phenotypes of interest. Significantly associated SVs will be assessed for their predicted functional consequences and the extent to which they are independent of known GWAS signals. The results of these analyses have the potential to highlight new drug target opportunities. SVs with direct impact on functional elements provide additional variant-to-gene-to function evidence for existing target-indication pairs, enabling us to generate better therapeutic hypotheses. The inferences from SVs will demonstrate one of the benefits of WGS over traditional GWAS approaches.

Complex Traits Posters - Thursday

PB1276. Association between a type 2 diabetes polygenic risk score and type 2 diabetes-related risk factors and complications

Authors:

B. Guo^{1,2}, Y. Cai², R. A. Smit³, L. Stalbow³, Q. Sun⁴, L. Raffield⁴, S. Raghavan⁵, J. B. Meigs⁶, C. S. Carlson², S. Rich⁷, L. R. Wilkens⁸, L. L. Marchand⁸, D. O. Stram⁹, M. Graff⁴, C. Arehart⁵, C. Gignoux⁵, K. E. North⁴, S. Buyske¹⁰, R. J. F. Loos³, C. Haiman⁹, U. Peters^{2,1}, C. Kooperberg^{2,1}, H. M. Highland⁴, B. Darst², the Population Architecture using Genomics and Epidemiology (PAGE) study; ¹Univ. of Washington, Seattle, WA, ²Fred Hutchinson Cancer Ctr., Seattle, WA, ³Icahn Sch. of Med. at Mount Sinai, New York, NY, ⁴Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁵Univ. of Colorado, Anschutz Med. Campus, Aurora, CO, ⁶Massachusetts Gen. Hosp., Boston, MA, ⁷Univ. of Virginia, Charlottesville, VA, ⁸Univ. of Hawai'i Cancer Ctr., Honolulu, HI, ⁹Univ. of Southern California, Los Angeles, CA, ¹⁰Rutgers Univ., Piscataway, NJ

Abstract Body:

It is estimated that >13% of the US population will have diabetes by 2030. With increasing GWAS sample sizes and advances in the development of polygenic risk scores (PRS), we can estimate genetic risk of type 2 diabetes (T2D), which may inform preventive strategies. However, it is unclear whether T2D PRS holds prognostic value for T2D risk factors or major complications, particularly across multi-ancestry populations. We investigated these questions using a PRS of 582 genome-wide significant variants identified in a previously conducted large multi-ancestry T2D GWAS of 228,499 cases and 1,178,783 controls. Regression models were used to evaluate the association between the T2D PRS and T2D-related risk factors and complications (fasting glucose and insulin, HOMA-IR and HbA1c, cardiovascular traits, and lipid traits) in an independent sample of 14,823 T2D cases and 25,278 controls from African American (AA), East Asian (EA), Hispanic (HA), and Native Hawaiian (NH) populations from the Population Architecture using Genomics and Epidemiology (PAGE) Study, with models run separately for each population and stratified by T2D status. We found improvements in the percent phenotypic variance explained (R^2) for all continuous traits and small to no improvements in the area under the curve (AUC) for the two binary traits (incident heart attack and stroke) evaluated when the T2D PRS was added to a base model of age, sex, body mass index, and the first ten principal components of ancestry. The highest R^2 improvements across continuous traits were consistently observed in HA (ranging from 0.02% to 0.99%) or EA (ranging from 0.03% to 2.89%) and typically the lowest in AA (ranging from 0% to 0.37%) among T2D controls. Relative to R^2 estimated in the base model (7.06%, 10.23%, 9.97%, and 4.18% in AA, EA, HA, and NH, respectively), the trait with the largest R^2 improvement was fasting glucose (7.41%, 10.93%, 10.96%, and 4.84% in AA, EA, HA, and NH, respectively). Compared to the R^2 estimated in the base model (1.57%, 24.89%, 18.64%, and 9.05%, in AA, EA, HA, and NH, respectively), the trait with the lowest R^2 improvement was HOMA-IR (1.58%, 24.87%, 18.66%, and 9.10% in AA, EA, HA, and NH, respectively). We are conducting additional analyses to evaluate the association between the T2D PRS and other T2D comorbidities, including retinopathy, neuropathy, kidney and liver function, and cancer. These results suggest that the T2D PRS may help identify individuals across diverse populations with a greater risk of developing T2D-related risk factors and complications, which could have important clinical and prognostic implications.

Complex Traits Posters - Wednesday

PB1277. Association between Blood Pressure-related Polygenic Risk Scores and Hypertension among White and Black Individuals Across the Life Course

Authors:

X. Sun; Tulane Univ., New Orleans, LA

Abstract Body:

Genetic information may help to identify individuals at increased risk for hypertension in early life, prior to the manifestation of elevated blood pressure (BP) values. We examined 369 Black and 832 white Bogalusa Heart Study (BHS) participants recruited in childhood and followed approximately 37 years. Multi-ancestry genome-wide polygenic risk scores (PRSs) for systolic BP (SBP), diastolic BP (DBP), and hypertension were tested for association with incident hypertension and stage-2 hypertension using Cox proportional hazards models. Race-stratified analyses adjusted for baseline age, age², sex, body mass index, genetic principal components, and BP. In Black participants, each standard deviation increases in SBP and DBP PRS (respectively) conferred 38% (P=0.009) and 22% (P=0.02) increased risk of hypertension and 74% (P<0.001) and 50% (P<0.001) increased risk of stage 2 hypertension, while no association was observed with the hypertension PRS. In whites, each standard deviation increases in SBP, DBP, and hypertension PRS (respectively) conferred 24% (P<0.05), 29% (P=0.01) and 25% (P<0.001) increased risk of hypertension, and 27% (P=0.08), 29% (0.01), and 42% (P<0.001) increased risk of stage 2 hypertension. Addition of BP PRSs to the covariable-only models generally improved C-statistics (P<0.05). Multi-ancestry BP PRSs demonstrate the utility of genomic information in early life prediction of hypertension.

Complex Traits Posters - Thursday

PB1278. Association between genomic runs of homozygosity and complex traits in the Taiwan Biobank

Authors:

Y-C. Feng¹, Y-J. Lai¹, T-T. Chen², C-Y. Chen³, T. Ge⁴, H. Huang⁵, Y-F. Lin⁶; ¹Natl. Taiwan Univ., Taipei, Taiwan, ²Natl. Hlth.Res. Inst.s, Zhunan Town, Miaoli County, Taiwan, ³Biogen, Cambridge, MA, ⁴Massachusetts Gen. Hosp., Boston, MA, ⁵MGH, Boston, MA, ⁶Natl. Hlth.Res. Inst.s, Taiwan, Zhunan Town, Miaoli County, Taiwan

Abstract Body:

Inbreeding results in elevated levels of autozygosity (homozygosity-by-descent) widely known to increase the risk of Mendelian conditions. Using high-density genotype data, earlier studies have shown that contiguous stretches of homozygous segments—runs of homozygosity (ROH)—are also common in modern outbred human populations with an impact on complex traits and diseases. The emergence of global biobanks has now made it possible to investigate the phenotypic spectrum associated with genome-wide homozygosity at the population level, especially beyond the European ancestry. Here, we leveraged the population-based Taiwan biobank (TWB) to study the distribution of ROH and its relationship with 46 health-related quantitative traits (anthropometric, biometric, and reproductive) among 93,598 unrelated individuals of Han-Chinese descent. ROHs were called using PLINK based on published recommendations (≥ 65 consecutive homozygous SNPs spanning at least 1,000 Kb) in the LD-pruned imputed genetic dataset. The total length of ROHs and the fraction of the autosome in ROHs (FROH) were calculated for each individual. Using linear mixed-effects models, we estimated the effect of FROH on phenotype (standardized within each sex) adjusting for age, sex, and the first 20 principal components with batch treated as a random factor. To account for potential confounding by socioeconomic status (SES), we performed additional analysis including education and family income as SES covariates. The longest ROH detected in TWB individuals spanned 9.8Mb containing 1,788 homozygous SNPs. FROH ranged from 0-0.04 and averaged at 0.003. After Bonferroni correction, our results suggested FROH was significantly, negatively associated with inspiratory capacity ($\beta_{\text{FROH}} = -10.7$ in phenotypic SD, $P = 1.3 \times 10^{-4}$), height (-7.8SD , $P = 1.9 \times 10^{-4}$), and total cholesterol (-7.1SD , $P = 1.2 \times 10^{-4}$). The inverse association was observed at nominal significance ($P < 0.05$) for triglyceride, red blood cell count, forced vital capacity, and number of times pregnant. However, the top associations became non-significant and attenuated after adjusting for SES; notably, the effect of FROH on the number of times pregnant became slightly stronger. Overall, these findings are consistent with previously reported associations with increased autozygosity among individuals of distant relatedness and implicate possible directional selection for these traits. We will use whole-genome sequencing data to map recessive loci within ROH tracts and extend the analysis to disease phenotypes to understand the role of ROHs in the genetic architecture of complex traits.

Complex Traits Posters - Wednesday

PB1279. Association between polymorphisms in *CD32*, *CD36*, *CD40*, *CD54* and Malaria phenotypes among under-fives in Ibadan South-west Nigeria.

Authors:

T. Olajide¹, S. A. Ademola¹, O. J. Bamikole¹, M-D. B. Olufeagba¹, B. A. Adedeji², N. O. Bukoye¹, O. K. Amodu¹; ¹Univ. of Ibadan, Ibadan, Nigeria, ²Modibbo Adama Univ., Yola, Nigeria

Abstract Body:

Background: Malaria remains one of the most prevalent infectious diseases in the world. Africa and Nigeria account for 95% and 27% of the world's malaria prevalence respectively. The significant rise from the 2019 figures can be attributed to service destruction associated with the Covid-19 pandemic. Children under age 5 years and pregnant women constitute the greater proportion of malaria-associated mortalities in the world. Cell surface receptors and messengers are factors that play important roles in parasite clearance during malaria infection. Polymorphisms in *FCGR2A(CD32)*, *CD36*, *CD40LG*, and *ICAM-1(CD54)* have been associated with increased susceptibility to malaria. Methods: A total of 470 children between the ages of 6 and 59 months (247 males and 223 females) from a section of the Nigerian population were recruited into case and control groups based on WHO guidelines. The participants were examined for clinical parameters and genotyped for seven SNP variants *CD40LG* (rs1126535 and rs3092945), *FCGR2A* (rs1801274), *CD36* (rs3211938 and rs201346212) and *ICAM-1* (rs5498 and rs1799969). The results were analyzed for conformity with HWE and genotypes were tested for association with the phenotypes using logistic regression under different genetic models. Results: Packed cell volume, parasitemia, respiratory rates, and body temperature differed across the clinical groups. Genetic variants (*CD40LG* rs1126535) and (*CD36* rs3211938) deviated from HWE. Here we report that (*CD36* rs3211938) was a risk for uncomplicated malaria when two copies of the risk allele were inherited while it protects against severe malaria anemia when just a copy of the protective allele is inherited. Conclusions: *CD36* (rs3211938) is associated differentially with existing malaria phenotypes in our population. However, this study could not establish a genetic association between the other SNPs and malaria phenotypes.

Complex Traits Posters - Thursday

PB1280. Association of a genetic risk score with incident coronary heart disease in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) cohort.

Authors:

C. Hutten¹, R. A. Durazo-Arvizu¹, S. Wassertheil-Smoller², C. R. Isasi², J. Cai³, J. T. Unkart⁴, J. Sun¹, V. Persky¹, T. Sofer⁵, M. L. Daviglus¹, M. Argos¹; ¹Univ. of Illinois at Chicago, Chicago, IL, ²Einstein Coll. of Med., Bronx, NY, ³UNC at Chapel Hill, Chapel Hill, NC, ⁴SUNY Downstate, Brooklyn, NY, ⁵Harvard Med. Sch., Boston, MA

Abstract Body:

Background: Coronary heart disease (CHD) is a leading cause of death for the Hispanic/Latino population living in the United States (US). Genetic risk scores (GRS) have been used in European ancestry populations to identify those at higher risk for CHD events. **Objectives:** This study evaluates a GRS with incident CHD in a Hispanic/Latino cohort. **Methods:** The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a prospective longitudinal study that enrolled 16,415 Hispanic/Latino adults aged 18 to 74 years from four US communities between 2008-2011 with a follow-up visit in 2014-2017. We assessed the association between a 71-SNP GRS and incident CHD among 9,030 HCHS/SOL participants with complete data and consent to conduct genetic research. Incident CHD was defined as no self-reported physician diagnosis or evidence of electrocardiographic abnormalities indicative of CHD at baseline and self-reported CHD at Visit 2. CHD events and time to events (such as incident myocardial infarction (MI)) were adjudicated through 2015, with further adjudication in progress. We considered an unweighted (sum of risk alleles), European-weighted, and Latino-weighted GRS. European and Latino weights were derived from a prior publication employing this GRS (Ke et al. 2018). Cox proportional hazards models were used to derive hazard ratios and 95% CIs for the association between the GRS and incident CHD, adjusted for a priori confounders and survey weights. **Results:** Preliminary analysis shows a trend for greater than 50% increased risk of incident CHD for those in the highest quintile of unweighted GRS compared to the lower four quintiles combined (HR 1.57, 95% CI 0.91, 2.71). The trend for incident MI events was even higher for those in the top quintile of GRS compared to all others (HR 2.08, 95% CI 0.75, 5.76). Stratified analyses to explore heterogeneity by sex showed males had a 2.5-fold increased risk of incident CHD when comparing the highest quintile of the GRS to all other quintiles (HR 2.58, 95% CI 1.29, 5.14); females showed no increased risk (HR 0.86, 95% CI 0.40, 1.84). Furthermore, stratification by European genetic ancestry showed those with greater than 67% European ancestry had a 4-fold increased risk of incident CHD (HR 4.09, 95% CI 1.87, 8.94); lower proportions of European ancestry showed no increased risk. Results were comparable using the other GRS constructs. **Conclusion:** The results highlight the need for additional studies to evaluate and optimize risk prediction of CHD based on cumulative genetic risk in the Hispanic/Latino population. Adjudication of further CHD events may confirm the increased risk of incident CHD in those with high GRS.

Complex Traits Posters - Wednesday

PB1281. Association of African Ancestry with risk for fibroproliferative diseases is consistent with selection for a Th2 favored genome in African derived populations

Authors:

G. Hampton¹, J. Keaton², J. N. Hellwege³, K. Kiryluk⁴, N. Shang⁵, G. Hripacsak⁶, I. Stanaway⁷, S. Asgari⁸, R. Knevel⁹, D. Crosslin¹⁰, J. Denny¹¹, D. Roden¹², H. Hakonarson¹³, J. Harley¹⁴, G. P. Jarvik¹⁵, G. Hayes¹⁶, P. Sleiman¹⁷, C. Weng¹⁸, S. Raychaudhuri¹⁹, D. Edwards²⁰, T. Edwards²¹; ¹Vanderbilt Univ., Nashville, TN, ²Ctr. for Precision Hlth.Res. NHGRI, NIH; Div. of Epidemiology, Dept. of Med., VUMC, Nashville TN, Bethesda, MD, ³Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁴Div. of Nephrology, Dept. of Med., Columbia, New York, NY, ⁵Div. of Nephrology, Dept. of Med., Columbia Univ., New York, NY, ⁶Dept. of BioMed. Informatics, Columbia Univ., New York, NY, ⁷Kidney Res. Inst., Harborview Med. Ctr., Div. of Nephrology, Sch. of Med., Univ. of Washington, Seattle, WA, ⁸Icahn Sch. of Med. at Mount Sinai, New York, NY, ⁹Harvard Med. Sch., Boston, MA, ¹⁰BioMed. Informatics and Genomics, John W. Deming Dept. of Med., Tulane Univ. Sch. of Med., New Orleans, LA, ¹¹NIH, Kensington, MD, ¹²Professor of Med., Pharmacology, and BioMed. Informatics Senior Vice President for Personalized Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, ¹³Children S Hosp. of Philadelphia, Philadelphia, PA, ¹⁴US Dept. of VA, Cincinnati, OH, ¹⁵BioMed. Informatics and Genomics and Genomic Sci., Univ. of Washington Med. Ctr., Seattle, WA, ¹⁶Northwestern Univ., Chicago, IL, ¹⁷CHOP, Philadelphia, PA, ¹⁸Dept. of BioMed. Informatics, Columbia Univ., New York, NY, ¹⁹Harvard Med. Sch.; Broad Inst. of MIT, Boston; Cambridge, MA, ²⁰DRVE affiliation: Div. of Quantitative Sci., Dept. of Obstetrics and Gynecology, Dept. of BioMed. Informatics TLE: Div. of Epidemiology, Dept. of Med., Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN, ²¹Vanderbilt Genetics Inst., Div. of Epidemiology; Dept. of Med., Vanderbilt Univ. Med., Nashville, TN

Abstract Body:

Fibroproliferative diseases (FPDs) are characterized by excessive scarring and overgrowth of connective tissue resulting from overactive T-helper cell type 2 (Th2) fibrotic immune response. In contrast, **inflammatory diseases (IFDs)** are characterized by an overactive T-helper cell type 1 (Th1), a Th2 antagonist, immune response that causes excessive inflammation. Many FPDs are more prevalent in African ancestry (AA) individuals, while many IFDs are more common in European Ancestry (EA) individuals. The increased prevalence of FPDs in AA populations is hypothesized to be due to selection for alleles that intensify the Th2 immune response. However, certain Th1-driven IFDs (i.e., alopecia areata and multiple sclerosis) are more prevalent in AA individuals, contradicting a Th2-favored genome. Here we calculated global ancestry proportions from 1000 Genome Phase 3 global populations [West African (WAFR), East African (EAFR), Southern European (SEUR), and Northern European (NEUR)] using Admixture software and data from electronic health record biobanks, BioVU and eMERGE. We tested for association with eight FPDs and seven IFDs defined by phecodes. Logistic regression analyses were conducted on combined BioVU and eMERGE datasets, per phenotype, for self-identified non-Hispanic blacks and non-Hispanic whites with case/control status defined by FPD and IFD phecodes, with the primary exposure being ancestry proportions, adjusted for sex, age, and body mass index (BMI). Effect estimates are reported per 10% increase in genetically inferred ancestry proportions. We detected positive associations of AA with all FPDs, with the largest effects being observed for keloids (EAFR Odds Ratio [OR]=1.26, 95% confidence interval [CI]1.22, 1.33; WAFR OR=1.27, 95% CI 1.23, 1.35) and uterine fibroids (EAFR OR=1.29, 95% CI 1.26, 1.32; WAFR OR=1.33, 95% CI 1.29, 1.36). We also observed a negative association of AA with all IFDs, with the strongest effects for celiac disease (EAFR OR=0.47, 95% CI 0.36, 0.58; WAFR OR=0.47, 95% CI 0.36, 0.57). We detected negative associations for each FPD in EAs, with the strongest effects for uterine fibroids (SEUR OR=0.81, 95% CI 0.80, 0.82;

ASHG 2022 Annual Meeting Poster Abstracts

NEUR OR=0.74, 95% CI 0.71, 0.77) and glaucoma (SEUR OR=0.84, 95% CI 0.83, 0.85; NEUR OR=0.93, 95% CI 0.90,0.95). We also detected a positive association for EA with all IFDs, with the strongest effects for Celiac (SEUR OR=1.32, 95% CI 1.23, 1.42; NEUR OR=1.42 95% CI 1.28, 1.57)). This evidence supports the role of a Th2 favored genome that contributed to the increased frequency of FPDs while also demonstrating a potentially protective relationship between AA and IFD development.

Complex Traits Posters - Thursday

PB1282. Association of genetic predisposition and physical activity with risk of gestational diabetes mellitus in nulliparous women.

Authors:

P. Radivojac¹, K. A. Pagel², H. Chu¹, R. Ramola¹, R. F. Guerrero³, J. H. Chung⁴, S. Parry⁵, U. M. Reddy⁶, R. M. Silver⁷, J. G. Steller⁴, L. M. Yee⁸, R. J. Wapner⁹, M. W. Hahn¹⁰, S. Natarajan¹¹, D. M. Haas¹²; ¹Northeastern Univ., Boston, MA, ²Johns Hopkins Univ., Baltimore, MD, ³North Carolina State Univ., Raleigh, NC, ⁴Univ. of California, Irvine, CA, ⁵Univ. of Pennsylvania, Philadelphia, PA, ⁶Yale Univ., New Haven, CT, ⁷Univ. of Utah, Salt Lake City, UT, ⁸Northwestern Univ., Chicago, IL, ⁹Columbia Univ., New York, NY, ¹⁰Indiana Univ., Bloomington, IN, ¹¹Univ. of Texas, Dallas, TX, ¹²Indiana Univ., Indianapolis, IN

Abstract Body:

Introduction: Polygenic risk scores (PRS) for Type II Diabetes Mellitus (T2DM) can improve risk prediction for Gestational Diabetes Mellitus (GDM), yet the strength of the relationship between genetic and lifestyle risk factors has not been quantified. In this work, we assess the effects of PRS and physical activity on existing GDM risk models and identify patient subgroups who may receive the most benefits from receiving a PRS or activity intervention.

Methods: The nuMoM2b cohort was established to study individuals without previous pregnancy lasting 20 or more weeks (nulliparous) and to elucidate factors associated with adverse pregnancy outcomes. In this work, a sub-cohort of 3,533 participants with European ancestry was used for risk assessment and performance evaluation of GDM. The key factors under investigation include PRS and self-reported total physical activity in early pregnancy that was quantified as metabolic equivalent of tasks (METs). PRS were calculated for T2DM using contributions of 84 single nucleotide variants, weighted by their association in the DIAGRAM Consortium data. Prediction of the development of GDM was performed based upon clinical, genetic, and environmental variables collected in early pregnancy.

Results: A total of 3,533 women met inclusion criteria (mean age = 28.6 +/- 4.9 years), of which 132 (3.7%) received GDM diagnosis. In high-risk population subgroups (body mass index ≥ 25 or age ≥ 35), individuals with high PRS (top 25th percentile) or low activity (METs < 450) have significantly increased odds of GDM diagnosis. Participants with both high PRS and low activity have three times higher odds of GDM diagnosis than the population, while those with either low PRS or high activity do not have increased odds of GDM diagnosis.

Conclusions: In this cohort study, the addition of PRS resulted in stratified risk of GDM diagnosis among high-risk patient subgroups, suggesting the benefits of targeted PRS ascertainment to encourage early intervention.

Complex Traits Posters - Wednesday

PB1283. Association of Genetically-Predicted Placental Gene Expression and Diseases of the Cardiovascular System

Authors:

A. Pigg^{1,2}, J. Hellwege², E. Jasper², T. Edwards², D. Velez Edwards²; ¹Meharry Med. Coll., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Mother-to-child transfer of disease risk may happen in utero and influence the health of offspring throughout early life and adulthood. The placenta plays a vital role in the prenatal environment and is an important biological conduit for not only maintaining a healthy pregnancy but the health of offspring. The Developmental Origins of Health and Disease (DOHaD) hypothesis describes the implications of the prenatal environment on the offspring's risk of disease later in life like obesity, type II diabetes, psychiatric disorders, and cardiovascular disease. Several cardiovascular diseases run in families which indicates an inherited genetic risk by offspring. Gene expression in the placenta could play a role in the risk of the development of cardiovascular diseases throughout the course of life. Previous studies have described the genetic control of placental gene expression through expression quantitative trait loci (eQTL) studies. In this study, we evaluated the relationship between placental gene expression and cardiovascular diseases. We constructed genetically predicted gene expression (GPGE) models for 25,885 genetic variants associated with expression of 15,154 genes from reference placental eQTL data (Peng et al., 2018). Using these models, we investigated whether placental GPGE is associated with seven adult cardiovascular diseases (Sakaue, S. et al., 2021) using GWAS summary statistics. Predicted gene expression analyses were conducted using SPrediXcan software. We respectively identified five, two, seven, three, and one genes whose predicted expression was significantly associated ($p < 1.0 \times 10^{-5}$) with angina, cardiac valvular disease, myocardial infarction, peripheral arterial disease, and subarachnoid hemorrhage after correcting for multiple comparisons. When compared to predicted expression models from 48 adult tissues from GTEx v7, nine genes were uniquely associated in the placental models. Notable genes identified are TOMM40, CDKN2B, and MAN2A2, which were previously reported to be associated with LDL levels, BMI, type II diabetes, and coronary artery disease. These results suggest unique genetic mechanisms by which the placenta contributes to later-in-life cardiovascular disease beyond that observed through adult gene expression.

Complex Traits Posters - Wednesday

PB1285. Association of Structural Variants with Coronary Artery Disease

Authors:

K. Iyer¹, R. Guarischi-Sousa¹, S. Clarke¹, G. Jun², P. de Vries², T. Assimes¹; ¹Stanford Univ Sch Med., Palo Alto, CA, ²UTHlth.Sch. of Publ. Hlth., Houston, TX

Abstract Body:

Genome-wide association studies (GWAS) have identified several hundred susceptibility single nucleotide polymorphisms (SNPs) for coronary artery disease (CAD). Despite SNP-based GWAS revolutionizing our understanding of the genetics of CAD, the contribution of structural variants (SVs) to the risk of CAD remains largely unclear. We leveraged SVs detected from high-coverage whole-genome sequencing (WGS) data in a diverse group of participants from the NHLBI Trans-Omics for Precision Medicine (TOPMed) program. Using single variant tests, we performed an SV-based GWAS including $n=75,733$ SVs that passed QC filters and had a minor allele count ≥ 10 in 13,504 CAD cases and 45,078 controls. These tests were performed using GENESIS after adjusting for age, sex, self-identified race/ethnicity, study, sequencing center, and relatedness. Population structure was also adjusted for using the first 10 principal components derived from SV calls. We identified two statistically significant overlapping duplications near *RFPL4B* on chromosome 6q21 (DUP_6:112405701-112406900 and DUP_6:112405201-112406900 with $p=7.12E-10$ and $p=2.59E-9$, respectively). These biallelic SVs were observed to have a minor allele frequency (MAF) of 5.5% and 2.5% in our participants. Both duplications spanned over 1000 bp, and housed a candidate cis-Regulatory Element EH38E2495040 with a distal enhancer-like signature along with nine common intergenic SNPs from 1000 Genomes Phase3. These SNPs were found to be nominally associated with cardio-metabolic traits like weight circumference, body mass index (BMI), hypertension, and Type 2 diabetes (T2D) on the phenome-wide associations (PheWAS) catalog. One of the SNPs in this region - rs9387078 almost reached the PheWAS significance threshold for T2D with renal manifestations ($p=7.1E-5$) in White British participants of the UK Biobank. Our analysis also identified an association between CAD and a rare 487bp biallelic deletion (MAF: 0.08%) within *CSMD2* on 1p35.1 (DEL_1:34097323-34097810, $p=9.69E-7$), a gene previously implicated in several brain-related conditions. Two intronic SNPs - rs12031634 and rs9426003 located less than 20kb and 40kb from the deletion region were found to be significantly associated with BMI ($p=1E-10$; $p=3.2E-10$) among people of European ancestry. These two variants were found not to be in linkage disequilibrium with our deletion. Our results showcase the nature of the relationship between rare and common SVs among a highly genetically diverse population and suggest that structural variation may influence the risk of CAD through well-established cardiometabolic risk factors.

Complex Traits Posters - Thursday

PB1286. Associations between polygenic risk scores and gene expression identify core genes for bipolar disorder.

Authors:

S. Lapinska¹, T. Schwarz¹, T. Boltz², M. Bot³, R. Kahn⁴, M. Boks⁵, R. Ophoff⁶, B. Pasaniuc⁷; ¹Bioinformatics InterDept.al Program, Univ. of California Los Angeles, Los Angeles, CA, ²Dept. of Human Genetics, David Geffen Sch. of Med., Univ. of California Los Angeles, Los Angeles, CA, ³Ctr. for Neurobehavioral Genetics, Semel Inst. for NeuroSci. and Human Behavior, Univ. of California, Los Angeles, Los Angeles, CA, ⁴Dept. of Psychiatry, Icahn Sch. of Med., Mount Sinai, NY, ⁵Dept. of Psychiatry, Brain Ctr. Univ. Med. Ctr. Utrecht, Univ. Utrecht, Utrecht, Utrecht, Netherlands, ⁶Dept. of Computational Med., David Geffen Sch. of Med., Univ California Los Angeles, Los Angeles, CA, ⁷Dept. of Computational Med., David Geffen Sch. of Med., Univ. of California Los Angeles, Los Angeles, CA

Abstract Body:

Bipolar disorder (BP) is a complex, highly heritable mental illness. Approximately 20% of BP heritability can be explained by common SNP variants, while in twin studies, the overall heritability of BP has been estimated to be over 70%. Even though BP is highly heritable, only a fraction of the genetic loci that influence the risk of developing BP has been identified and the biological mechanisms that play a key role in the development of BP are mostly unknown. A recently proposed omnigenic model suggests that the heritability of most complex traits is heavily influenced by weak trans effects that converge onto a smaller set of trait-relevant genes that seem to be at the core of the molecular pathway. Given that numerous risk variants have downstream consequences in the same pathway, we investigate the associations between polygenic risk scores (PGS) and gene expression, otherwise known as expression quantitative trait scores (eQTS), to determine how genetic effects regulate genes and pathways and provide insights into the biology of bipolar disorder.

We derive PGS for bipolar disorder from sample data composed of individuals of European ancestries collected from psychiatric hospitals and institutions throughout the Netherlands via the University of California, Los Angeles. The cohort consisted of 916 individuals diagnosed with BP, 358 controls, and 216 related individuals. We further utilized this cohort to perform bulk RNA-seq of whole blood tissue to quantify estimates for gene expression as well as identify eQTL associations.

We determined a significant correlation between PGS and covariate-adjusted expression for 195 genes (Bonferroni adjusted $P < 2.62 \times 10^{-6}$), which we refer to as eQTS genes. The overrepresented Gene Ontology categories included those related to hormone receptor binding, antigen binding, and the major histocompatibility complex. Also, of these 195 eQTS genes, 20 were located within 1MB of an existing GWAS risk loci while the remaining eQTS genes indicate potential new risk regions that regulate the expression of BP. LoF-intolerant genes ($pLI > 0.9$) comprise ~14% of all genes and are enriched in the eQTS gene set with ~44% being LoF-intolerant while only ~13% of significant TWAS are LoF-intolerant. Also, ~6% of all genes have significant overlap with transcription factors (TFs) (Fisher's exact test, $OR = 1.2$, $P < 1.6 \times 10^{-3}$) and are enriched in eQTS genes with ~15% being strongly associated with TF ($OR = 2.2$, $P < 1.8 \times 10^{-4}$). Analysis with significant TWAS genes found no significant overrepresentation with TF ($OR = 1$, $P = 0.52$). Our results demonstrate that eQTS has the ability to identify potential candidate genes that capture meaningful signals.

Complex Traits Posters - Wednesday

PB1287. Associations between polygenic risk scores, alcohol use, and differences in brain measures in the UK Biobank

Authors:

V. Thornton^{1,2}, Y. Chang¹, L. Fox¹, S. Hartz², J. Bijsterbosch², A. Anokhin¹, L. Bierut³; ¹Washington Univ. in St. Louis, St. Louis, MO, ²Washington Univ. Sch. of Med., St. Louis, MO, ³Washington Univ. Sch. of Med., St Louis, MO

Abstract Body:

Purpose: The UK Biobank offers an unprecedented resource with self-reported alcohol use, genotypes, and 936 imaging derived phenotypes (IDPs). We leverage this opportunity to explore the relationship between genetics, brain differences, alcohol use behavior to determine the direction of causality between brain and alcohol use. Our analysis includes IDPs representing global and regional gray matter volume and cortical thickness derived from T1 MRI, white matter hyperintensity lesions derived from T2 MRI, white matter integrity from diffusion MRI, and 6 independent components from rfMRI. **Methods:** Survey responses, genotyping, and MRI were all acquired in accordance with UK Biobank protocols. We performed six runs in PRSice-2 to generate polygenic risk scores for drinks per week. We then performed linear regression in R while controlling for systolic and diastolic blood pressure, waist hip ratio, BMI, income, education years, pack years, brain volume, site, sex, age, imaging date, head size, and rfMRI motion. We modelled the association between alcohol use and brain imaging, PRS and alcohol use, and PRS and brain imaging. **Results:** 102 out of 936 IDPs are significantly ($p \leq 5.34 \times 10^{-5}$ after Bonferroni correction for multiple comparisons) associated with alcohol use measured as drinks per week. These include decreased total brain volume, total grey matter volume, and total white matter volume, as well as numerous regional measures. The association between PRS for drinks per week and alcohol use in drinks per week is positive, large (Beta = 14.6) and highly significant ($p = 4.3 \times 10^{-43}$). 9 out of 936 IDPs are significantly associated with PRS for drinks per week, these do not include total brain volume. Two regions, the left fusiform and left superiorfrontal are associated with both PRS and alcohol phenotype. This is double what is predicted by chance alone. **Conclusion:** The regions associated with drinks per week PRS, alcohol use, and brain imaging show that in some regions genetics contributes to brain structure which is associated with patterns of alcohol use. However, many regions remain which are associated with alcohol use and not PRS, consistent with alcohol affecting brain structure.

Complex Traits Posters - Thursday

PB1288. Autism and Cognitive Ability CNV- Genome Wide Association study (CNV-GWAS).

Authors:

C. Poulain^{1,2}, C. Proulx^{1,2}, E. Douard^{1,2}, J-L. Martineau², Z. Saci², Z. Pausova³, T. Paus², L. Almasy⁴, D. C. Glahn⁵, G. Huguet², S. Jacquemont^{1,2}; ¹Univ. of Montréal, Montréal, QC, Canada, ²CHU Sainte-Justine Res. Ctr., Montréal, QC, Canada, ³The Hosp. for Sick Children, Toronto, ON, Canada, ⁴Univ. of Pennsylvania, Philadelphia, PA, Canada, ⁵Boston Children's Hosp., Boston, MA

Abstract Body:

Background: All rare genomic variants (including CNVs) associated with Autism Spectrum Disorder (ASD) have also been associated with Intellectual Disability (ID). Moreover, 30% of autism patients have intellectual disability. In fact, we have shown that any CNV genome-wide encompassing genes intolerant to haploinsufficiency increase ASD risk even after adjusting for their negative effects on cognitive ability. The relationship between genetic risk, cognitive ability, and ASD remains contentious in part because controls in most genetic studies were not assessed for cognitive ability. **Knowledge gap:** The specific effects of genes within CNV are not well known. Indeed, some genes seem to be more associated with ASD than ID and vice versa. **Hypothesis :** CNV impact can be assessed by genes characteristics with regressions models. **We aim** to identify rare CNVs that confer ASD risk, while carefully controlling for their effects on cognition. **Methods:** We performed an association study (CNV-GWAS) on an aggregate dataset (2 cohorts with ASD and 6 control cohorts) of ~450 000 individuals, to identify CNVs implicated in ASD. For genome-wide significant CNVs, we recomputed the association study after adjusting for cognitive ability.

Results: We replicated previous associations of 33 recurrent CNVs with ASD. We also identified 100 new regions (17 for deletion and 83 for duplication) and corresponding genes not previously associated with ASD. When adjusting for cognitive ability, 19 of the new regions (6 for deletion and 13 duplication) remain significant with ASD.

Conclusion : These datasets allowed us to detect ultra rare variants with better precision than before. This unique design including controls with cognitive assessments demonstrated that CNVs deletions and duplications remained associated with ASD even after adjusting for their effects on cognitive ability.

Complex Traits Posters - Wednesday

PB1289. Autism Spectrum Disorder (ASD) in Qatar: Whole genome sequencing of 100 affected families highlights Dominant and Recessive risk genes of ASD

Authors:

M. Abdi^{1,2}, E. Aliyev², B. Trost³, M. Kohailan^{2,1}, W. Aamer², N. Hoang³, R. Shaath^{3,2}, J. Howe³, O. Rennie³, A. Syed², J. Lakshmi², S. Hussein², S. Padmajeya², A. Hussein², I. Poggiolini², A. Akil², M. Kamal⁴, S. Scherer³, K. Fakhro²; ¹Hamad Bin Khalifa Univ., Doha, Qatar, ²Dept. of Genetics, Sidra Med., Doha, Qatar, ³Hosp. for Sick Children, Toronto, ON, Canada, ⁴Dept. of Pediatrics, Sidra Med., Doha, Qatar

Abstract Body:

Abstract:

Autism spectrum disorder (ASD) is a neurodevelopmental condition associated with reduced social and communication skills, along with restricted interests and repetitive behaviors. The prevalence of ASD among children in Qatar has been estimated to be 1.14%. Large-scale genomics studies of ASD have identified causative variants that span the entire mutational spectrum, ranging from single nucleotide variants (SNVs) to large structural variants (SVs). In outbred populations, ASD often appears to be caused by *de novo* mutations; however, in consanguineous populations, recessive biallelic variants may contribute to substantial risk of ASD. Whole genome sequencing of 104 families (total samples=402) with children clinically diagnosed with ASD (n=107) was performed at a pediatric tertiary academic center. Using an integrative genetic approach, we examine the genomes for rare SNVs, indels, and SVs, and prioritized deleterious (loss-of-function (LoF) or predicted damaging missense) variants affecting known or candidate ASD genes. In total, we discovered 13 potentially pathogenic SNVs in previously reported ASD genes, including *SCN2A*, *STAG1*, *ANK3*, *KDM5B*, *MTOR*, and *PTCHD1*. The majority were *de novo* (8/13) or X-linked (3/13), while biallelic variants were present in 2 of the families. In addition, we identified deleterious variants in 22 candidate ASD genes, including 12 *de novo* variants in genes such as *MOV10*, *NPAS3*, *CHD9*, and *HDAC7* and 10 biallelic events in genes such as *TRAPPC9*, *SYNE2*, *CDH23*, and *METTL2A*. Moreover, we identified 17 SVs (12 *de novo* and 5 biallelic) affecting ASD candidate genes, including *NIPBL*, *CSNK1A*, *ELOVL2*, *MED12L*, *CHRH1*, and *CLN3*. Overall, we identified a putative genetic risk factor in 39.2% (42/107) of ASD cases—29 dominant, 10 recessive, and 3 X-linked. Our study highlights the importance of WGS studies for providing a molecular diagnosis for ASD patients in tertiary care settings and provides a growing resource and first comprehensive look into the genetic structure underlying ASD in the middle east population.

Complex Traits Posters - Thursday

PB1290. Automated identification, GWAS, and ExWAS of enlarged perivascular space burden in human brain MRI in the UK Biobank

Authors:

N. Parikshak¹, B. Geraghty¹, S. Gelfman², V. Rajagopal³, G. Coppola¹, J. Marchini¹, GHS-RGC DiscovEHR Collaboration, Regeneron Genetics Center; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Regeneron Genetic Ctr., Tarrytown, NY, ³Regeneron Genetics Ctr., White Plains, NY

Abstract Body:

Background: Perivascular spaces (PVS) are fluid-filled compartments surrounding blood vessels in the brain. They may function in circulating interstitial fluid and clearing metabolites in order to maintain brain tissue health. Numerous visible PVS on magnetic resonance imaging (MRI) may represent pathology. Prior studies have associated PVS burden with cardiovascular risk factors, stroke, and poor cognitive function. We pursued a genetic investigation of quantitative PVS burden in a large cohort to identify new risk factors related to human disease. **Methods:** We obtained T1 MRI scans from the UK Biobank (UKB) and applied a novel image processing algorithm based on connected-component labelling to identify PVS in the centrum semiovale (CS, cortical white matter) and basal ganglia (BG). We validated the quality of segmentation by visual inspection and performed a genome wide association study with PVS counts across common (GWAS) and rare protein-coding (ExWAS) variants in 35,043 individuals. We used genetic correlations, phenome-wide associations, and gene expression data to annotate our findings. **Results:** GWAS identified 22 significant loci ($P < 5e-8$) in the CS and 16 in the BG. Comparing our PVS CS results to prior work, we found significant genome-wide signal +/-500kB of 12/21 loci of a prior GWAS of binarized high PVS burden in cortical white matter, including those near SLC13A3 and GFAP. We find that many genes near significant loci are expressed in the human cerebrovasculature. Phenotypic and genetic correlations revealed distinct signal compared to other MRI-derived traits and region-specific signals in PVS with respect to cardiovascular risk factors, stroke, and cognitive function. ExWAS identified a significant association between rare variant burden in GPR20 and increased PVS burden in the BG but not the CS ($P < 2.5e-6$, driven by missense variants with MAF < 0.01). GPR20 is highly expressed in arterial cell types and specifically in vascular smooth muscle cells of the human brain. Missense variants in GPR20 exhibited no association with blood pressure traits, a nominal association with reduced cognition, and a trend with elevated stroke risk.

Conclusion: Automated PVS quantification in large cohorts can enable large-scale genomic investigations at a biobank scale. Missense mutations in GPR20 likely modulate PVS burden via changes in vascular smooth muscle and may result in cognitive changes and elevated stroke risk. Further exploration of GPR20 and the genetics of region-specific PVS burden are warranted to understand the relationship between the human cerebrovasculature and health.

Complex Traits Posters - Wednesday

PB1291. Burden of functional variants in epilepsy patients using a deep learning approach.

Authors:

A. Girard¹, C. Moreau¹, J. Michaud², B. Minassian³, P. Cossette⁴, S. Girard¹; ¹Univ. of Quebec in Chicoutimi, Chicoutimi, QC, Canada, ²CHU Sainte-Justine Res. Ctr., Montreal, QC, Canada, ³Hosp for Sick Children, Toronto, ON, Canada, ⁴CHUM Res. Ctr., Montreal, QC, Canada

Abstract Body:

Background: Epilepsy is a neurological disease with a strong genetic component. Nevertheless, classical methods, like GWAS, are not as effective at detecting new causal loci as in other diseases. Non-coding regions have largely been ignored in the study of the disease, so far only the role of non-coding RNA was studied. We used a deep learning approach to predict the tissue specific functional effect of non-coding variants in epilepsy patients. **Methods:** Whole genome sequence data are available for 241 epilepsy patients and 388 controls. We computed the predicted expression change of variants using the deep learning algorithm ExPecto (Zhou *et al.* 2018). We used python's statsmodels to conduct logistic regressions by comparing patients and controls functional variants' burden across the genome in different sets of genes (all genes, epilepsy genes, genes intolerant to loss-of-function variants and ion channels genes). **Results:** Variants were filtered based on an expression change threshold. It corresponds to the value at which directionality prediction becomes perfect when compared to known eQTLs from GTEx. Using only variants that passed the threshold, we performed a logistic regression to compute the functional variants' burden. Preliminary results show no difference between groups. Nevertheless, we observed that the odds ratios tend to increase when we use a larger threshold. Thus, showing that non-coding regulatory regions might play a role in the underlying genetic mechanism of the disease. **Conclusion:** Deep learning algorithms are incredibly powerful tools to predict variant functional effects. Those methods are especially useful in neurology since brain tissues are hard to access. ExPecto is an important asset in our study of epilepsy, and it has the potential to be so for many other illnesses. **References:** Zhou, J. et al. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat. Genet.* 50, 1171-1179 (2018).

Complex Traits Posters - Thursday

PB1292. Causal effects of gut-related metabolites on human psychiatric disorders.

Authors:

S. Ihejirika¹, H. Xu², K. Ye¹; ¹Univ. of Georgia, Athens, GA, ²Univ. of Georgia, ATHENS, GA

Abstract Body:

Psychiatric disorders are currently the major cause of disability globally, and the human gut microbiome has been linked to several psychiatric disorders. However, the underlying mechanism by which the gut contributes to the etiology and progression of these diseases is still largely unknown. One hypothesis is that the gut microbiome produces certain metabolites which then travel to the brain and influence brain functions. In this study, the shared genetic basis and potentially causal relationship between metabolites produced by the gut microbiome and psychiatric disorders will be investigated. Summary statistics of GWAS in European populations were downloaded and after quality control, 17 psychiatric disorders were included. These disorders are major depression, schizophrenia, Alzheimer's disease, bipolar disorder, obsessive-compulsive disorder, anxiety disorders, post-traumatic stress disorder, anorexia nervosa, autism spectrum disorder, Tourette syndrome, attention deficit hyperactivity disorder, mood disorder, sleeplessness/insomnia, opioid dependence, alcohol use/dependence, cannabis use disorder, and neuroticism. A literature review was performed to identify metabolites that are related to the abundance of specific microbes or metabolic pathways in the gut microbiota. Only circulating metabolites with existing GWAS in UK Biobank were retained in our analysis to ensure sufficient sample size for genetic correlation analysis. We will first perform a genetic correlation analysis to examine shared genetic bases for all pairs of select metabolites and psychiatric disorders. Moreover, a two-sample Mendelian randomization analysis will be conducted to determine the causal effect of select metabolites on the 17 psychiatric disorders. The findings of this study will be useful in selecting metabolites and pathways of interest, to further investigate in the role they may play in the pathogenesis of psychiatric disorders.

Complex Traits Posters - Wednesday

PB1293. Cell-type-specific activity-dependent multiomic profiling in single hiPSC-derived neurons for neuropsychiatric disorders

Authors:

S. Zhang¹, C. Li², W. Wood¹, B. V. Jamison¹, H. Zhang¹, A. Kozlova³, A. Sanders⁴, Z. Pang⁵, X. He⁶, J. Duan⁷; ¹NorthShore Univ. Hlth.System, Evanston, IL, ²Northshore Res. Inst., Evanston, IL, ³The Res. Inst., NorthShore Univ. Hlth.System, Evanston, IL, ⁴North Shore Univ Hlth.System, Evanston, IL, ⁵Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ, ⁶Univ Chicago, Chicago, IL, ⁷Northshore Univ Hlth.system/Univ of Chicago, Evanston, IL

Abstract Body:

Neuronal activity-dependent (NAD) gene regulation plays a vital role in brain development, learning, and memory processes that are impaired in neuropsychiatric disorders. While many NAD response genes, such as FOS (Fos Proto-Oncogene, AP-1 Transcription Factor Subunit), and late-response genes, such as BDNF (Brain-Derived Neurotrophic Factor), are well-known, their cell-type-specific response to neural activity and the underlying regulatory mechanisms are not well understood. Here, we co-cultured excitatory (Ex; glutamatergic) and inhibitory (Inh; GABAergic) neurons derived from human (hiPSC) of five donors, mimicked neuronal activities using 50 mM KCl stimulation, and performed single-cell multi-omics (scRNA-seq, scATAC-seq). We also performed CRISPR/Cas9-based gene editing on a specific open (accessible) chromatin region (OCR). Our integrative analysis of scRNA-seq and scATAC-seq data identified more than 9,000 cell-type-specific, differentially expressed (DE) genes after 1 hr (early response) and 6 hrs (late response) of KCl treatment, of which approximately 50% were either Ex or Inh neuron-specific. The NAD transcriptomic changes were accompanied by substantial alterations in genome-wide chromatin accessibility, with 11-18% and 23-32% differentially accessible peaks across cell types. We found an Ex neuron-specific KCl-responsive OCR (~700 bp) that is ~50 kb upstream of the BDNF transcription starting site. It exhibited increased accessibility at 1 hr and reached its peak intensity at 6 hrs. Peak-gene linkage analysis also correlated this OCR with BDNF expression. Additional transcription factor (TF) footprinting within this OCR revealed specific binding sites of the components of AP-1 TF heterodimer such as FOS and JUN families (early response genes) at 1 hr of KCl stimulation, while the TF binding pattern at 6 hrs of KCl stimulation shifted to POU2F1/POU5F1/POU3F4 and TCF3/4 families, both have been implicated in neurodevelopmental disorders. These results were further confirmed using 18 donor lines. Finally, we deleted the OCR by CRISPR/Cas9 editing in hiPSC with ~60% editing efficiency. We observed ~42% reduction of BDNF, suggesting it is required for the NAD BDNF expression. In summary, our study identified widespread cell-type-specific NAD transcriptomic and chromatin accessibility changes in a hiPSC-derived neuronal model and novel mechanistic insight into NAD-regulated BDNF expression, which itself plays part in neurodevelopment and neuropsychiatric disorders.

Complex Traits Posters - Thursday

PB1294*. Cellular context-specific gene regulation of neuropsychiatric disorders in single human neurons

Authors:

W. Wood¹, S. Zhang², H. Zhang¹, L. Liang³, C. Li⁴, B. Jamison¹, A. Kozlova¹, A. Sanders², Z. Pang⁵, X. He⁶, J. Duan⁷; ¹NorthShore Univ. Hlth.System, Evanston, IL, ²North Shore Univ Hlth.System, Evanston, IL, ³Dept. of Human Genetics, Univ. Of Chicago, Chicago, IL, ⁴Northshore Res. Inst., Evanston, IL, ⁵Dept. of NeuroSci. & Cell Biology, Rutgers Univ., New Brunswick, NJ, ⁶Dept. of Biophysics, Univ. of Chicago, Chicago, IL, ⁷Northshore Univ Hlth.system/Univ of Chicago, Evanston, IL

Abstract Body:

Genome-wide association studies (GWAS) of neuropsychiatric disorders have identified a plethora of neuropsychiatric risk loci that affect chromatin accessibility and gene expression (Zhang, Science 2020). However, functional interpretation of possible causal variants/genes remains challenging, especially in the interaction between polygenic risk effects and the environment. Regulatory variants often act in specific biological contexts, e.g., only in stimulated cells. We hypothesize that many neuropsychiatric risk variants may alter cell-type-specific chromatin accessibility and gene expression preferentially in activated/stimulated neurons. We model neural activation by potassium chloride (KCl) stimulation in co-cultured excitatory and inhibitory neurons of ~100 hiPSC lines, then assaying activity-dependent single-cell multiomes (scRNA/ATAC-seq) by mapping activity-specific expression and chromatin QTLs. With five donor lines as the start, we identified cell-type-specific expression changes of thousands of genes (40-50%) after 1hr (e.g., FOS) or 6hrs (e.g., BDNF) post-KCl stimulation in three neuron subtypes (GABAergic inhibitory, NEFM+ or NEFM- excitatory neurons). MAGMA analysis showed that the up-regulated genes are enriched for neuropsychiatric GWAS risk, with the strongest enrichment for schizophrenia (SZ). Most neuropsychiatric risk genes showed a later response and were predominantly up-regulated except for post-traumatic stress disorder (PTSD) risk genes. Moreover, the expression dynamics of risk genes shared by different disorders tended to be conserved between cell types; for instance, risk genes shared by SZ and bipolar (CACNA1C, FURIN, MAD1L1, and SP4) were up-regulated in all cell types. Consistent with the transcriptomic landscape, KCl-stimulation also led to changes in chromatin accessibility, with 11-18% and 23-32% differentially accessible (dynamic, DA) peaks across cell types at 1hr and 6 hrs post-stimulation. sLDSC analysis further showed that DA peaks showed stronger heritability enrichment of SZ and other neuropsychiatric disorders than static peaks. Furthermore, 604 SNPs in GABAergic and 608 SNPs in excitatory neurons showed differential allelic chromatin accessibility (i.e., allele-specific open chromatin, ASoC), most of which were stimulation-specific, including 4 SZ GWAS risk SNPs that showed activity-dependent ASoC. Our study provides novel mechanistic insights into how neuropsychiatric risk variants confer disease risk, expanding the repertoire of regulatory neuropsychiatric risk variants that may affect chromatin accessibility and gene expression in activated neurons.

Complex Traits Posters - Wednesday

PB1295. Characterization of a diverse Frontotemporal Dementia cohort, enriched for Caribbean Hispanic patients.

Authors:

K. Nuytemans^{1,2}, A. Martinez¹, F. RAJABLI¹, E. Gu¹, S. Tejada¹, H. Acosta³, B. Baumel⁴, C. Camargo⁴, X. Sun⁴, J. Vance^{1,2}, M. Cuccaro^{1,2}, M. Pericak-Vance^{1,2}; ¹John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL, ²Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami, Miami, FL, ³Caribbean Ctr. for the Study of Memory and Cognition, San Juan, Puerto Rico, ⁴Dept. of Neurology, Univ. of Miami, Miami, FL

Abstract Body:

Background: The vast majority of biomedical data currently available for any disease is derived from studies in non-Hispanic white (NHW) populations. Specifically, clinical information, genetic factors as well as biomarkers for frontotemporal dementia (FTD) have been studied predominantly in those NHW populations. To increase representation in biomedical research, we set out to enroll and characterize a diverse FTD patient cohort enriched for Caribbean Hispanic patients. **Methods:** Our current cohort consists of 89 FTD patients (30% NHW, 67% Hispanic), with continuing enrollment from the University of Miami Hospital Neurology Department in Miami, FL and the Caribbean Center for the Study of Memory and Cognition in San Juan, PR. All patients were evaluated using NACC approved Uniform DataSet (UDS) or equivalent in their preferred language. For ~65% of the cohort we also completed the NACC FTD module forms. We generated genotyping data (Illumina GDA+Neurobooster array) as well as whole genome sequencing and plasma biomarker data (Quanterix Simoa Neuroplex-3; A β 40, A β 42, and total tau) for a subset of the cohort. **Results:** Initial genetic analyses showed none of the Hispanic patients are carriers of known FTD mutations originally identified in NHW patients, including the *C9orf72* repeat expansion and reported pathogenic variants in *MAPT* or *GRN*. We did not identify a significant difference in age-at-onset or Clinical Dementia Rating scores at time of enrollment between NHW and Hispanic patients. Additionally, biomarker data on A β 40/A β 42 ratio and total tau levels in a subset of 22 FTD patients (~12/10 Hispanic/NHW) did not show significantly different levels between patients of both ethnicities. **Conclusions:** Genetic analyses of FTD in underrepresented population groups is necessary as genetic information from research in NHW is not always generalizable across race/ethnicity. We are currently working to expand our efforts to include identification of novel genetic risk factors for FTD in the Hispanic patients using whole genome sequencing, full evaluation of the Neuroplex as well as p-tau181 and NfL biomarkers in the complete cohort and comparison of clinical presentations between ethnicities. The biomedical characterization of FTD across race/ethnicity will help the understanding of disease mechanisms in all patients ultimately preventing further health disparities.

Complex Traits Posters - Thursday

PB1296. Characterization of gene expression profile and molecular pathways involved in type 2 diabetes among self-identified Hispanic American individuals

Authors:

M. Yaser¹, H-H. Chen², H. Polikowsky³, L. Petty³, M. Lee⁴, J. McCormick⁴, S. Fisher-Hoch⁴, H. Highland⁵, M. Graff⁶, J. Seo¹, K. Young¹, K. North⁶, J. Below⁷, J. Below⁸; ¹Univ. of North Carolina Chapel Hill, Chapel Hill, NC, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Vanderbilt Univ., Nashville, TN, ⁴The Univ. of Texas Hlth.Sci. Ctr. at Houston Sch. of Publ. Hlth., Brownsville, TX, ⁵Univ North Carolina at Chapel Hill, Chapel Hill, NC, ⁶Univ North Carolina, Chapel Hill, NC, ^{7,8}Vanderbilt Univ Med Ctr., Nashville, TN

Abstract Body:

Background In the United States, 37.3 million adults have type 2 diabetes (T2D), with longstanding disparities affecting minorities, particularly Hispanic/Latino (H/L) populations. Both environmental and genetic contributions to T2D are widely acknowledged, and while genome-wide association studies (GWAS) have identified hundreds of loci associated with T2D, the function of much of this variation is unknown. Gene expression measures can illuminate the link between genetic variation and T2D, but to-date, only a handful of studies has examined the role of gene expression to identify molecular signatures associated with T2D, particularly in those populations most burdened by disease. In this study, we characterized the differential gene expression profiles and dysregulated pathways associated with T2D among H/L. **Methods** We leveraged whole blood cells derived RNA-sequencing data in 181 T2D with T2D and 424 normoglycemic controls collected from randomly selected study participants from the Cameron County Hispanic Cohort (CCHC) to identify patterns of differentially expressed (DE) genes. We used established protocols and alignment, yielding 15,694 protein coding genes after quality control. We applied *EdgeR* to assess DE, adjusted for sex, age, BMI, and 10 probabilistic estimation of expression residual (PEER) factors, followed by functional annotation and enrichment analyses of significantly dysregulated genes. **Results** After false discovery rate (FDR<0.05) adjustment, 126 genes were DE, including 36 downregulated and 90 upregulated genes in cases compared to controls. Of this set, 15 genes illustrated fold change of >5% in cases versus controls. We observed upregulation of genes *RETN*, *HP*, *ACE* and *ITG1A* genes, previously reported associated with T2D. Interestingly, novel gene expression associations were mapped, including for upregulated *CPT1A* which has previously been associated with gestational diabetes. Amongst people with T2D we observed increased WB expression of *NPAS4*, a neuronal transcription factor that is known to regulate expression of *BDNF*, part of the hunger/satiety pathway. Although less understood, *NPAS4* is also regulates transcription in pancreatic β -cells. Pathway enrichment analyses identified pathways involved in the immune response (activation and migration of leukocytes), stress signaling, and cell secretion activities. **Conclusions** Collectively, these data demonstrate how transcriptomic studies may identify novel genes important for T2D pathogenesis, elevating our understanding of molecular pathways involved and provide candidate therapeutic targets in this understudied population.

Complex Traits Posters - Wednesday

PB1297. Characterization of sequencing-based HLA alleles in a Quebec COVID-19 biobank

Authors:

P. McClelland¹, P. Nadukkalam Ravindran¹, V. Mooser², S. Zhou¹, C. Bhérier¹, D. Taliun¹; ¹McGill Univ., Montreal, QC, Canada, ²McGill Genome Ctr., Montreal, QC, Canada

Abstract Body:

Recent genetic studies have reported associations between HLA alleles and related polymorphisms with susceptibility, severity and progression of COVID-19. However, further investigation is needed to replicate these findings in other populations and better understand the role of HLA in the pathogenesis of COVID-19. Our project aims to characterize HLA alleles in the people from the province of Quebec, Canada, and identify their role in disease progression. We performed HLA typing (class I and II) using high-depth WGS data from 2,074 participants (1,558 European (EUR), 172 African (AFR), 118 East Asian (EAS), 90 South Asian (SAS), and 104 Admixed American (AMR) inferred genetic ancestries and 32 of other origins) in the Biobanque québécoise de la COVID-19 (BQC19), which collects high-quality biosamples and clinical data of hospitalized and non-hospitalized SARS-CoV-2 PCR positive and negative individuals. In total, we identified 403 HLA alleles, which passed quality checks: 71 A, 128 B, 53 C, 11 DPA1, 39 DPB1, 10 DQA1, 21 DQB1, and 70 DRB1. Of these, 347 HLA alleles were also present in the most extensive publicly available multi-ancestry high-resolution HLA-reference panel (N = 21,546). 47 HLA alleles in BQC19 had frequencies different from the HLA-reference panel (P-value < 14×10^{-5}): 29 in EUR, 17 in EAS, and 1 in AFR. We performed logistic regression (covariates: sex, age, 1-5 PCs) to test for HLA alleles association with COVID-19 severity (cases/controls = 401/980) and hospitalization (cases/controls = 816/575) among unrelated (>3rd degree) patients who tested positive. None of the associations reached the statistical significance threshold (P-value < 12×10^{-5}) corrected for multiple-testing, however, HLA-A*32:01 showed evidence of a protective effect against hospitalization (OR = 0.42) with moderate significance (P-value = 17×10^{-5}). This replicates previous suggestive evidence of HLA-A*32:01 at increased frequencies in healthy controls compared to COVID-19 patients. When stratified by ancestry, the HLA-A*32:01 allele showed the lowest P-value = 0.003 among all alleles for association with hospitalization in Europeans. All other ancestry-stratified analyses were underpowered. We are currently building a population-specific HLA reference panel for imputing HLA alleles using genotype-array data, allowing us to explore HLA associations in large-scale biobanks with tens of thousands of genotype individuals in the province and worldwide.

Complex Traits Posters - Thursday

PB1298. Characterizing 31 cardiometabolic candidate genes for a role in early vascular inflammation and foam cell formation using CRISPR-Cas9, *in vivo* imaging and deep learning

Authors:

M. den Hoed¹, E. Mujica¹, H. Zhang¹, A. Emmanouilidou¹, E. Mazzaferro¹, M. Nyberg², C. Metzendorf¹, M. Bandaru¹, N. Cook¹, J. Campos-Costa¹, G. Alavioon¹, S. Gry Vienberg², A. Larsson³, D. Djordjevic², A. Allalou⁴; ¹Beijer Lab. and Dept. of Immunology, Genetics and Pathology; Uppsala Univ. and SciLifeLab, Uppsala, Sweden, ²Novo Nordisk A/S, Målöv, Denmark, ³Dept. of Med. Sci.; Div. of Clinical Chemistry; Uppsala Univ., Uppsala, Sweden, ⁴Dept. of Information Technology; Div. of visual information and interaction; Uppsala Univ. and SciLifeLab, Uppsala, Sweden

Abstract Body:

INTRODUCTION: Genome-wide association studies have identified hundreds of loci that are robustly associated with risk of coronary artery disease. For most loci, the causal genes remain to be functionally characterized. Cell-based model systems provide valuable insights, but cannot model the complexity of a live, intact organism, while mouse models are too costly and time consuming to characterize hundreds of genes. In addition, mice lack *CETP* and only develop atherosclerosis in an *Apoe*^{-/-} or *Ldlr*^{-/-} background. To overcome these challenges, we developed and validated image- and CRISPR/Cas9-based model systems in zebrafish larvae, which we use here to functionally characterize the role of candidate genes in early-stage vascular inflammation and foam cell formation. **METHODS:** We characterized 21 genes for a role in early-stage vascular inflammation (i.e., vascular accumulation of lipids, neutrophils and macrophages), and 21 genes for a role in early-stage foam-cell formation (i.e., vascular accumulation of lipids, oxidized LDL and macrophages) using fluorescently labelled transgenes and a lipid staining dye. From day 5 post-fertilization, zebrafish larvae were fed on regular dry food (29 genes) or on regular dry food enriched with 4% extra cholesterol (10 genes). At day 9, 10 or 11, we acquired optical sections of the vasculature and liver using a fluorescence microscopy, and whole-body images in bright field. Outcomes were quantified using deep learning. After imaging, larvae were sacrificed and homogenized for enzymatic assessment of LDLc, triglyceride, total cholesterol and glucose levels. Crispants were distinguished from controls using a fragment length analysis. **RESULTS AND DISCUSSION:** By imaging 7270 CRISPR/Cas9 founders and controls for 31 genes, we show that cholesterol challenged *adamts7* crispants have less vascular infiltration of oxLDL and macrophages than controls; while CRISPR/Cas9-induced mutations in *gucylal* and *il6r* result in more vascular infiltration by oxLDL and macrophages, even without a cholesterol challenge. In line with the effects of SGLT2 inhibitors on glucose metabolism and cardiovascular risk, mutations in *slc5a2* result in lower glucose and LDLc levels in cholesterol-challenged larvae, and in less vascular infiltration by neutrophils in unchallenged larvae. **CONCLUSION:** By systematically characterizing candidate genes for a role in a range of atherosclerosis-related traits in zebrafish larvae, we can prioritize genes for further in-depth characterization and begin to understand the mechanisms by which loci identified by genome-wide association studies influence disease risk.

Complex Traits Posters - Wednesday

PB1299. Cholesterol and LDL-C levels are elevated prior to symptom onset in Alzheimer's disease independent of APOE-4.

Authors:

B. Chase^{1,2}, R. Frigerio³, D. Maraganore⁴, C. J. Yucus⁵, S. Patel⁵, J. S. Castle⁵, A. Epshteyn¹, A. R. Sanders⁶, J. Duan⁷, K. Markopoulou^{5,2}; ¹NorthShore Univ. Hlth.System, Skokie, IL, ²Univ. of Chicago, Chicago, IL, ³NorthShore Univ. Hlth.System, Evanston, IL, ⁴Tulane Univ., New Orleans, LA, ⁵NorthShore Univ. Hlth.System, Glenview, IL, ⁶North Shore Univ Hlth.System, Evanston, IL, ⁷Northshore Univ Hlth.system/Univ of Chicago, Evanston, IL

Abstract Body:

Background: Many of the >70 Alzheimer's disease (AD) risk loci, including *APOE*, impact lipid metabolism. Polygenic risk scores of low-density lipoprotein-cholesterol (LDL-C) levels, but not high-density lipoprotein-cholesterol (HDL-C), are also associated with AD risk. Although LDL-C plasma levels are elevated in early-onset AD, their association with the strongest AD risk allele, APOE-4, and their utility as a biomarker for later-onset AD (LOAD) progression remains controversial. **Methods:** To test associations of plasma lipid levels with AD, mild cognitive impairment (MCI), and *APOE* at times 5 yr before, at, and 5 yr after symptom onset, we utilized clinical data from a community-based practice setting gathered over up to 9 yr using standardized clinical documentation support tools embedded in the electronic medical record (EMR), and retrospective lipid panel data. Patients were genotyped as part of the DodoNA study using a custom Axiom SNP array. We compared lipid levels in AD (n = 390, 93% LOAD), MCI (n = 245), and at-risk (baseline = visit 1, n = 293) cohorts to 2-3 fold more controls matched for age, sex, and statin use. Controls were from other neurological disease cohorts and lacked ICD codes for cognitive impairment, dementia or Parkinson's disease. We report nominally significant associations between lipid levels and the conditions and *APOE* at each time, controlling for age, sex, body-mass index, and race. **Results:** AD was associated with increased cholesterol [pre: β (95% CI) = 10.1 (4.6, 15.6), $p < 0.0001$; onset: $\beta = 7.1$ (1.1, 13.2), $p = 0.021$; post: $\beta = 11.1$ (4.8, 17.3), $p = 0.001$], LDL-C [pre: $\beta = 7.9$ (3.2, 12.6), $p = 0.001$; post: $\beta = 5.94$ (0.5, 11.3), $p = 0.031$] and non-HDL-C [pre: $\beta = 8.0$ (3.0, 13.0), $p = 0.002$; onset: $\beta = 6.0$ (0.3, 11.6), $p = 0.038$; post: $\beta = 9.5$ (3.7, 15.4), $p = 0.001$]. While high cholesterol and LDL-C were associated with AD 5 yr before onset, no lipids were associated with MCI at any time point. Associations of lipids with statin use, sex, and age were consistent across time and cohort. Associations with at-risk cohort membership were inconsistent. APOE-4 was associated with cholesterol [$\beta = 11.8$ (4.3, 19.4), $p = 0.002$], non-HDL-C [$\beta = 12.1$ (5.2, 19.1), $p = 0.001$] and LDL-C [$\beta = 13.2$ (6.5, 19.9), $p < 0.001$] in MCI, but only 5 yr before onset, while in AD, these associations were not seen at any time. APOE-2 had consistent negative associations with cholesterol and LDL-C. **Conclusions:** Higher LDL-C and cholesterol levels are associated with increased AD risk independent of APOE-4. Finding higher LDL-C and cholesterol levels in AD but not MCI five years prior to symptom onset suggests that plasma LDL-C and cholesterol levels may serve as a biomarker for LOAD progression.

Complex Traits Posters - Thursday

PB1300. Choroidal thickness as a biomarker for age-related macular degeneration progression: Genome-wide Association Study in Amish

Authors:

J. Cooke Bailey¹, F. RAJABLI², Y. Song¹, O. García-Rodríguez³, L. Gómez³, S. Slifer³, K. Miskimen¹, M. Nittala⁴, S. Sadda⁴, J. Haines¹, M. Pericak-Vance³, D. Stambolian⁵, W. Scott²; ¹Case Western Reserve Univ., Cleveland, OH, ²Univ. of Miami, Miami, FL, ³Univ. of Miami Miller Sch. of Med., Miami, FL, ⁴Doheny Imaging Reading Ctr., Doheny Eye Inst., Los Angeles, CA, ⁵Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Prevalence of age-related macular degeneration (AMD), the second leading worldwide cause of irreversible blindness, continues to rise with the growth of the population over 65 years of age. Progression from early to late AMD varies substantially, and factors influencing this progression are not yet fully discerned. Choroidal thickness (CT), which is significantly heritable and inversely associated with AMD severity, is a promising, potential AMD biomarker. To identify genetic loci influencing CT change, we performed a genome-wide association study (GWAS) in the Amish Eye Study (AES). Amish are a founder population with less genetic and environmental variation than the general population. AES recruits from the three largest Amish communities in the United States: Holmes County, Ohio, Elkhart and LaGrange Counties, Indiana, and Lancaster County, Pennsylvania. For this analysis, genotype data from the Illumina MEGAex array were generated for a family-based prospective cohort of 499 Amish adults over age 65. Association between single nucleotide polymorphism (SNP) allele dosage and two year difference in CT, measured on Heidelberg Spectralis optical coherence tomography (OCT) instruments, was assessed with the GWAF program via linear mixed modeling adjusting for sex, age, spherical equivalent refraction, intraocular pressure, and population and pedigree structures. Although no loci were associated with two year difference in CT at genome-wide significance ($p\text{-value} < 5 \times 10^{-8}$), we identified six genes with SNPs generating suggestive association ($p\text{-value} < 5 \times 10^{-5}$). At four SNPs, the rare allele was suggestively associated with decreasing CT: rs17418461 ($p\text{-value} = 3.9 \times 10^{-6}$, *DABI*) and rs17535312 ($p\text{-value} = 8.6 \times 10^{-6}$, *TLL7*) on chromosome 1, and rs576371849 ($p\text{-value} = 4.8 \times 10^{-6}$, *LINC00900*) and rs116309326 ($p\text{-value} = 4.8 \times 10^{-6}$, *LINC02698*) on chromosome 11. At two SNPs, rare alleles were suggestively associated with increasing CT: rs115627559 ($p\text{-value} = 5.5 \times 10^{-6}$, *TBC1D1*) on chromosome 4 and rs35892074 ($p\text{-value} = 9.22 \times 10^{-6}$, *LRRC28*) on chromosome 15. Two known loci for advanced AMD (or appropriate proxy SNPs) were nominally significant: rs10117842 ($p\text{-value} = 0.02$, *TRPM3*) on chromosome 9 and rs20145990 ($p\text{-value} = 0.04$, *C20orf85*) on chromosome 20. *DABI*, *TLL7*, and *TBC1D* are expressed in the retina and, thus, we have identified three genes that may be associated with CT change, and, therefore, AMD progression. Follow up studies are needed to replicate these findings. Identification of novel genes associated with CT and/or AMD progression could provide insight into mechanisms underlying progression to AMD and inform future therapeutic developments.

Complex Traits Posters - Wednesday

PB1301. Classical HLA Alleles are Associated with Fasting Glucose and Type 2 Diabetes in Multiple Populations

Authors:

G. Chen¹, J. Zhou², J. Zhang³, A. Bentley³, A. Doumatey², D. Shriner⁴, C. Rotimi⁵, A. Adeyemo⁵; ¹NHGRI/CRGGH, Bethesda, MD, ²NHGRI, Bethesda, MD, ³NIH, Bethesda, MD, ⁴Natl. Inst.s Hlth, Bethesda, MD, ⁵NIH, Bethesda, MD

Abstract Body:

Introduction: Class II major histocompatibility complex (MHC-II) loci play a key role in the pathophysiology of type 1 diabetes (T1D). However, several GWAS for type 2 diabetes (T2D) have found genome-wide significant variants in both class 1 and 2 HLA genes. The relationship between classical HLA alleles and T2D has rarely been investigated in large cohorts. We investigated the association between classical human leukocyte antigen (HLA) alleles and fasting glucose (FG) and T2D in large study cohorts. **Methods:** Participants from 11 datasets were included in this study for a total of 30,376 individuals: 5,080 Africans, 9,357 African Americans, 12,045 European Americans, 697 Asian Americans, 1,907 Chinese, and 1,290 Hispanic Americans. Of these, 17.6% had T2D. Classical HLA alleles (*A*, *B*, *C*, *DP*, *DQA*, *DQB* and *DR*) were imputed using the HIBAG algorithm. Association analysis was done by study under an additive genetic model with covariates of sex, age, body mass index, and significant principal components of population structure, followed by a fixed-effects meta-analysis. P values were adjusted using the Bonferroni method. **Results:** The most significant associations with FG were with HLA-DQA1*01:02 ($P_{adj}=0.0050$), HLA-DRB1*15:01 ($P_{adj}=0.0059$) and HLA-DQB1 ($P_{adj}=0.015$) while HLA-DQA1*01:02 was associated with T2D with $p_{adj}=0.038$. These findings remained significant after adjusting for a SNP-derived PRS for FG and T2D, as well as after conditioning on the most significant SNPs within the MHC region. Haplotype analysis showed that DRB1*15:01-DQA1*01:02-DQB1*06:02 was associated with FG ($p=0.003$) and T2D ($p<0.0001$). In Africans, in whom additional data was available, this haplotype was also strongly associated with higher insulin resistance (IR) ($p<0.0001$), lower Gastric Inhibitory Polypeptide (GIP) ($p=0.0014$), and Glucagon-Like Peptide-1 (GLP-1) levels ($p<0.0001$). **Conclusions:** This study reveals that classical HLA alleles are significantly associated with higher FG and T2D in multiple populations. Our finding in Africans of an HLA haplotype associated with higher IR and lower GIP and GLP-1 levels suggests that the associations are mediated via mechanisms that result in insulin resistance and/or reduced β -cell stimulation in response to the gut sensing glucose or a meal. This work adds to a growing body of literature demonstrating some overlap in the pathophysiology and genetic architecture of T1D and T2D.

Complex Traits Posters - Thursday

PB1302. CLEC16A regulates the adipogenic and lipolytic capacity of adipocytes

Authors:

M. Bakay¹, R. Pandey², B. Strenkowski², H. Hakonarson¹; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Children's Hosp. of Philadelphia, Philadelphia, PA, United States, Philadelphia, PA

Abstract Body:

CLEC16A has been shown to play a role in autophagy/mitophagy processes and genetic variants in *CLEC16A* are implicated in multiple autoimmune diseases. To investigate the role of CLEC16A we have generated an inducible ubiquitous *Clec16a* knockout (KO). The global *Clec16a* KO mice (*Clec16a*^{AUBC}) exhibit a complex phenotype with marked immune dysfunction, rapidly progressing sensory neurodegeneration and severe lipodystrophy. Recently we demonstrated that JAK/STAT inhibitor therapy partially rescues the lipodystrophic autoimmune phenotype of *Clec16a*^{AUBC}. Here, we investigate CLEC16A contribution to adipocyte functions.

Methods: We generated *Clec16a*-deficient 3T3-L1 adipocytes using the CRISPR-Cas9 system and continued the exploration of *Clec16a*^{AUBC} mice for loss-of-function studies. **Results:** We established *Clec16a*-deficient 3T3-L1 cell line (Clec-KO) and confirmed the *Clec16a* deletion for DNA, RNA, and protein. The rate of cell proliferation of Clec-KO preadipocytes was significantly faster comparing to Ctrl cells. Autophagy markers (LC3 and p62) as well as mTOR activity were not affected in Clec-KO preadipocytes when comparing to Ctrl in starvation-induced autophagy. To examine the roles of CLEC16A on adipogenesis, Clec-KO preadipocytes were differentiated into adipocytes, and analyzed for lipid accumulation using Oil-Red-O staining. While Clec-KO cells showed comparable lipid droplet accumulation within the adipocytes to that of controls, secreted adiponectin levels in the conditioned media of Clec-KO were slightly decreased versus Ctrl. To further elucidate the roles of CLEC16A in adipogenesis, the mRNA levels of key transcription factors for adipogenesis such as C/EBPs and PPAR γ were analyzed: C/EBP β and C/EBP α were dysregulated; PPAR γ at terminal differentiation was significantly decreased in Clec-KO. mRNA levels of mature adipocyte marker genes such as FABP4, leptin, and Adipoq were also significantly decreased in Clec-KO cells. The mRNA levels of marker genes encoding adipocyte lipolysis, including adipocyte triglyceride lipase (ATGL), hormone-sensitive lipase (HSL), monoacylglycerol lipase (MGL), and lipoprotein lipase (LPL), were significantly elevated compared to Ctrl. **Conclusions:** Mature adipocyte differentiation was attenuated in Clec-KO adipocytes, indicating that CLEC16A is important for supporting adipogenesis. While these finding require further investigation, they concur with previous results in *Clec16a*^{AUBC} mice and suggest that adipocytes are important in regulating various biological functions, including weight gain and may play a role in lipodystrophies.

Complex Traits Posters - Wednesday

PB1303*. Clinical utility of polygenic scores for cardiometabolic disease in Arabs

Authors:

I. Shim^{1,2,3}, H. Kuwahara⁴, N. Chen⁴, M. Hashem⁵, L. AlAbdi⁶, X. Gao⁴, M. AbouElHoda⁶, H-H. Won³, P. Natarajan^{1,2}, P. T. Ellinor^{1,2}, A. V. Khera^{1,2,7}, F. S. Alkuraya⁶, A. C. Fahed^{1,2}; ¹Div. of Cardiology, Dept. of Med., Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, ²Cardiovascular Disease Initiative of the Broad Inst. of Harvard and MIT, Cambridge, MA, ³Dept. of Digital Hlth., SAIHST, Sungkyunkwan Univ., Samsung Med. Ctr., Seoul, Korea, Republic of, ⁴Computational BioSci. Res. Ctr., King Abdullah Univ. of Sci. and Technology, Thuwal, Saudi Arabia, ⁵Dept. of Computational Sci., Ctr. for Genomic Med., King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, ⁶Dept. of Translational Genomics, Ctr. for Genomic Med., King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, ⁷Verve Therapeutics, Cambridge, MA

Abstract Body:

Background Arabs account for 5% of the world population and have a high burden of cardiometabolic disease, yet they remain underrepresented in genetic studies. What is the clinical utility of polygenic scores for cardiometabolic disease in Arabs?

Method In 5,399 Arab patients (age 54.8±14.8 years, 35.7% female), we tested the performance of European-derived polygenic scores for 3 diseases (coronary artery disease [CAD], type 2 diabetes [DM-2], and cardiomyopathy [CM]), and 7 traits (LDL cholesterol, HDL cholesterol, triglyceride, systolic and diastolic blood pressures [BP], body mass index [BMI], and height). To optimize better scores for Arabs, in a validation subset (N=2,700) we selected the score with the largest AUC or R² out of multiple candidate scores calculated using best-available methods (PRSice2, LDpred2, lassosum2, PRS-CS, and PRS-CSx). We constructed a population reference distribution (N=1,017) from ethnically matched Arabs to create ancestry-adjusted polygenic scores and define percentiles. In a testing subset (N=2,699), we studied the association of selected scores with disease, the proportion of individuals with high genomic risk, and the interplay of genomic risk and clinical risk factors. Models were adjusted for age, sex, and genetic ancestry.

Results Compared to European-derived scores, we identified better performing scores for all 10 cardiometabolic traits, some with comparable performance to Europeans despite the lack of large Arab GWAS. Optimized scores for CAD, DM-2, and CM had odds ratio per standard deviation (SD) of 1.49, 1.74, and 1.32 and AUC of 0.79, 0.73, and 0.64, respectively. Optimized scores for LDL cholesterol, HDL cholesterol, triglycerides, SBP, DBP, BMI, and height had adjusted R² of 0.04, 0.13, 0.07, 0.11, 0.04, 0.08, and 0.51, respectively, and each score SD was associated with an increase of the trait by 9.9 mg/dL, 3.5 mg/dL, 28.3 mg/dL, 2.9 mmHg, 1.8 mmHg, 1.1 kg/m², and 0.025 m, respectively. The proportion of people with more than two-fold increased genomic risk was 21% for CAD, 29% for DM-2, 5% for CM, 10.6% for obesity, 4% for hypercholesterolemia, and 7.5% for hypertension. Genomic risk for CAD was independent of conventional risk factors and the polygenic score identified individuals who had early-onset and more severe disease.

Conclusion We describe a pragmatic approach for optimizing polygenic scores for 10 cardiometabolic traits in Arabs and show a strong promise for clinical utility. This framework leverages publicly available resources - short of much-needed large GWAS in non-Europeans - and could be generalizable to other populations to advance equitable clinical implementation of polygenic scores.

Complex Traits Posters - Thursday

PB1304. Clonal behavior across multiple timepoints in clonal hematopoiesis

Authors:

T. Mack^{1,2}, K. Von Beck¹, A. Silver¹, M. Savona¹, A. Bick^{1,2}; ¹Vanderbilt Univ. Sch. of Med., Nashville, TN, ²Vanderbilt Genetics Inst., Nashville, TN

Abstract Body:

Clonal hematopoiesis of indeterminate potential (CHIP) is a common age-related condition characterized by an acquired somatic mutation in a hematopoietic stem cell that confers a growth advantage. This “driver” mutation can lead to increased proliferation relative to other hematopoietic stem cells, which results in a distinct sub-clonal blood cell population that increases risk of developing blood cancer, cardiovascular disease, and death. While DNA biorepositories have the ability to identify large cohorts of individuals with CHIP, they typically only contain blood from a single timepoint, limiting the ability to infer how CHIP clones changes over time. Additionally, clones with faster expansion rates have been linked to more detrimental outcomes, but little is known about why certain clones grow faster than others. In this study, we utilized multi-timepoint samples from 103 individuals with CHIP in Vanderbilt’s biorepository (BioVU) to characterize clonal behavior over time. We used the linked electronic health record to exclude patients with prevalent hematologic malignancy. In this cohort, the driver mutations occurred 47% of the time in *DNMT3A*, 23% of the time in *TET2*, and 9% of the time in *JAK2*. 32 of these individuals had more than one CHIP driver mutation. The mean difference in time between the two timepoints was 5.2 years (SD=2.9). Surprisingly, we observed both clonal expansion and clonal shrinkage across the timepoints with 30% of *DNMT3A*, 0.6% of *TET2*, and 46% of *JAK2* clones shrinking over time. The fastest average expansion was seen in *TET2* clones (7% growth/year) and the slowest growth in *JAK2* clones (0.6% growth/year), but there was a significant amount of variation between individuals. The average expansion rate for *DNMT3A* was 2% growth/year, and there were no differences observed between loss of function mutations, missense mutations or *DNMT3A* R882 hotspot mutations. In summary, we leverage electronic health record linked serial biospecimens to quantify clonal expansion rate on a per individual basis. We find significant individual level heterogeneity in CHIP clonal dynamics even amongst individuals with the same CHIP driver mutation. Future work will evaluate how the phenome and drug exposome affects how clones change over time.

Complex Traits Posters - Wednesday

PB1305. Collagen gene cluster expressions role on liver fibrosis in biliary atresia patients

Authors:

F. Gunadi¹, A. Makhmudi¹, K. Iskandar²; ¹Univ.s Gadjah Mada, Yogyakarta, Indonesia, ²UGM Hosp., Yogyakarta, Yogyakarta, Indonesia

Abstract Body:

Background: Liver fibrosis progression might continue even after Kasai surgery in biliary atresia. Our study associated the expressions of the collagen gene cluster, including *COL6A1*, *COL6A2*, *COL6A3*, and *COL1A1*, with liver fibrosis in patients with BA. **Methods:** We extracted total RNA from liver of BA patients and controls, followed by qPCR of *COL6A1*, *COL6A2*, *COL6A3*, and *COL1A1* genes. **Results:** Our study involved 20 patients with BA and 18 controls. A significant downregulated of *COL6A1* (ΔC_T 9.06 ± 2.64 vs. 5.42 ± 2.41 ; $p=0.0009$), *COL6A2* (ΔC_T 8.25 ± 2.07 vs. 5.77 ± 3.51 ; $p=0.02$), *COL6A3* (ΔC_T 11.2 ± 6.08 vs. 6.78 ± 3.51 ; $p=0.024$), and *COL1A1* (ΔC_T 3.26 ± 1.71 vs. 0.19 ± 2.76 ; $p=0.0015$) expressions were noted in liver of patients with BA compared to controls. In addition, there was a significant association between collagen gene cluster expressions and liver cirrhosis ($p=0.0085$, 0.04 , and 0.0283 for *COL6A1*, *COL6A2*, and *COL6A3*, respectively). **Conclusion:** Our study reveals the collagen gene cluster expressions aberrant in patients with BA, implying the role of these gene clusters in the liver fibrogenesis of BA.

Complex Traits Posters - Thursday

PB1306. Colocalization and variant-to-gene mapping nominates pleiotropic bone mineral density effector genes

Authors:

M. Conery¹, J. A. Pippin², W. P. Bone³, Y. Wagley⁴, M. C. Pahl⁵, K. D. Hankenson⁴, S. Grant¹, B. Voight³; ¹Children's Hosp. of Philadelphia/Univ. of Pennsylvania, Philadelphia, PA, ²Children's Hosp. of Philadelphia, Philadelphia, PA, ³Univ. of Pennsylvania, Philadelphia, PA, ⁴Univ. of Michigan, Ann Arbor, MI, ⁵Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

53 million Americans live with reduced bone mineral density (BMD) and elevated risk of osteoporotic fractures, yet the genetic etiology of BMD and how it relates to those of other complex traits is poorly understood. Specifically, while some BMD loci, such as *3q21.1*, have been found to exhibit pleiotropy, or affect multiple phenotypes, it is unknown in general how prevalent pleiotropy is across BMD loci. To elucidate these issues, we used coloc to test for colocalization between BMD signals identified by a genome-wide association study (GWAS) and the GWAS signals of 41 related complex traits. We identified 194 BMD loci that colocalize with one or more other traits, including 58 with height and 50 with serum alkaline phosphatase levels. At colocalizing loci, we nominated possible effector genes using two approaches. First, we used ColocQuiaL to test for colocalization of the BMD associations with Genotype-Tissue Expression (GTEx) v8 expression quantitative trait loci (eQTLs). Second, to supplement the lack of bone eQTLs in GTEx, we used ATAC-seq (Assay for Transposase-Accessible Chromatin) and Capture-C datasets from human mesenchymal stem cell derived osteoblasts to nominate additional genes. We identified open-chromatin regions that contained proxies of the lead, colocalized BMD SNPs and that either overlapped the promoter of an expressed gene or interacted with one or more such promoters. The two methods nominated effector genes at 94 loci including known BMD modulating genes like *SOST* and *ALPL* as well as *FDFT1*, a target of BMD-increasing bisphosphonate therapy. We hypothesize that loci prioritized specifically by the second method mostly mediate their associations via effects on osteoblast expression of nominated genes. We conclude that these observations warrant further investigation, and we have already initiated a framework that uses CRISPR interference in hFOB1.19 cells to knockdown these nominated genes and test for effects on bone deposition with paired ALP and Alizarin Red S Assays. Our approach is implicating new osteoporosis drug targets and providing insight into the mechanisms by which BMD loci exhibit cross-phenotypic effects.

Complex Traits Posters - Wednesday

PB1307. Combining genetic risk with machine learning derived biological age improves prediction of type 2 diabetes

Authors:

J. Leiby¹, M. Shivakumar¹, M. Lee¹, D. Kim¹, E. Choe²; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract Body:

Introduction: Chronological age is a strong risk factor for death and disease. However, individuals can age biologically at different rates. The difference between the two ages, delta age, represents an important feature related to overall health and may have implications for clinical care and disease prevention. In this study, we developed a machine learning model to calculate delta age. We performed association analyses with 108 phenotypes, corroborated by comprehensive health check-ups in a Korean population, including metabolic and lifestyle factors. Lastly, we examined how delta age can be used to stratify an individual's risk of type 2 diabetes (T2D) in combination with their genetic risk for disease. **Methods:** The target dataset in this study was the Korean GENIE cohort, consisting of physiological and genetic data. To generate biological age, we selected healthy individuals from the KOGES dataset (The Korean Genome and Epidemiology Study cohort) and used a random forest model to estimate chronological age. We used the trained model to calculate delta age in the target dataset. Association analyses were performed using logistic or linear regression with 108 phenotypes, correcting for chronological age. For T2D genetic risk, we used LDpred to estimate variant weights from BioBank Japan summary statistics and used these to calculate the polygenic risk score for T2D (PRS-T2D) in the target dataset. To validate PRS-T2D and delta age, we perform logistic regression. To determine if delta age is a significant feature to stratify risk of T2D in combination with genetic risk, we performed logistic regression for T2D within each quantile correcting for covariates and genetic principal components. **Results:** The mean delta age was -2.66 (+/- 5.15). Delta age is significantly associated with 14 phenotypes including T2D. Our models showed that including delta age as a feature to predict instance of T2D significantly increased the performance of the PRS-T2D only model from an AUC of 0.777 to 0.786. In the quantile analysis biological age remains a significant feature for detecting T2D, with an odds ratio of 1.82 ($p=1.48e-2$) in the lowest fifth quantile PRS-T2D, and 1.67 ($p=5.48e-6$) in the highest quantile PRS-T2D. **Discussion:** We used machine learning to calculate biological age and found it is significantly associated with clinically relevant phenotypes. The T2D analysis showed that delta age significantly improved prediction of disease and can be used to further stratify an individual's risk in the highest and lowest genetic risk quantiles. Replication in more diverse datasets will further support the clinical utility of biological age.

Complex Traits Posters - Thursday

PB1308. Combining polygenic and monogenic causes for improved osteoporotic fracture risk prediction.

Authors:

T. Lu¹, V. Forgetta², S. Zhou¹, B. Richards¹, C. Greenwood³; ¹McGill Univ., Montreal, QC, Canada, ²Lady Davis Inst., Montreal, QC, Canada, ³Jewish Gen. Hosp., Montreal, QC, Canada

Abstract Body:

Osteoporosis and osteoporotic fractures affect >200 million people worldwide, cause significant morbidity and mortality, and incur a heavy socioeconomic burden. Predicting bone density by polygenic risk score (PRS) has demonstrated potential utility in refining osteoporotic fracture risk prediction and management that rely on traditional clinical risk factors. However, existing PRSs do not account for rare monogenic causes, because of a lack of power to accurately estimate their effects in previous studies. Leveraging UK Biobank whole-exome sequencing data from 454,787 individuals, we developed a PRS, called ggSOS, that harnesses both common and rare variants to predict heel ultrasound speed of sound (SOS), a quantitative measure of bone density. Using a combination of SnpEff annotation and Combined Annotation Dependent Depletion (CADD) score, we first generated multiple gene burden masks based on predicted functional impacts of detected variants. Next, we assigned European ancestry individuals into a training dataset (N = 317,434) and a test dataset (N = 74,825). In the training dataset, we conducted a genome-wide association study for SOS and developed a common variant-based PRS, aggregating variants with a minor allele frequency (MAF) >0.001. We then regressed measured SOS on this PRS and conducted burden testing for 19,308 RefSeq genes using the residualized SOS, targeting rare pathogenic variants with a MAF ≤0.001. Finally, the common variant-based PRS was additively combined with significant gene burden masks to derive ggSOS, with common-rare interaction effects incorporated via forward selection.

Fourteen genes harbored rare pathogenic variants associated with SOS (p-value <2.5x10⁻⁶). These variants cumulatively affected 6.6% of individuals. Amongst 4,949 carriers in the test dataset, a one standard deviation decrease in ggSOS was associated with 1.52 (1.27-1.81)-fold increased odds of osteoporosis and 1.45 (1.25-1.69)-fold increased hazard of major osteoporotic fracture. Compared to a common variant-based PRS, ggSOS had increased C-indices for predicting osteoporosis (0.625 vs. 0.619) and major osteoporotic fracture (0.620 vs. 0.613) risks. Furthermore, ggSOS demonstrated mildly improved predictive performance in populations of African and East Asian ancestries, but not in populations of South Asian and other admixed ancestries.

In conclusion, identifying monogenic causes may assist polygenic prediction for osteoporotic fracture risk. Nonetheless, the relatively high cost of sequencing, limited number of carriers, and small magnitude of predictive power gain entail careful considerations in research and in clinics.

Complex Traits Posters - Wednesday

PB1309. Common and rare variant associations with spinal stenosis in 29,488 cases.

Authors:

J. Otto, E. Stahl, N. Lin, GHS-RGC DiscovEHR Collaboration, Penn Medicine BioBank, Regeneron Genetics Center, C. Paulding, G. Coppola, K. Praveen; Regeneron Genetics Ctr., LLC, Tarrytown, NY

Abstract Body:

Background: Chronic back pain is a common musculoskeletal problem, affecting up to 7.5-15% of individuals worldwide, and is a leading cause of global disability. Numerous medical conditions contribute to symptoms of back pain, including spinal stenosis. Genetic factors account for 21-67% of the variation in risk for low back pain, and up to 80% of the variation in risk for spinal stenosis, warranting further investigation into the specific gene regions contributing to the development of these traits.

Methods: Case-control status was defined using ICD-10 M48.0 diagnosis codes and included 29,488 cases and 774,235 controls across five cohorts (UK Biobank, Geisinger DiscovEHR, University of Pennsylvania - Penn Medicine BioBank, Mount Sinai's BioMe Personalized Medicine Cohort, and FinnGen). Cross-ancestry genome-wide and exome-wide association analyses (GWAS and ExWAS, respectively) were conducted for single common and rare variants, as well as their gene burdens using REGENIE, and results across cohorts were pooled using inverse-variance weighted meta-analysis.

Results: After correction for multiple testing, 24 genome-wide significant ($P < 5e-08$) loci were identified that included effects of common variants near novel and genes previously identified in GWAS of spinal stenosis and related conditions, such as back pain. Amongst novel loci, we observed a protective association of a missense variant (rs6330; Ala35Val) within the *NGF* gene ($OR = 0.93$; $P = 3.70e-11$). Three genome-wide significant rare variant ($MAF < 0.005$) associations were also observed, including a rare missense variant (rs28931614, Gly380Arg; $MAF = 9.47e-06$) in the *FGFR3* gene ($OR = 71.54$; $P = 1.11e-08$). Gly380Arg is a known pathogenic variant for Achondroplasia, an autosomal-dominant skeletal dysplasia disorder in which spinal stenosis is a common complication. Finally, in gene burden tests, rare ($MAF < 0.001\%$) predicted loss-of-function and deleterious missense variants in a gene involved in the regulation of skeletal development pathways were significantly associated with increased risk of spinal stenosis ($OR = 2.49$; $P = 6.40e-09$).

Discussion: The current study represents the largest investigation of genetic factors underlying spinal stenosis to date. Identified genes and variants likely represent biological pathways involved in both cause (e.g., osteoarthritis) and consequence (e.g., reported back pain) of spinal stenosis. Given *NGF*'s role in mediating chronic pain response and osteoarthritis disease progression, future studies might seek to elucidate whether the current genetic signal reflects its role in disease pathophysiology or pain outcomes.

Complex Traits Posters - Thursday

PB1310. Comparison of clinical characteristics among survivor and non-survivor of covid-19 infected patients: Scoping review and meta-analysis.

Authors:

M. Adebisi^{1,2,3}, O. Ropo⁴, A. Adebisi^{1,2,3}, A. Shekari^{5,2}, M. Enwere^{5,2}; ¹Computer Sci. Dept., Coll. of Pure and Applied Sci., Landmark Univ., Omu-aran, Nigeria, ²Covenant Applied Informatics and Communications- Africa Ctr. of Excellence (CAPIIC-ACE), Ota, Nigeria, ³Landmark Univ. SDG 3: Good Hlth.and Well-being, Omu-aran, Nigeria, ⁴Biostatistics Unit, Discipline of Publ. Hlth.Med., Sch. of Nursing and Publ. Hlth., Univ. of KwaZulu-Natal (UKZN), Durban, South Africa, ⁵Dept. of Computer and Information Sci., Coll. of Sci. and Technology, Covenant Univ., Ota, Nigeria

Abstract Body:

The Coronavirus Disease of 2019 was identified as an emerging infectious disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV 2) and was named COVID-19 by the WHO. This study used a meta-analysis approach to explore the proportion of systemic disorder with the development of COVID-19 infection. The aim is to summarize recent evidence for the prevention of COVID-19 by published studies on patients with severe COVID-19 infection. Literature search of databases using PubMed and Web-of-Science till 4th April 2021 as search engines were adjusted to identify relevant studies. A meta-analysis of studies pooled prevalence of clinical symptoms. Random-effects models were used, heterogeneity and various meta-analyses were carried out for four systemic disorders using Stata version 16.0. Of the 599 patients included in this analysis, compared with the non-survivor patient, the pooled odds ratio (OR) of Cough, Breathlessness and Fever in survivor patients were (OR: 11.2%, 95% CI: 8.2-15.2%) and (OR: 10.5%, 95% CI: 3.8-13.8%), while that of Myalgia and dyspnoea are also crucial. There is a significant heterogeneity for clinical symptoms ($I^2 = 56\%$). Re-affirming the most prominent clinical-symptoms proportion after One and a half years of COVID-19 existence, cough, breathlessness and fever still remains the most-common symptoms in COVID-19 infected patients.

Complex Traits Posters - Wednesday

PB1311. Concordance of directional effects between sexes suggests pervasive epistasis and omnigenic regulation of complex metabolic traits in mice

Authors:

A. Miller¹, C. Pan², A. J. Lusis², D. C. Crawford¹, D. A. Buchner¹, S. M. Williams³; ¹Case Western Reserve Univ., Cleveland, OH, ²UCLA, Los Angeles, CA, ³Case Western Reserve Univ, Cleveland, OH

Abstract Body:

Epistasis in humans has proven difficult to investigate in part due to limited statistical power, relatively small sample sizes, and computational demands. The well-characterized inbred strains of the Hybrid Mouse Diversity Panel (HMDP) offer a unique and powerful alternative to systematically test for non-linear epistasis in a mammalian system. We used previously collected data from over 100 inbred strains from the HMDP that were genotyped using the Mouse Diversity Genotyping Array encompassing >20,000 polymorphic loci and phenotyped for 8 complex traits related to obesity and type 2 diabetes. Using FaST-LMM and Bonferonni significant thresholds corrected for the number of haplotype blocks, we analyzed the HMDP for both main effects (additive) and epistatic interactions that regulate these traits. In females, no main effects were detected whereas in male mice, main effect QTLs were identified for adiposity (n=58), body weight (n=22), and fat mass (n=60). Interaction effects were identified in females for fat mass (n=8) and in males for adiposity (n= 56,112), body weight (n=66,296), fat mass (n=111,749), cholesterol (n=57), and triglycerides (n=123). The numbers of interactions at first seems extraordinary but are in line with the number of QTLs discovered for main effects, given the orders of magnitude more SNP-SNP interactions assessed relative to main effects. The direction of these significant effects in males (i.e. increased or decreased trait values) were concordant with the direction of effect in females, even when the effects were not statistically significant. Between 63 and 100% of main effect directions and 79 and 100% of interaction effects showed the same direction. This concordance rate is substantially greater than the 50% expected by chance, validating these interactions. Remarkably, this shared directionality of effects between males and females was also detected for non-statistically significant main effect and interacting loci until p-values ~ 0.8. This finding is consistent with an omnigenic model, where most loci contribute to the regulation of these metabolic traits with both additive and epistatic effects. These data from a mammalian model organism support an omnigenic inheritance pattern comprised of pervasive epistasis, and indicate that interactions may be critical to understanding the inheritance of complex traits in humans.

Complex Traits Posters - Thursday

PB1312. Considerations in developing a framework for defining clinical actionability of polygenic risk scores

Authors:

J. Hunter¹, A. Buchanan², J. Bulkley³, M. Garcia-Closas⁴, K. Goddard⁵, S. Massimino⁶, B. Powell⁷, I. Kullo⁸, ClinGen; ¹RTI Intl., Durham, NC, ²Geisinger, Danville, PA, ³Kaiser Permanente Ctr. for Hlth.Res., Portland, OR, ⁴Natl. Cancer Inst, Rockville, MD, ⁵Natl. Cancer Inst., Bethesda, MD, ⁶Kaiser Permanente Northwest, Portland, OR, ⁷Univ North Carolina at Chapel Hill, Chapel Hill, NC, ⁸Mayo Clinic, Rochester, MN

Abstract Body:

There is substantial ongoing effort to generate and validate polygenic risk scores (PRS). Here we focus on their use to identify individuals at modified risk for developing a health condition where changes in management may improve health outcomes. The ClinGen Actionability Working Group developed a standardized process for assessing clinical actionability of monogenic conditions, and is adapting this framework for PRS. We consider how assessment of clinical actionability may differ for the PRS context, which have population-scale impact, compared to rare monogenic conditions, and whether risk information is returned as primary or secondary genomic findings. For the four domains of clinical actionability, points of consideration include:

Domain 1: Severity of the Outcome. Similar to the monogenic framework, we consider the impact of the health condition on individual-level morbidity and mortality. The polygenic adaptation also accounts for population impact based on health condition prevalence.

Domain 2: Likelihood of the Outcome. As with the monogenic framework, we identify estimates of absolute or relative risk. For the polygenic adaptation, we also account for variability in risk across population groups, an important consideration for all PRS.

Domain 3: Effectiveness of the Intervention. Similar to the monogenic framework, we seek recommendations for clinical interventions in practice guidelines. Since PRS-based guidelines may not be available, we use guidelines based on risk thresholds even if not specific to monogenic or polygenic sources of risk, and evidence in primary literature. For the polygenic adaptation, we consider whether evidence of effectiveness differs across population groups. To account for the population-scale impact of implementation, we summarize cost-effectiveness evidence.

Domain 4: Nature of the Intervention. We anticipate minimal adaptation of this domain for PRS, since it is not specifically related to the underlying etiology of the condition.

We will pilot test the draft PRS clinical actionability framework using exemplar health conditions, define a protocol to identify and curate evidence, and use an iterative approach to refine the framework.

Complex Traits Posters - Wednesday

PB1313. Construction of clinically significant mutation map and their spatio-temporal single cell transcriptomics trajectories in congenital heart disease

Authors:

R. Tambi¹, S. Nandkishore¹, S. Sharafat¹, F. Kader¹, N. Nassir¹, B. Ashraf¹, N. Mohamed¹, A. Ahmed¹, R. Hameid¹, A. Alsheikh-Ali^{1,2}, R. Donato³, M. Uddin⁴, B. Berdiev¹; ¹Mohammed Bin Rashid Univ. Of Med. and Hlth.Sci., Dubai, United Arab Emirates, ²Dubai Hlth.Authority, Dubai, United Arab Emirates, ³Al Jalila Children's Hosp., Dubai, United Arab Emirates, ⁴Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates

Abstract Body:

Congenital heart disease (CHD), one of the most prevalent neonatal congenital anomalies, involves a spectrum of birth defects affecting the normal functions of heart. The genetic component of CHD plays a critical role, yet the pathophysiological mechanisms remain unknown. In this study we aim to identify putative candidate risk genes to CHD and analyze the heterogeneity of these genes across cell types using spatiotemporal heart single cell transcriptome data. We have conducted rigorous literature review and collected 16,350 single nucleotide variants (SNVs) impacting 8,308 genes in 3194 CHD cases that were reported in 31 genomic studies on CHD published between January 2000 and December 2020. We have used ANNOVAR for functional annotation and after quality control (removed 891 synonymous mutations) retained only 16206 exonic/splice site rare variants (Exac_all < 0.01) for downstream analysis. The mutations were classified as benign/likely benign (0.1%), pathogenic/likely pathogenic (P/LP - 7.0%) and variance of uncertain significance (VOUS - 92.9%) following the ACMG classification guidelines. 83% of the filtered mutations were loss of function (LOF), and rest 17% were missense. In our data, 16% of these mutations were of *denovo* origin. In order to identify the candidate risk genes, we scanned the 15306 P/LP and VOUS mutations and filtered the LOF recurrent mutations which impacted 257 genes. We identified highly constraint (pLI > 0.9) 12 LOF genes with significant enrichment of pathogenic mutations, including seven known CHD risk genes: *NOTCH1*, *ANKHD1*, *TNS3*, *ARHGEF12*, *RANBP9*, *DLC1* and *FLT4*; and five novel candidate CHD risk genes *MAG11*, *MTMR12*, *SH3RF3*, *SLC12A2*, and *EDC4*. These CHD risk genes were also found to be significantly enriched in the organ development and signaling network pathways. Our analysis from spatio-temporal single cell transcriptomics analysis show *ARHGEF12*, *MAG11* and *RANBP9* genes are highly upregulated in cardiomyocyte cell types in fetal developmental period. Our study identifies the core CHD genes and maps their cellular heterogeneity, that will require further investigation to pin point cellular level causation into the aetiology of CHD.

Complex Traits Posters - Thursday

PB1314. Construction of mutational landscape and characterization of associated genes in a Bangladeshi cohort of neurodevelopmental disorders

Authors:

H. Akter^{1,2}, M. Rahman³, S. Sarker^{4,5}, M. Basiruzzaman^{4,6}, M. Chowdhury⁴, M. Rahaman¹, T. Eshaque¹, R. Mim¹, M. Morshed¹, N. Jahan⁶, A. Islam^{1,7}, A. Mondol⁶, G. Kundu⁸, S. Kanta⁵, M. Rahman⁹, M. Hosen², Z. Ashraf¹⁰, N. Nassir¹¹, A. Nabi², K. Uddin^{1,12}, M. Uddin^{10,7}; ¹Genetics and Genomic Med. Ctr., NeuroGen Hlth.care, Dhaka, Bangladesh, ²Dept. of Biochemistry and Molecular Biology, Univ. of Dhaka, Dhaka, Bangladesh, ³Dept. of Paediatric Neurology, Bangabandhu Sheikh Mujib Med. Univ., Dhaka, Bangladesh, ⁴Dept. of Child Neurology, NeuroGen Hlth.care, Dhaka, Bangladesh, ⁵Dept. of paediatric neuroSci., Dhaka Shishu Hosp., Dhaka, Bangladesh, ⁶Ctr. for Precision Therapeutics, NeuroGen Hlth.care, Dhaka, Bangladesh, ⁷Cellular Intelligence Lab, GenomeArc Inc., Toronto, ON, Canada, ⁸Dept. of Child Neurology, Bangabandhu Sheikh Mujib Med. Univ., Dhaka, Bangladesh, ⁹Dept. of Pharmaceutical Sci., North South Univ., Dhaka, Bangladesh, ¹⁰Coll. of Med., Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates, ¹¹Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates, ¹²Dept. of Biochemistry, Holy Family Red Crescent Med. Coll., Dhaka, Bangladesh

Abstract Body:

Background: Short nucleotide variations (SNVs) and copy number variations (CNVs) play a critical role into the pathogenesis of neurodevelopmental disorders (NDD) among children. In this study, we aimed to identify clinically relevant variants, genes and their phenotypic characteristics in an ethnically underrepresented homogenous NDD patients of Bangladesh.

Methods: We have conducted WES (33.05 Mb human coding regions) and genome-wide CMA (642,824 probes spanning the genome) analysis for 545 NDD patients to identify SNVs and CNVs (deletion/duplication/translocation and rearrangements). We have used human genome build GRCH38/UCSC hg38 as reference. Variant classification analysis was conducted based on the American College of Medical Genetics (ACMG) guidelines.

Results: The cohort comprises 68.12% male and 31.88% female, with male to female ratio of 2.1:1. Of the 545, 44.95% (245), 21.83% (119) and 33.22% (181) patients' genomic analysis were done using CMA, WES and combined method (CMA plus WES) respectively. The diagnostic yield by only CMA, only WES and combined for broader NDD is 14.69% (36/245), 31.09% (37/119) and 37.57% (68/181) respectively. In the full cohort, further burden test identified females are significant carriers of pathogenic CNVs in comparison to males (OR=2.5; p=1.98×10⁻⁵). In the CMA cohort, we have found 3.27% (8/245) pathogenic CNVs located at the subtelomeric regions. Autism Spectrum Disorder (ASD) subset (91) of CMA cohort shows severe social communication deficit (p=0.014) and overall ASD symptoms severity (p=0.026) among the patients carrying duplication CNV compared to the CNV negative group. Candidate gene analysis identified Critical Exon Gene (CEG) PSMC3 as a potential candidate gene for ASD. Moreover, we hypothesized that KMT2B gene duplication might be associated with intellectual disability. Our custom pathway enrichment analysis of the impacted genes of all pathogenic variants identified GO:0019787, GO:0099003, GO:1901214 and GO:0099536 pathways to be highly significant (FDR P<6.53 × 10⁻⁷), (FDR P<7.1 × 10⁻⁵), (FDR P< 5.34 × 10⁻⁷) and (FDR P<1.72× 10⁻⁴) after correction for multiple tests.

Conclusions: Our results show the utility of WES and CMA for precise genetic diagnosis and its integration into the diagnosis therapeutics and management of NDD patients. To our knowledge, this is the first genomic research using larger number of NDD patients of Bangladesh.

Keywords: Neurodevelopmental Disorders (NDD); Short nucleotide variations (SNVs); Chromosomal

ASHG 2022 Annual Meeting Poster Abstracts

Microarray Analysis (CMA); Copy number variation (CNV); Autism Spectrum Disorder (ASD); Critical Exon Gene (CEG); Autism Spectrum Disorder (ASD)

Complex Traits Posters - Wednesday

PB1315. Copy number variants and mental health in a community pediatric sample

Authors:

M. Zarrei¹, C. Burton¹, **W. Engchuan**¹, D. Merico^{2,1}, E. Higginbotham¹, J. Wei¹, S. Shaikh¹, N. Roslin¹, J. MacDonald¹, G. Pellecchia¹, T. Nalpathamkalam¹, S. Lamoureux¹, R. Manshaei¹, J. Howe¹, B. Trost¹, B. Thiruvahindrapuram¹, R. Yuen^{1,3}, J. Vorstman^{1,3}, C. Marshall^{1,3}, R. Wintle¹, L. Strug^{4,3}, D. Stavropoulos^{1,3}, P. Arnold⁵, M. Woodbury-Smith^{6,1}, J. Crosbie^{1,3}, R. Schachar^{1,3}, S. Scherer^{1,3}; ¹The Hosp. for Sick Children, Toronto, ON, Canada, ²Deep Genomics Inc., Toronto, ON, Canada, ³Univ. of Toronto, Toronto, ON, Canada, ⁴Hosp. for Sick Children, Toronto, ON, Canada, ⁵Univ. of Calgary, Calgary, AB, Canada, ⁶Newcastle Univ., Jesmond, Newcastle upon Tyne, United Kingdom

Abstract Body:

‘Spit for Science’ is a cohort of 22,515 youth (4-18 years of age) in whom we collected neurodevelopmental traits measurements and DNA extracted from saliva. Quantitative phenotype data were collected and trait scores were generated for attention deficit hyperactivity disorder (ADHD), obsessive compulsive disorder (OCD), and response inhibition using Stop-Signal Reaction Time (SSRT, a cognitive biomarker for neurodevelopmental traits). We genotyped 7,100 unrelated individuals of European (n=5,686) and East Asian (n=1,414) ancestry using either the Illumina HumanCoreExome or Illumina Global Screening Array. Among these, 1,233 (17.5%) participants self-reported a diagnosis of a neurodevelopmental disorder (NDD; e.g. autism, OCD, ADHD), and 2,412 (34.2%) participants had either a self-reported diagnosis of NDD or high quantitative scores on NDD trait measures. We performed global and functional gene-set burden analyses, and locus association test separately for East Asian and European participants on rare (<0.5%) and less-rare (1-5%) CNVs using linear regression with different traits. We identified CNVs with clinical relevance to NDDs in 279 (3.9%) participants, and found elevated standardized T-scores for all traits in this group of participants. These CNVs included i) sex-chromosome and autosomal aneuploidies in 7 (0.1%) individuals, ii) 13 genomic loci previously implicated in NDD in 48 (0.7%) participants, and iii) CNVs affecting genes with known brain functions such as *ASTN2*, *DLGAP2*, *MACROD2*, *NRXN1*, and others. Individuals carrying these CNVs had three-fold increases in having self-reported or community diagnoses of NDD (odds ratio: 3.1, $p=10^{-5}$). Rare deletions were globally associated with inattention traits in Europeans ($\beta=0.15$, $p=0.04$). In Europeans, we found an association of ADHD and inattention traits with increased burden of rare deletions impacting genes that are highly expressed in brain or involved in nervous system development or brain function. In East Asians, ADHD and inattention were associated with less-rare deletions for synaptic and neurofunctional genes. Also, within East Asian participants, we observed an association of response inhibition (cognitive biomarker) with rare duplications in a gene linked with brain function. These results highlight the importance of assessing the prevalence and impact of rare and less rare variants in community samples to provide a baseline to improve understanding of pediatric mental health.

Complex Traits Posters - Thursday

PB1316. Coronary artery disease risk assessment in Asian Indian Punjabis using a genome wide polygenic risk score

Authors:

M. Rout; Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK

Abstract Body:

BACKGROUND: Polygenic risk score (PRS) has been shown to be highly effective in predicting coronary artery disease (CAD) risk in Western European populations. However, such studies on South Asians are scarce despite the fact that they account for 50% of the global burden of CAD. In this study, we evaluated the predictive efficacy of PRS derived from the Asian Indian (AI) ancestry-specific score (PRSAI) and European-derived PRS (PRSEU). Also, we compared these with the clinical risk score (CRS).

METHODS: The study used 4602 participants (791 CAD cases and 3790 controls) from the Asian Indian Diabetic Heart Study/Sikh Diabetes Study. Weighted PRS was constructed using 100 significant SNPs from our Punjabi/Sikh CAD GWAS and 75 SNPs identified from the European GWAS catalog. The CRS was derived using the clinical risk factors described for the Framingham risk score.

RESULTS: Ancestry-specific PRSAI showed an enhanced efficacy of over 30% in estimating the relative risk for CAD over PRSEU. In sensitivity analysis, the area under the curve (AUC) for PRSAI and PRSEU were 0.84 and 0.72, respectively, while the AUC remained unchanged on combining the PRS (AI+EU), i.e., 0.84. PRSAI also predicted the risk for increased waist to hip ratio (Beta=0.11, $p=4.1 \times 10^{-13}$) in both gender, while fasting glucose levels (Beta=0.08, 3.0×10^{-3}) were confined to females.

CONCLUSIONS: The results highlight evidence for the utility of PRS for identifying genetically predisposed high-risk individuals and attest to its broader clinical value. Also, sex differences may play a role in determining the risk for CAD, and increased fasting glucose and type 2 diabetes (T2D) are known to increase the risk of heart disease in women more than men, especially after menopause.

Complex Traits Posters - Thursday

PB1317. Cross-tissue single-nucleus RNA-sequencing reveals distinct tissue-specific expression profiles of adipocytes, depending on their residency in human adipose tissue or heart

Authors:

S. Das¹, S. Rajkumar¹, S. Lee¹, M. Alvarez¹, A. Koka¹, K. Pietiläinen², P. Pajukanta^{1,3}; ¹Univ. of California Los Angeles, Los Angeles, CA, ²Univ. of Helsinki, Helsinki, Finland, ³Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract Body:

Understanding the molecular mechanisms of complex diseases needs extensive characterization of cell-types across heterogenous tissues. Adipocytes have been recognized as a vital player in metabolic regulation. Furthermore, dysfunctional adipocytes confer complex cardiovascular diseases (CVDs). Although adipose tissue is the predominant site of adipocytes, they are also residents of other tissues, including the myocardial tissue of each heart chamber i.e., left atrium (LA), left ventricle (LV), right atrium (RA), and right ventricle (RV). However, it is not known whether gene expression profiles of adipocytes differ depending on whether they reside in the adipose tissue or heart. To study the tissue-specific profiles of adipocytes and their associations with cardiovascular disease, we investigated single-nucleus RNA sequencing data from a human myocardial tissue cohort (n=7) (Tucker et al., 2020) and subcutaneous adipose tissue (n=15). We clustered the adipose tissue droplets using Seurat and assigned them into 11 main adipose cell-types, including adipocytes, using SingleR. For myocardial tissue, we clustered droplets of each chamber separately, and then annotated the clusters of each chamber into 10 main heart cell-types, including adipocytes. Identification of adipocyte marker genes (AMGs) and further comparison between adipose and each of the heart chambers resulted in 655 tissue-specific AMGs (162 in LA, 170 in LV, 139 in RA, 63 in RV, and 121 in adipose tissue, respectively), and in 291 shared AMGs. Furthermore, KEGG pathway analysis showed that the heart-specific AMGs of all four chambers are significantly enriched for the propanoate metabolism pathway (FDR<0.05) that is not enriched among the adipose tissue -specific AMGs. Three genes (ACSS2, ECHDC1, and PCCA) of this pathway are shared across all chambers, among which variants in PCCA and ECHDC1 predispose to dilated cardiomyopathy. Adipose tissue-specific AMGs are in turn enriched for metabolic pathways, such as the PPAR signaling pathway, and regulation of lipolysis in adipocytes. Using the GTEx data, we identified cis-expression quantitative trait loci (eQTL) variants, associated with CVD traits, regulating the shared and tissue-specific AMGs across both tissues. Noteworthy, the RA-specific AMGs are also enriched for coronary artery disease and heart failure GWAS SNPs, and the LA-specific AMGs for heart failure GWAS SNPs using Magenta (FDR<0.05). Overall, our results provide insights into the distinct, tissue-specific expression profiles of heart and adipose tissue resident adipocytes, pathways involved in their tissue-specificity, and their genetic links to CVDs.

Complex Traits Posters - Wednesday

PB1318. Curation burden mitigation: strategies for empowering high throughput gene panel analysis using whole genome sequencing

Authors:

S. Strom¹, L. M. Amendola¹, A. Coffey², C. Brown³, J. Lowry⁴, A. Kesari¹, J. AVECILLA¹, S. Ajay¹, R. J. Taft⁵, D. Perry³; ¹Illumina Inc., San Diego, CA, ²Illumina Inc, Great Thurlow, United Kingdom, ³Illumina, San Diego, CA, ⁴Illumina Inc., Salt Lake City, UT, ⁵Illumina Inc, San Diego, CA

Abstract Body:

Background: Whole genome sequencing (WGS) is an ideal platform for high throughput screening of gene panels due to the simplicity of the wet laboratory preparation, uniformity of data, access to multiple variant types, and unbiased genome-wide coverage of genetic loci. Without adequate preparation, however, each test can result in a high curation burden with many rare variants present per test subject. As only pathogenic/likely pathogenic variants are typically reported in screening settings, pre-curation and variant filtering strategies are critical for efficiency. We have developed a scalable clinical WGS test for patients with CVD that includes a CVD gene panel. Here we discuss our bioinformatics approach to reducing this workload while maintaining high accuracy. **Methods:** We first reviewed a set of genes potentially associated with CVD for evidence supporting a Definitive or Strong -gene-disease relationship (GDR). For 182 genes with at least one Strong or Definitive GDR CVD-associated condition, we performed a literature review to estimate the prevalence and penetrance of that condition. We then calculated population minor allele frequency (MAF) cutoffs for auto-classification of benign (B), likely benign (LB) and VUS variants. Variants in these genes present in the gnomAD database with overall MAF <5% were then subjected to auto-scoring and used to populate a variant database. Heuristic filters were developed to exclude variants with MAF >5% and variants unlikely to impact the protein sequence. These filters were tested on a 20 negative and 1 positive WGS datasets. **Results:** A total of >7.4 million variants from 182 CVD-associated disease genes with MAF <5% were scored (~4.4M B/LB and ~3M VUS). Across the 21 WGS case datasets, the average number of variants requiring review per case was 3.2 (range 0-11 LB/VUS), with 6 cases (28.5%) having zero variants requiring review. The expected variant in the known positive case was correctly identified. **Conclusions:** Pre-evaluation of GDRs and automated exclusion of B/LB/VUS variants supports scalable reporting of WGS-based gene panels. If done on an *ad hoc* basis, the required curation could extend turnaround time and overall interpretation cost. Screening of a large cohort is planned as a to verify these finding and to support deployment of this test.

Complex Traits Posters - Thursday

PB1319*. Cytokine treatment increases *APOEε4* expression in reactive astrocytes in European local ancestry but not in African local ancestry

Authors:

O. Oron¹, M. Vasquez¹, A. J. Griswold¹, B. DeRosa¹, F. RAJABLI¹, K. Celis¹, B. Feliciano-Astacio², G. W. Beecham¹, L. D. Adams¹, C. Silva², S. J. Tejada¹, P. Mena¹, P. Whitehead¹, K. Hamilton-Nelson¹, T. Starks³, G. S. Byrd³, M. L. Cuccaro¹, j. young¹, M. Pericak-Vance¹, J. Vance¹, D. Dykxhoorn¹; ¹John P. Hussmann Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, ²Dept. of Internal Med., Univ. Central Del Caribe, Bayamón, Puerto Rico, ³Maya Angelou Ctr. for Hlth.Equity, Wake Forest Sch. of Med., Winston-Salem, NC

Abstract Body:

Background: Local genomic ancestry (LA) determines the difference in Late-Onset Alzheimer's Disease (LOAD) risk between European (EU) and African (AF) carriers of *APOEε4/4*. Single-Nuclei RNA sequencing (snRNA-seq) from frontal cortex suggest that this risk difference is at least partially explained by increased *APOEε4* expression in *APOEε4/4* homozygotes with EU LA compared to *APOEε4/4* homozygotes with AF LA. In some of the EU LA patients, an astrocyte cluster with the highest *APOEε4* expression also expressed a signature of genes consistent with A1 Reactive Astrocytes (A1RA).

Understanding if astrocyte reactivity is a precursor for an increase in *APOEε4* expression in specific astrocyte subtypes and contributes to LOAD risk is crucial for identifying new therapeutic targets in AD.

Methods: To determine if cytokine treatment affects *APOEε4* expression in astrocytes or astrocyte subtypes, iPSC lines were derived from *APOEε4/4* patients with either EU LA (n=1) or AF LA (n=1) and differentiated into iPSC-derived astrocytes (iAstrocytes). Mature (days in-vitro 49) iAstrocytes were treated for 24 hours with a cytokine cocktail (IL-1a, hTNFα and C1q) to induce A1RA and then subjected to Single-Cell RNA sequencing (scRNA-seq) along with control iAstrocytes. scRNA-seq was performed with the 10X Chromium Next GEM Single Cell 3' v3.1 kit and sequencing on the Illumina NovaSeq 6000. Following raw data processing via CellRanger v6.0 the data was analyzed using the Seurat v4.0 pipeline to identify differences in expression between treated and untreated iAstrocytes carrying either EU or AF LA.

Results: Overall, 29,836 cells were sequenced at a median depth of ~135,000 reads per cell detecting on average ~5900 genes per cell. Integration of all four samples at a resolution of 0.4 in the Seurat FindClusters command resulted in 9 distinct clusters. The EU LA *APOEε4/4* carrier had a significant increase in *APOEε4* expression in 7 out of the 9 clusters upon cytokines treatment compared to control. Moreover, in these 7 clusters, multiple A1RA biomarkers (*C3*, *IFITM3*, *IL6*, *GBP2*, *IL1R1* and *ICAM1*) also increased in expression. Conversely, in the AF LA *APOEε4/4* carrier, no clusters showed increased *APOEε4* expression but did show increased expression of the A1RA biomarkers *C3*, *IL6* and *ICAM1*.

Conclusions: In this preliminary study, we suggest a relationship between cytokine induced increase in *APOEε4* and A1RA gene expression in EU LA *APOEε4/4* carriers but not in AF LA *APOEε4/4* carriers. This suggests a difference in molecular response to inflammatory cues depending on *APOEε4/4* local ancestry which may modulate the differences in AD risk.

Complex Traits Posters - Wednesday

PB1320. Data-driven Clustering of Human Cardiovascular Disease (CVD) Related Variables

Authors:

M. Mandal¹, J. Levy², C. Ives³, S. Hwang⁴, Y. Zhou⁵, A. Motsinger-Reif⁶, H. Pan⁷, W. Huggins⁸, C. Hamilton⁷, F. Wright⁹, S. Edwards¹; ¹RTI Intl., RTP, NC, ²Levy Informatics, RTP, NC, ³RTI Intl., Apex, NC, ⁵North Carolina State Univ., Raleigh, NC, ⁶Natl. Inst. of Environmental Hlth.Sci., Durham, NC, ⁷RTI Intl., Durham, NC, ⁸RTI Intl., Research Triangle Park, NC, ⁹North Carolina State Univ, Raleigh, NC

Abstract Body:

While large nationwide studies such as All of Us represent a new paradigm for understanding human health and disease, there is a wealth of existing data that can provide valuable insights potentially missed in the larger cohorts. This study aims to programmatically identify related variables within each study, thus streamlining the integration of data from existing human cohorts.

The Atherosclerosis Risk in Communities (ARIC) study was used to evaluate automated methods for identifying biologically relevant groups of variables. First, correlations between ARIC variables were calculated and used to assemble the variables into clusters. Cluster quality was evaluated via a manual review of the 391 clusters generated from 3,285 variables. Based on the results of the manual review process, 95% of the clusters containing more than one variable were determined to be related to some degree. To assess the possibility of confirmation bias among the manual reviewers, natural language processing (NLP) was used to generate a semantic similarity score for each cluster. There was a good correspondence between the reviewer evaluation and the semantic similarity score with most exceptions explained by the limitations of the NLP. Second, a network of variable clusters was created based on the average inter-cluster correlations. Community-finding algorithms were used to group the variable clusters into higher order communities. These communities were evaluated for biological coherence and used to identify the major themes addressed by the study.

We have developed a fully automated workflow that hierarchically groups variables from the ARIC study. Manual evaluation of this automated grouping confirmed the biological relevance of individual clusters and showed that the hierarchy corresponds well with the themes expected from the original study design. This workflow is intended as an aid to human annotators when performing data harmonization and curation on existing studies. Since the variables and clusters from the data driven approach captured the major themes of the study, this should also reduce the overall number of variables considered by curators. Streamlining this labor-intensive step should unlock a wealth of existing data for new analyses to complement ongoing efforts in new data collection.

Complex Traits Posters - Thursday

PB1321. Data-driven genome-wide association study of longitudinal pain and mobility trajectories among cases of Parkinson's disease from the Fox Insight Data Exploration Network

Authors:

S. Liu¹, D. D. Gunzler^{2,1}, S. A. Gunzler³, D. C. Crawford^{1,4,5}, F. Briggs¹; ¹Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, ²Ctr. for Hlth.Care Res. and Policy, Sch. of Med., Case Western Reserve Univ., Cleveland, OH, ³Neurological Inst., Univ. Hosp. Cleveland Med. Ctr. and Case Western Reserve Univ. Sch. of Med., Cleveland, OH, ⁴Cleveland Inst. for Computational Biology, Cleveland, OH, ⁵Genetics and Genome Sci., Case Western Reserve Univ., Cleveland, OH

Abstract Body:

Parkinson's disease (PD) is the second most common progressive neurodegenerative disease. It affects movement control and leads to both motor and non-motor symptoms. There is no cure or neuroprotective medication to slow PD progression, which is heterogeneous and unpredictable. Genome-wide association studies (GWAS) have identified ~90 risk variants, but these loci are not associated with pain, a PD non-motor symptom that broadly impacts daily functioning, or mobility deficits, a common PD motor symptom that causes disability. To identify genetic variants specific to pain or mobility deficit trajectories, we accessed longitudinal patient-reported measures from the Fox Insight Data Exploration Network (Fox DEN) linked to 23andMe genome-wide data. FoxDen PD participants (n=37,769) are on average 66 years of age, European descent (96%), and male (~65%). Self-reported pain and mobility deficits from routine longitudinal assessments using the questionnaire items were available for 8,612 persons with PD (PWP) early in their disease course (<3 years). Five clusters were obtained with the latent class growth analysis leveraging the longitudinal patient-reported outcome of pain for the 8,612 PWP. For mobility deficits, we identified four clusters. Among those with genotypes, we stratified the data by genotyping platform (v4, a fully customized Illumina array, n = 443; v5, a customized Illumina Infinium Global Screening Array, n=3,696) for quality control and imputation using the TOPMed reference panel. We performed tests of association using logistic regression comparing the cluster of PWP of very low pain (n=977, 22.1%) with the cluster experiencing severe pain (n=105, 2.4%), and the cluster of PWP of low mobility deficits (n=1,743, 39.5%) with that of severe mobility deficits (n=137, 3.15%), adjusting for age at recruitment, sex, and 20 principal components. We did not identify associations at genome-wide significance. The most significant association identified for pain was rs72794357 (OR=5.1; 95% CI=2.8, 9.6; p=3.0x10⁻⁷), an intronic variant of *MAPK8* (this gene has been associated with opioid dependence), and rs11047777 (OR=3.5; 95% CI=2.2, 5.6; p=5.2x10⁻⁸) for mobility deficits, an intronic variant of *IRAG2*. A multinomial GWAS comparing all the clusters constructed is ongoing, as are gene-based and pathway enrichment analyses. The successful identification of genetic variants associated with subgroups from their longitudinal trajectories has the potential to lead to a better understanding of the mechanisms underlying PD progression, providing potential novel targets for therapeutics.

Complex Traits Posters - Wednesday

PB1322. *De novo* CNVs impacting constraint overlapping genes in neurodevelopmental disorders and congenital anomalies.

Authors:

S. Safizadeh Shabestari¹, N. Nassir¹, S. Sopariwala², I. Karimov³, R. Tambi¹, B. Zehra¹, N. Kosaji¹, H. Akter⁴, B. K. Berdiev¹, M. Uddin^{1,5}; ¹Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates, ²Univ. of Guelph, Guelph, ON, Canada, ³Univ. of Bremen, Bremen, Germany, ⁴NeuroGen Hlth.care Ltd., Dhaka, Bangladesh, ⁵GenomeArc Inc, Toronto, ON, Canada

Abstract Body:

Background: Neurodevelopmental disorders (NDDs) and congenital anomalies (CAs) are a collection of rare disorders with complex etiology. While both conditions have strong genetic component and phenotypes often co-manifest in patients, their underlying genomic overlap is less explored. In this study, we investigated the genomic overlap of copy number variants (CNVs) in two large cohorts of NDD and CA patients to identify *de novo* CNVs and candidate genes that are strongly associated to both NDD and CA phenotypes. **Methods:** We analyzed clinical microarray CNV data from 10,620 NDD and 3,176 CA cohorts, recruited from Sickkids Hospital, Toronto and Credit Valley Hospital, Mississauga, Canada. We annotated CNVs using GenomeArc Analytics, a platform to annotate and delineate CNVs, and then used rigorous downstream analysis to evaluate overlapping genes from NDD and CA CNVs. The gene enrichment analysis, human developmental transcriptomic landscape and descriptive statistics were carried out for the overlapped candidate gene set to infer biological pathways and brain developmental trajectory. **Results:** Out of 154 patients (NDD and CA cases), we extracted overlapping genes from 217 *de novo* CNVs (47.3% (80/169) of NDD and 64.6% (31/48) CA cases) and performed constraint gene analysis. 79 genes were found to have significantly enriched genomic overlap between NDD and CA patients within rare *de novo* pathogenic deletions (P-value = 0.01, OR = 1.58) and 45 genes identified within *de novo* pathogenic duplications (P-value = 0.01, OR = 1.97). The *de novo* CNVs encompass 138 overlapping genes impacted in both disorders, of which at least one exon is impacted by pathogenic deletions, and this shows significant (P-value = 2.87×10^{-90} , OR = 14.69) functional disruptions. Analysis of spatiotemporal transcriptome demonstrated both pathogenic deletion and duplication genes to be highly expressed during prenatal stage in human developmental brain (P-value = 4.95×10^{-6}). Analysis from the total cohort including pathogenic and variant of uncertain significance (VUS) identified 1086 overlapping constraint genes from CNVs and 90 genes (8%) had no Online Mendelian Inheritance in Man (OMIM) entries. From the list of overlapping genes, *EHMT1*, an interesting known NDD gene encompassed *de novo* pathogenic CNVs from both NDD and CA patients, whereas *FAM189A1*, and *FSTL5* are new candidate genes from non-OMIM entries which may have common etiology in NDD and CA. **Conclusion:** In this study of two large cohorts, we have identified overlapping CNVs and genes in NDD and CA patients. We propose a list of constraint genes that have the potential to play important role into the common etiology of NDD and CA.

Complex Traits Posters - Thursday

PB1323. *De novo* germline and somatic variants as modifiers of cleft lip/palate severity in mice

Authors:

S. Rao¹, E. Farrow², T. Cox¹; ¹Univ. of Missouri-Kansas City, Kansas City, MO, ²Children's Mercy Hosp, Kansas City, MO

Abstract Body:

Cleft lip/palate (CL/P) is one of the most common human birth defects affecting 1 in 700 live births around the world each year. Non-syndromic CL/P accounts for about two-thirds of all cases and is typically characterized by low penetrance and highly variable presentation. Population-level genetic and epidemiological data as well as teratogenic studies support a threshold model of susceptibility, with genetic and environmental (or epigenetic) contributions. Common variants at over 40 loci contribute ~30% of the genetic risk of non-syndromic CL/P in different ethnic populations. In contrast, rare, high-impact, inherited and *de novo* germline pathogenic variants underlie as many as 40% of familial and ~5% of isolated cases. Notably, marked inter-individual phenotypic variability is also seen in inbred mouse CL/P models that have long been presumed to be genetically identical. While there is growing evidence to support the role of epigenetics and maternal nutrition in cleft risk, we hypothesize that *de novo* germline variants (ie. individual-specific genetic variation) and/or postzygotic (ie. somatic) mutations arising during embryogenesis may also contribute to penetrance and variability in presentation of CL/P, that is, they may function as phenotypic modifiers of cleft risk and severity even in the presence of an inherited risk allele. To begin to investigate this, we have used the *Wnt9b* KO mouse CL/P model, which exhibits the full variability of cleft presentations as seen in humans, including incomplete penetrance. Using both whole genome sequencing and total RNA-sequencing approaches on the same embryonic facial tissue, we have identified a high incidence of both *de novo* germline and tissue-specific mutations, indicating that significant numbers of somatic variants are present in all embryonic tissues. Using standard filters for coding region pathogenicity and predicted regulatory region function, we find dozens of variants that could potentially impact the expression or function of genes or pathways crucial for facial development. While the true impact of such variants is still to be functionally tested, if proven, the data will have significant implications for all birth defects, including relative risk calculations, counselling of families, and personalized medicine interventions.

Complex Traits Posters - Wednesday

PB1324. *De novo* variant analysis of childhood-onset obsessive-compulsive disorder in the French-Canadian population.

Authors:

K. Bornais^{1,2}, J. P. Ross^{1,2}, Z. Schmilovich^{1,2}, M. Medeiros^{1,2}, D. Spiegelman^{1,2}, P. A. Dion^{1,2}, G. A. Rouleau^{1,2}; ¹McGill Univ., Montreal, QC, Canada, ²Montreal Neurological Inst. and Hosp., Montreal, QC, Canada

Abstract Body:

Obsessive-compulsive disorder (OCD) is a neuropsychiatric disorder with a global prevalence of 1-2%. Affected individuals experience cycles of intrusive thoughts or sensations (obsessions) and repetitive behaviours (compulsions). There is a genetic component to early-onset OCD, with estimates of heritability being 45-65%, but most of this heritability cannot currently be explained by known genetic factors. *De novo* variants (DNVs) arise post-zygotically in an individual but are not present in either biological parent and may contribute significantly to OCD risk. However, the number of identified DNVs in OCD remains limited and the level of genetic contribution to OCD suggests that more are likely to exist. Here, we aimed to identify DNVs in the probands of French-Canadian (FC) childhood-onset OCD trios. Given that leveraging the genetically homogenous FC population increases the power for genetic discovery, we hypothesized that DNVs observed in the FC childhood-onset OCD probands would identify OCD genetic risk factors.

Whole-exome sequencing (WES) was performed on 53 FC OCD trios ($n = 159$). Initial inclusion criteria selected only trios where all three trio participants self-reported as FC. DNVs were called by filtering for variants that are heterozygous in the probands and not carried by either parent. Candidate DNVs were selected based on their estimated impact on gene function, a minor allele frequency of less than 0.1% in the general population, and biological relevance to OCD. DNVs were collapsed into genes to identify candidate childhood-onset OCD risk genes. This candidate gene list was then used to identify enriched molecular and biological pathways underlying childhood-onset OCD.

1,441 candidate DNVs across 504 genes were observed in the probands of the trios. 44 of these genes were previously identified in large-scale DNV analyses of OCD (Halvorsen et al., 2021; Cappi et al., 2020). Additionally, our analysis identified 260 previously unobserved genes in OCD carrying DNVs that rank in the top 0.1-1% of deleterious variants in the human genome (average CADD score = 26.3). The top enriched Gene Ontology (GO) terms of interest include “regulation of neuron projection” ($p = 0.003$), “histone acetylation” ($p = 0.0009$), “calcium channel activity” ($p = 0.0006$), and “cadherin binding” ($p = 0.0004$). Findings from this study will shed light on the genetic and biological factors underlying childhood-onset OCD risk, further highlighting the ability to leverage the FC population to elucidate the genetics of complex neuropsychiatric disorders.

Complex Traits Posters - Thursday

PB1325. Deciphering the causal relationship between blood pressure and white matter integrity: a Mendelian Randomization study in the UK Biobank

Authors:

Z. Ye¹, C. Mo², S. Liu³, S. Gao¹, B. Zhao⁴, T. Canida⁴, Y-C. Wu⁴, K. Hatch¹, Y. Ma¹, B. Mitchell¹, E. HONG¹, P. Kochunov¹, C. Chen¹, B. Zhao⁵, S. Chen¹, T. Ma⁴; ¹Univ. of Maryland, Baltimore, Baltimore, MD, ²Harvard Med. Sch., Boston, MA, ³Qilu Univ. of Technology (Shandong Academy of Sci.), Jinan, China, ⁴Univ. of Maryland, College Park, MD, ⁵Purdue Univ., West Lafayette, IN

Abstract Body:

Background: Elevated arterial blood pressure (BP) is one of the most common risk factors for cerebrovascular and cardiovascular diseases. It has been associated with the atrophy of cerebral white matter (WM) in the brain, but no causal relationship has been established.

Methods: In this study, we performed a two-sample Mendelian Randomization (MR) analysis to evaluate the causal effects of BP on WM integrity within different brain regions, which were measured by fractional anisotropy (FA) of diffusion tensor imaging. We followed the MR assumptions to select genetic variants as valid instrumental variables and used two non-overlapping sets of epidemiological samples collected by the UK Biobank to estimate gene-exposure association (N = 203,111; mean age = 56.71 years) and gene-outcome association (N=16,156; mean age = 54.61 years). The causal effect of BP on regional FA was then estimated by combining the two association estimates using an inverse-variance weighted approach. We also performed two additional MR analyses to exclude the possibility of reverse causality.

Results: We found significantly negative causal effects (adjusted $p < 0.05$) of two BP measures (systolic blood pressure (SBP) and diastolic blood pressure (DBP)) on FA in eight brain regions: bilateral posterior thalamic radiation (PTR-R/PTR-L), fornix (cres)/stria terminalis left (FX/ST-L), superior fronto-occipital fasciculus right (SFO-R), bilateral sagittal stratum (SS-R/SS-L), Body of corpus callosum (BCC) and retrolenticular part of internal capsule (RLIC-L), which are related to memory and cognitive functions.

Conclusion: Our study extended the previous findings of associations between BP and WM integrity to the causation and provide more insights into the pathological processes of elevated BP that might chronically alter the brain microstructure within specific brain regions.

Complex Traits Posters - Wednesday

PB1326. Deep embedded clustering by phenotype and genome-wide association study in autism from SPARK individuals

Authors:

F. Ueno¹, T. Onuma¹, I. Takahashi¹, H. Ohseto¹, A. Narita¹, T. Obara¹, M. Ishikuro¹, K. Murakami¹, A. Noda¹, F. Matsuzaki¹, H. Metoki^{1,2}, G. Tamiya^{1,3}, S. Kure^{1,4}, S. Kuriyama¹; ¹Tohoku Univ., Sendai, Japan, ²Tohoku Med. and Pharmaceutical Univ., Sendai, Japan, ³RIKEN Ctr. for Advanced Intelligence Project, Tokyo, Japan, ⁴Miyagi Children's Hosp., Sendai, Japan

Abstract Body:

Autism spectrum disorder (ASD) has heterogeneous phenotypic and genetic characteristics, with associated variants in ~1,000 genes. Stratifying ASD patients may help identify more genetically homogeneous subgroups. Using a deep embedded clustering algorithm, we conducted cluster analyses of Simons Foundation Powering Autism Research for Knowledge (SPARK) datasets and performed genome-wide association studies (GWAS) of the clusters. We observed no significant associations in the conventional GWAS comparing all patients to all controls, whereas in cluster-based GWAS (cGWAS), we identified 90 chromosomal loci that satisfied the $P < 5.0 \times 10^{-8}$. Many of these loci were located within or near previously reported candidate genes for ASD. Many other revealed genes were previously suggested to be associated with other brain disorders. These findings suggest that clustering may successfully identify subgroups with relatively homogeneous disease etiologies.

Complex Traits Posters - Thursday

PB1327. Deep learning models of gene regulation in Alzheimer's disease improve functional fine-mapping and thereby cross-ancestry PRS portability

Authors:

T. Lin¹, C. M. Lakhani¹, T. Raj², D. Knowles¹; ¹New York Genome Ctr., New York, NY, ²Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Alzheimer's Disease (AD) is the most prevalent type of dementia and has been shown to have a substantial genetic component. Previous genome-wide association studies (GWAS) found many, predominantly non-coding, loci that are associated with AD and which can be used to predict AD via polygenic risk score (PRS). However, due to European-centric studies and the different linkage disequilibrium (LD) patterns and allele frequencies across various ethnic backgrounds, these results are not necessarily applicable to non-European populations. We hypothesized that using functional fine-mapping to detect the causal variants underlying GWAS loci could improve PRS cross-ancestry portability for AD. Hence we used deep learning models of gene regulation in AD-relevant cell types to empower functional fine-mapping and improve cross-ancestry AD prediction. We used three different published GWAS for downstream analysis. We combined 48 functional annotations selected by stratified LD score regression from Roadmap and DeepSEA with brain ATAC-seq from cell types that are reported to be associated with AD progression. We also included 187 "baseline" LD-related annotations to perform our study. Tests were built using functional fine-mapping (PolyFun) and PRS, compared with (1) statistical Fine-mapping (SuSiE), (2) baseline using clumping and P-value thresholding (PLINK), and (3) other published methods (SBayesR and PRSice). Using whole-genome sequencing data from the AD Sequencing Project, 5,122 cases and 6,599 controls were included after filtering out samples under the age of 65. European (EUR), African American (AFR), and Hispanic (AMR) groups account for 57.2%, 22.6%, and 18.9% of the total samples, respectively. Since our target data has a relatively diverse ancestral composition, we tested the performance of each ethnic group independently. We achieved our best performance using GWAS from Kunkle (2019); after removing the two most significant APOE- ϵ 4 loci rs429358 and rs7412, the AUC and R-squared reached 0.58 and 3.9% in the EUR cohort, which is consistent with previous studies. We found functionally-informed fine-mapping with deep learning annotations did increase trans-ethnic portability of PRS in AFR and AMR cohorts. Using a PRS trained on the Bellenguez (2021) GWAS, we found AUC increased from 0.528 to 0.55 in the AMR population and from 0.515 to 0.516 in the AFR population. While the model trained on summary statistics from Kunkle (2019) increased PRS AUC from 0.512 to 0.517 in the AFR cohort but decreased from 0.53 to 0.524 in the AMR cohort. To sum up, our work provides new insight into the applicability of functional annotations in cross-ethnic AD prediction.

Complex Traits Posters - Wednesday

PB1328. Deep venous thrombosis: genotyping and whole-exome sequencing data to improve genetic risk prediction

Authors:

V. Lo Faro, T. Johansson, J. Höglund, F. Hadizadeh, Å. Johansson; Uppsala Univ., Uppsala, Sweden

Abstract Body:

Background: Deep venous thrombosis (DVT) is the formation of blood clots in the deep veins. Blood clots traveling to the lung can cause organ damage or sudden death. More than 60% of DVT risk is influenced by genetic factors, such as the Leiden mutation in *F5* (FVL). Characterizing the genetic contribution and stratifying individuals based on their genetic makeup can favorably impact risk prediction. **Methods:** We performed a genome-wide association study and constructed a polygenic risk score (PRS) in the 60% (N=284,591) of the UK Biobank cohort. The remaining 40% (N=198,362) was employed to evaluate the PRS and to perform a gene-based test on exome-sequencing data to investigate the effects of rare variants. **Results:** We identified and replicated a new variant (rs11604583) near the *TRIM51* gene, and a rare variant (rs187725533), associated with a 2.2-fold higher risk of DVT, in the *CREB3L1* gene. The top PRS decile was associated with a 3.4-fold risk of DVT, an effect that was still 2.3-fold when excluding FVL carriers. The cumulative risk of DVT at the age of 70 years for FVL carriers in the top PRS decile is 10%, contraposed to 5% for non-carriers. **Conclusion:** We showed that common and rare variants influence DVT risk and that the PRS improves risk prediction on top of FVL. This suggests that individuals classified with high PRS scores could benefit from early genetic screening.

Complex Traits Posters - Thursday

PB1329. Differential gene expression analysis of alcohol use disorder across postmortem human brain tissues.

Authors:

C. Willis¹, J. White¹, B. Quach¹, S. Han², R. Tao², J. Shin², A. Deep-Soboslay², R. Mayfield³, B. Webb¹, E. Johnson¹, J. Kleinman², L. Bierut⁴, D. Hancock¹; ¹RTI Intl., Research Triangle Park, NC, ²Lieber Inst. for Brain Dev., Baltimore, MD, ³University of Texas at Austin, Austin, TX, ⁴Washington Univ Sch Med, St Louis, MO

Abstract Body:

Alcohol use disorder (AUD) is a common psychiatric disorder that increases risk of many diseases, including liver disease and many cancers. AUD is 50–60% heritable, but specific genes contributing to AUD risk and gene regulation in response to alcohol exposure are not fully understood. Here, we conducted a differential gene expression analysis in two different human brain tissues, nucleus accumbens (NAc) and dorsolateral prefrontal cortex (DLPFC), to better understand AUD. Bulk RNA sequencing was performed on both NAc and DLPFC tissues from 122 deceased individuals with next-of-kin reporting to define cases (2+ AUD symptoms) and controls (no symptoms). Following quality control, 100 NAc samples (53 AUD cases, 47 controls) and 97 DLPFC samples (51 AUD cases, 46 controls), all of European ancestry, remained for analysis. 20,324 and 20,666 genes for NAc and DLPFC, respectively, were tested for differential expression by AUD status using negative binomial regression models that accounted for major depressive disorder status, smoking status, and surrogate variables, which serve as proxies for known covariates as well as unmeasured technical and biological confounders. We used false discovery rate (FDR) <0.05 to declare genome-wide statistical significance and removed potential false positives based on Cook's distance. We identified 105 and 637 genes as differentially expressed for NAc and DLPFC, respectively. Of these, 28 genes were differentially expressed in both NAc and DLPFC. We used our results to assess evidence for replication of genes previously reported in a similarly powered differential gene expression study of AUD in pre-frontal cortex (65 cases and 73 controls of European ancestry, Kapoor et al. *Transl Psychiatry* 2019). Of the 129 genes reported for differential expression, we replicated 24 genes in our DLPFC results at $P < 3.9 \times 10^{-4}$, based on Bonferroni correction for 129 genes tested. Six of the 24 were also implicated in NAc: *HMG2*, *MGEA5*, *ARRDC3*, *HMGB2*, *MDK*, and *ODCI*. In addition, 24 and 5 of the genes with differential expression in DLPFC and NAc, respectively, were identified as significant for gene-based association with alcohol consumption or use disorder, as derived from genome-wide association studies (GWAS) (Sey et al. *Mol. Psychiatry* 2022), suggesting that a subset of our findings represent genetically-driven differential expression that influences risk for AUD while other identified genes may result from alcohol exposure. The replication of differentially expressed genes in DLPFC, converging evidence with GWAS, and other differentially expressed genes identified in NAc helps expand understanding of the neurobiology of AUD.

Complex Traits Posters - Wednesday

PB1330. Differential splicing occurs in individuals with disseminated coccidioidomycosis infections.

Authors:

S. Jensen¹, S. J. Spendlove¹, A. V. Stephens¹, R. H. Johnson^{1,2}, A. Heidari^{1,2}, R. Kuran^{1,2}, H. Pimentel¹, M. J. Butte¹, V. A. Arboleda¹; ¹UCLA, Los Angeles, CA, ²Valley Fever Inst., Bakersfield, CA

Abstract Body:

Coccidioidomycosis is a disease caused by inhalation of the fungal spores from the fungi *Coccidioides immitis* and *posadasii*. The fungi infect approximately 150,000 people in the United States yearly, the vast majority of whom remain asymptomatic or develop mild respiratory symptoms and are considered to have uncomplicated Valley Fever (UVF). In an estimated 1-2% of infections, the fungi spread to other parts of the body, a potentially fatal form of the disease known as disseminated coccidioidomycosis (DCM). Recent work in patients with DCM has identified ultra-rare risk variants associated with the immune system and prompted the use of immunomodulatory drugs to successfully treat DCM. We hypothesized that analysis of whole blood transcriptomes could improve our identification of genes and pathways associated with susceptibility to DCM. In order to identify other patients with similar genetic susceptibilities, our study has recruited individuals with coccidioidomycosis from UC Davis and the Valley Fever Institute. Whole genome and RNA sequencing were performed for a cohort of 39 individuals who had (n=23) or did not have (n=16) a history of dissemination. Differential gene expression analysis was carried out with DESeq2 and differential splicing analysis with LeafCutter. Gene ontology enrichment and pathway analyses were carried out for significant results. We detected no significant differences in global gene expression profiles between individuals with history of DCM and those with more mild forms of coccidioidomycosis. Assessment of the proportion of splice isoforms between the severity groups revealed significant (FDR<0.05) differential splicing for six genes: *ARHGAP12*, *ATP6V1H*, *ERCC6*, *GSDMB*, *MRPL52*, and *PCGF5*. These transcripts are highly expressed in immune cells. Transcriptional analyses are an important first step to understanding the role splicing plays in determining risk factors for coccidioidomycosis dissemination. Future work will explore splicing QTLs and other genetic regulators of transcript abundance.

Complex Traits Posters - Thursday

PB1331. Differentiating Mesenchymal Stem Cells from iPSCs for Analysis of Type 2 Diabetes and Related Trait GWAS Loci

Authors:

C. Ventresca¹, A. Varshney¹, P. Orchard¹, A. Monteiro da Rocha¹, M. Laakso^{2,3}, J. Tuomilehto⁴, T. Lakka², K. Mohlke⁵, M. Boehnke¹, L. Scott¹, H. Koistinen⁴, F. Collins⁶, T. Herron¹, S. Parker¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Univ. of Eastern Finland, Kuopio, Finland, ³Kuopio Univ. Hosp., Kuopio, Finland, ⁴Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, ⁵Univ. of North Carolina, Chapel Hill, NC, ⁶NIH, Bethesda, MD

Abstract Body:

Background Some type 2 diabetes (T2D) and related trait genome wide association study (GWAS) signals likely impact T2D risk via skeletal muscle, a primary insulin responsive tissue. We previously generated in vivo mesenchymal stem cell (MSC) data through single nucleus (sn-)multi-omics (RNA+ATAC) profiling of 286 skeletal muscle biopsies and found that a subset of GWAS signals colocalize with cell-type specific e/caQTL active in skeletal muscle MSCs. One method to investigate these signals is to differentiate MSCs from induced pluripotent stem cell (iPSC) lines, allowing us greater flexibility to explore these pathways. From the same individuals with muscle sn-multi-omics, we derived 50 iPSC lines from fibroblasts. To develop a greater understanding of T2D and related trait mechanisms, we can use techniques on the iPSCs differentiated to MSCs to nominate target genes and regulatory elements at each GWAS signal. These approaches include sn-multi-omics and CUT&Tag, a sensitive method to analyze genome-wide histone marks. We have already optimized both techniques on multiple cell types. **Hypothesis** We hypothesize that MSCs can be differentiated from iPSC lines for the purpose of investigating the impact of MSC-specific molecular mechanisms on T2D risk and T2D-related traits. **Results** Here, we demonstrate that MSCs can be derived from iPSC lines for deep molecular phenotyping. These iPSC-derived MSCs display MSC morphology, expression of MSC marker genes, and loss of expression of pluripotency and hematopoietic markers. Over the course of the differentiation, *Oct3/4*, a pluripotency marker, drops from 73.4% of cells expressing this down to 0.42%. Meanwhile, *CD73*, a MSC marker, increases from 7.25% of cells expressing up to 60.2%, based on flow cytometry analysis. We are currently performing a time course analysis with this experimental system with ten independent iPSC lines all multiplexed to explore the trajectory from iPSC progenitors to fully-differentiated MSCs, and compare in vitro sn-multi-ome signatures to those observed in matched in vivo MSCs. **Future Directions** We will use a cohort of 50 iPSC-derived MSC samples to investigate stimulatory state specific genetic regulatory effects (e/caQTL) within T2D pathways and will validate our results by performing CRISPRi knockdown and CRISPRa activation on a subset of nominated loci.

Complex Traits Posters - Wednesday

PB1332. Direct participation of individuals scales the development of cardiometabolic polygenic risk score models.

Authors:

A. Lopez Pineda^{1,2}, M. Vernekar^{3,4}, S. Moreno Grau¹, B. Moatamed¹, M. Lee¹, M. Nava-Aguilar^{1,2}, G. Gonzalez-Arroyo^{1,2}, K. Numakura^{3,4}, Y. Matsuda^{3,4}, A. Ioannidis^{5,1}, N. Katsanis¹, T. Takano^{3,4}, C. Bustamante^{5,1}; ¹Galatea Bio, Inc., Hialeah, FL, ²Amphora Hlth., Morelia, Mexico, ³Genomelink, Inc., Berkeley, CA, ⁴Awakens Japan K.K., Tokyo, Japan, ⁵Stanford Univ., Palo Alto, CA

Abstract Body:

More than 30 million people have been genotyped by direct-to-consumer (DTC) companies, providing a potential mechanism for enabling biomedical research at a global scale. However, phenotypic information is sequestered in the initial provider networks. The scientific community has limited ability to integrate clinico-genomic datasets to fulfill the promise of precision health, catalyzing improvements in patient outcomes.

Here, we present a novel geno-pheno platform that integrates data from more than 300,000 participants who uploaded their genotype data files, and were invited to answer general health questionnaires regarding two common chronic disease conditions: Type 2 diabetes (T2D) and hypertension. Quality control, imputation, and genome-wide association studies were performed, also polygenic risk scores (PRS) were calculated using the BASIL algorithm.

Over a period of 6 months, we collected data on 4,550 participants who reported being affected for T2D and 4,528 participants for hypertension. We identified 164 (60%) variants in T2D, and 230 (63%) in hypertension showing identical effect direction to previously reported genome-significant findings in Europeans. Moreover, the performance of our PRS models (AUC of 0.68 in T2D, and 0.69 in hypertension) is comparable to previously published models trained with larger academic datasets.

The direct participation of individuals has shown the potential to generate rich datasets enabling the creation of PRS cardiometabolic models. The generation of clinico-genomic datasets, which leverages the direct participation of individuals, accelerates data acquisition for biomedical research. However, it is important to use proven quality control (QC) mechanisms to successfully enable traditional GWAS and PRS analyses.

Our study shows that predictive tools for both T2D and hypertension can be successfully trained with heterogeneous genotypic and phenotypic data generated outside of clinical environments. Our methods can recapture prior findings with fidelity, which are also highly comparable to those results obtained from controlled (and academic) datasets. DTC data provides a mechanism for scaling precision health care delivery beyond the small number of countries who can afford to finance these efforts directly.

Complex Traits Posters - Thursday

PB1333. Discovering disease context eQTLs in a patient cohort with active psoriatic arthritis

Authors:

M. Guan, A. Hart, W. Chen, K. Sweet, D. Waterworth; Janssen R&D, Spring House, PA

Abstract Body:

Psoriatic arthritis (PsA) is a heterogeneous, heritable, chronic inflammatory disease that occurs in ~20-30% of psoriasis patients. It variably affects synovial joints, tendons, entheses and axial sites in addition to skin, resulting in signs and symptoms including peripheral joint inflammation, enthesitis, and dactylitis. Genetic factors play an essential role in the dysregulation of key pathways in PsA, but information on genetic variants and their functional relevance associated with PsA susceptibility and pathology is limited. Expression Quantitative Trait Loci (eQTL) mapping is a tool to identify allelic variants associated with gene expression. Disease context eQTL mapping provides a deeper understanding of candidate susceptibility genes at risk loci and PsA etiology. We performed eQTL mapping using RNA sequence data from whole blood (WB) and imputed genetic data from 644 patients with active PsA to identify disease context eQTLs. A linear model, adjusting for age, sex, top 10 genetic principal components, and 3 surrogate variables capturing unknown sources of variability from expression data, was used for cis-eQTL analysis. Among 10,984 genes identified as eGenes (genes with significant cis-eQTLs), 8,046 (~73%) were overlapping with eGenes from a non-disease Genotype-Tissue Expression (GTEx) project v8 WB dataset (n=670 samples); 2,938 (~27%) eGenes were significantly associated with up-/downregulated genes in the PsA cohort (FDR $p < 0.05$). Pathway analysis of eGenes significant in PsA indicated their involvement in metabolism, signal transduction, and cytokine signaling in the immune system. Notable SNP-eGene pairs included rs10889664-IL23R, rs2837-ILDR2, rs4577297-IL1RL1, rs5744249-IL18, and rs78031781-S100A3. The identification of the eQTL rs10889664 further supports IL23R as a robust genetic susceptibility locus for PsA. In addition, 25 of 824 SNP-eGene pairs that were significant in both GTEx and PsA cohorts showed effect size differences ($< 25\%$ interquartile - 1.5 IQR or $> 75\%$ interquartile + 1.5 IQR). These include rs114351845-IL17RE, rs146047341-TGFB3 and rs77366070-SMURF2, which are involved in inflammatory conditions. IL-17RE encodes the IL-17C receptor implicated in skin inflammation and relevant to PsA. In summary, we identified eQTLs associated with PsA, enabling a better understanding of the association between genetic risk loci and PsA disease susceptibility. These data may facilitate the discovery of genetic factors to enable disease stratification and fine-mapping of PsA genetic associations for target identification.

Complex Traits Posters - Wednesday

PB1334*. Discovering the early molecular markers of lipid dysregulation using deep phenotyping and adipose tissue gene expression.

Authors:

M. Tomlinson, J. El-Sayed Moustafa, A. Hodgkinson, K. Small; King's Coll. London, London, United Kingdom

Abstract Body:

Cardiovascular disease is characterised by the accumulation of fatty deposits in the arteries and pro-atherogenic effects of adipose tissue signalling. Accordingly, dysregulation of lipid metabolism is a major contributor to the development of cardiovascular disease. To reduce the burden of cardiovascular disease, we need to improve prognostic approaches that identify individuals at risk of lipid dysregulation so they can modify their environmental exposures or start taking medication before it accelerates the onset of cardiovascular disease. Here we explore associations between the future use of lipid-regulating drugs and 287 intermediate quantitative phenotypes relevant for cardiovascular disease risk including biochemical measures, clinical diagnoses, and DXA scans of body composition at baseline (2001-2010) and follow-up (2014-2019) in a cohort of British twins. We find 27 phenotypes at baseline (2001-2010) and trajectories of 18 phenotypes are linked with future lipid-regulating drug use including measures of lipid metabolism and cardiovascular health (FDR 5%). Utilising transcriptional profiles from a subset of twins ($n = 766$), we find adipose tissue gene expression at baseline (2007-2009) is associated cross-sectionally with 16 of the phenotypes linked with future drug usage (FDR 5%). The genes with expression levels that are associated with 10 of these phenotypes also produce greater absolute log-fold changes in participants that go on to require lipid-regulating drugs ($p < 0.05/16$), and overall, these genes are enriched for processes related to atherosclerosis, inflammation, and lipid metabolism. Finally, we show 13 unique genes are differentially expressed at baseline in participants that go on to require lipid-regulating drugs by follow-up (2014-2019) compared to matched controls (FDR 5%), including top hit *ESM1* ($p < 1.82 \times 10^{-16}$), which has been implicated in angiogenesis and inflammation. Taken together, our findings highlight early phenotypic and molecular markers of lipid dysregulation that could help identify individuals at higher risk of lipid dysregulation and future cases of cardiovascular disease.

Complex Traits Posters - Wednesday

PB1335. Discovery of rare variants in 219 genes associated with adult human height via burden analysis in exomes of 785,210 individuals

Authors:

A. Locke¹, J. Kosmicki¹, D. Sharma¹, X. Bai¹, A. Marcketta¹, K. Watanabe¹, H. Kang¹, S. Balasubramanian¹, J. Torres², P. Kuri Morales³, R. Tapia-Conyer³, J. Alegre³, J. Berumen³, J. Emberson², R. COLLINS², J. Backman¹, J. Reid¹, Regeneron Genetics Center, RGC Research Partners, A. Baras¹, J. Marchini¹, G. Abecasis¹, M. Ferreira¹; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Univ. of Oxford, Oxford, Oxon, United Kingdom, ³Univ. Natl. Autónoma de México, Mexico City, Mexico

Abstract Body:

Human height is a highly heritable polygenic quantitative trait frequently used as a model to study the genetic contribution to individual variation. Through GWAS meta-analysis, the global GIANT Consortium has arguably explained all the additive variance in adult human height attributable to common genetic variation, implicating ~20% of the genome (Yengo et al. 2021). Yet the independent role of rare genetic variation (MAF<1%) in height remains largely unexplored. To this end, we performed a multi-cohort, multi-ancestry, joint association analysis of rare coding variation on adult human height in 785,210 individuals, including 209,676 individuals of non-European ancestry, from a joint call set of >1.1 million exomes. We tested rare predicted loss-of-function (pLoF) and deleterious missense variants, for association with height using burden and SKAT tests. After correcting for multiple testing ($p < 5 \times 10^{-8}$), we identified 219 unique genes associated with height, most of which remained significant after conditioning on common variant association signals. The effects of these genes ranged from -2.3 standard deviations (SD; *ACAN*) to 1.5 SD (*FBNI*), or ~24.7 cm shorter to 15.7 cm taller. The mean effect of these associated genes was -0.1 SD (-1.07 cm). As anticipated, effect sizes increased with greater predicted protein impact and/or decreasing allele frequency. For example, burden tests including only pLoF variants had mean effect 3x larger (-0.34 SD) than the average gene test, and the average effect size tripled again, to -0.99 SD (10.7 cm), when testing only pLoF variants that are singletons in our data. In an independent dataset of ~29,000 adults of European ancestry, putatively functional rare variants in these 219 genes explained ~1.4% of the variation in adult height. Thirty-six of 219 associations were with genes previously identified with extreme height phenotypes or skeletal growth disorders in the Online Mendelian Inheritance in Man database, and 77% (167 of 219) represent new gene-based associations compared to GIANT Consortium results from the exome chip (Marouli et al. 2017). While 169 of 219 height-associated genes found here are also near common variant GWAS signals, our analyses provide direct independent evidence for specific genes influencing mechanisms related to growth. Our results from coding variation in nearly 800,000 individuals provide strong evidence that rare variant associations are strongly complementary to, but dramatically more specific than, common variant GWAS association, and while individually rare, can, due to large effect sizes, account for a considerable proportion of population variation.

Complex Traits Posters - Thursday

PB1336. Discovery of Type 2 Diabetes genes using an accessible tissue.

Authors:

D. Davtian¹, J. Schwenk², M. McCarthy³, A. Mahajan⁴, M-G. Hong⁵, E. Dermitzakis⁶, H. Im⁷, E. Pearson¹, A. Viñuela⁸, A. Brown¹; ¹Univ. of Dundee, Dundee, United Kingdom, ²SciLifeLab, Stockholm, Sweden, ³University of Oxford, Oxford, United Kingdom, ⁴Genentech, South San Francisco, CA, ⁵Karolinska Inst, Stockholm, Sweden, ⁶GSK, Geneva, Switzerland, ⁷Univ. of Chicago, Chicago, IL, ⁸Univ. of Newcastle, Newcastle, United Kingdom

Abstract Body:

Transcriptome Wide Association Study (TWAS) methods use reference expression datasets to link genes to disease. Often these use samples from a disease relevant tissue, but these may be difficult to collect, limiting sample size. We evaluate the power to discover Type 2 Diabetes (T2D) causal genes using a well powered whole blood expression reference against smaller reference datasets from disease relevant tissues.

Predictive models were built using whole blood expression and proteomic data (DIRECT consortium, n = 3029), pancreatic islets eQTL data (INSPIRE consortium, n = 420) and eQTL data from 49 tissues (GTEx consortium). These models were combined with GWAS summary statistics (DIAGRAM consortium, n = 898,130) to calculate gene-T2D association scores using S-PrediXcan.

We indeed find that larger reference panels discover more genes (97 significant associations when using DIRECT, 43 with the pancreatic islets model and from 2 to 84 for GTEx results, Spearman correlation = 0.9 between sample size and number of significant associations) and relevance of the tissue did not appear as a significant factor. However, we do find that genes implicated by relevant tissues are more likely to be near genome wide significant GWAS associations such as *TCF7L2* and *IGF2BP2*, suggesting that these variants are tissue specific eQTLs. In contrast, genes implicated using a well powered reference dataset are more likely to have multiple lines of genetic evidence, with multiple eQTLs which also show moderate disease association. Finally, we find that controlling for BMI when looking at T2D associations reduces our power by around one-third, suggesting many of the implicated genes are involved in insulin resistance with BMI as a mediating factor.

Our findings indicate that large studies in non-relevant tissues identify disease causal genes but can miss relevant tissue specific signals. In addition, correcting for environmental factors such as BMI could help to investigate the underlying biology of these associations.

Complex Traits Posters - Wednesday

PB1337. Discrepancy between genetically-predicted and actual body mass index is a significant predictor of incident cardiovascular disease

Authors:

T-M. Rhee¹, J-B. Park¹, S-H. Kwak¹, H. Lee¹, J. Choi¹, N. Kim², S. Lee², S-P. Lee¹; ¹Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, ²Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract Body:

Background: Body mass index (BMI) is a key predictor of cardiovascular diseases (CVD) and related mortality. Polygenic risk scores (PRS) derived from genome-wide association studies (GWAS) can accurately predict BMI and genetically-predicted BMI (BMI-pred) could serve as a threshold for personal fat storage. We hypothesized that difference between BMI-pred and actual BMI (BMI-diff) could reflect severity of environmental factors affecting BMI and may predict incident CVD morbidity and mortality in general population. **Methods:** From the UK Biobank prospective cohort, we selected participants of white British descent without a history of the major CVD; myocardial infarction (MI), stroke, or heart failure (HF). The global extended PRS for BMI was calculated using polygenic prediction via Bayesian regression and continuous shrinkage priors (PRS-CS) with summary statistics from the Genetic Investigation of ANthropometric Traits consortium GWAS. According to the calculated BMI-diff, the 10-year risk of all-cause death and major adverse cardiovascular events (MACE, composite of cardiovascular death, MI, stroke, and HF) was compared using Cox proportional-hazards model. Optimal range of the BMI-diff was determined by restricted cubic spline curves for each outcome. **Results:** A total of 385,270 participants (mean age 56.7±8.0, male 45.0%) were randomly divided into train set (n=269,394) and test set (n=115,876) to establish the most accurate BMI-pred using PRS-CS. The BMI-pred could explain 7.3% of the variance of BMI in the test set. All-cause death and incident MACE occurred in 7,916 (6.8%) and 8,535 (7.4%) cases during 10-year follow-up, respectively. In the adjusted hazard ratio curves, the lowest risk of both all-cause death and MACE was observed in participants who were mildly underweight than genetically-predicted, providing the optimal BMI-diff range between 0 to 1.9 kg/m² for all-cause death, and 0 to 6.6 kg/m² for MACE, respectively. In the obese group with genotype-phenotype mismatch (BMI-diff <-5 kg/m²), the risk of both all-cause death and MACE was significantly higher than obese group with matched genotype-phenotype. However, the underweight group with mismatch (BMI-diff >5 kg/m²) was only associated with increased risk of all-cause death, not MACE. **Conclusion:** Deviation of actual BMI from BMI-pred which reflects severity of environmental factors is a significant predictor of incident CVD morbidity and mortality. As a step towards precision medicine, we suggest the optimal target range of body weight reduction for individuals determined by their own BMI-pred to prevent future risk of CVD.

Complex Traits Posters - Thursday

PB1338. Dissecting autism heterogeneity by genotype phenotype analysis among autism risk gene carriers in SPARK

Authors:

C. Shu^{1,2}, P. Feliciano³, L. Snyder³, N. Volfovsky³, Y. Shen², W. K. Chung¹, SPARK Consortium; ¹Dept. of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY, ²Dept. of Systems Biology, Columbia Univ. Irving Med. Ctr., New York, NY, ³Simons Fndn., New York, NY

Abstract Body:

There is significant heterogeneity on the severity of autism spectrum disorder (ASD) core symptoms and comorbid symptoms that complicate the behavioral manifestations of ASD. We comprehensively assessed the gene-specific contributions to ASD core symptoms and comorbidities based on ASD cases carrying *de novo* or inherited coding variants among known and newly identified ASD risk genes using the Simons Foundation Powering Autism Research for Knowledge (SPARK) cohort. To dissect the ASD phenotypic spectrum, we performed principal component analysis (PCA) among ~15,000 ASD cases on ~200 ASD phenotypic measures on medical and psychiatric history, developmental milestones and standardized surveys measuring levels of social communications, repetitive behaviors and motor skills. We represented ASD phenotypic spectrum by the top two principal components (PCs) and visualized the ASD risk gene carriers in phenotypic spectrum. We found that carriers with *de novo* heterozygous mutations in genes like *ADNP*, *SHANK3*, *CHD8*, *FOXP1* and *SCN2A* have greater severity of social communication impairment, delayed developmental milestones, more need for specialized ASD services. Gene carriers with mutations on *ARID1B*, *GIGYF1* and *NAV3* have less developmental issues but more severe repetitive behaviors. These results support that gene-level genotype-phenotype analyses are necessary rather than overall *de novo* burden analysis to dissect the ASD phenotypic heterogeneity.

Complex Traits Posters - Wednesday

PB1339. Dissecting single cell splicing isoforms in human hearts using single cell nanopore sequencing

Authors:

L. Lu¹, C-K. Shiau¹, R. Kieser², K. Youker³, R. Gao¹; ¹Northwestern Univ., Chicago, IL, ²Houston Methodist Res. Inst., Houston, TX, ³Houston Methodist Res. Inst., Houston, IL

Abstract Body:

Heart disease remains one of the leading causes of death in the world. Dysregulation of alternatively spliced isoforms plays critical roles in molecular etiologies of heart diseases. High throughput single cell nanopore sequencing (HT-scNanoSeq) of full-length mRNAs provides an ideal approach to investigate single cell splicing isoform profiles. However, detection of cell population specific isoforms has been challenged due to the high drop off rates in individual cells. To address this challenge, we developed a new computational toolkit, scNanoPSI (single cell Nanopore sequencing for Population Specific Isoform) to detect both differentially spliced and differentially expression isoforms in distinct cell types and cell states. In this study, we dissected the atlas of single cell splicing isoforms of major cell types in adult human hearts including cardiomyocytes, pericytes, fibroblasts, endothelial cells, lymphocytes, macrophages as well as neuronal cells. Moreover, we detected a list of aberrant isoforms in sub-cell types in failed human hearts. In summary, scNanoPSI enables robust detection of splicing isoforms traceable to cell types and cell states, and our findings in human hearts provide important evidence for developing new therapeutics by targeting abnormal isoforms to treat heart diseases.

Complex Traits Posters - Thursday

PB1340. Dissecting the relationship between recurrent pregnancy loss and type 1 and type 2 diabetes in Western Ukrainian population

Authors:

Y. Sharhorodska^{1,2,3}, V. Pascat⁴, Z. Balkhiarova^{2,3,5}, A. Ulrich^{3,5}, L. Chorna¹, I. Shymanska¹, D. Zastavna¹, O-R. Gnateyko¹, M. Kaakinen^{3,2,5}, H. Makukh¹, I. Prokopenko^{3,2,5}; ¹Inst. of Hereditary Pathology of Natl. Academy of Med. Sci. of Ukraine, Lviv, Ukraine, ²People-Ctr.d AI Inst., Univ. of Surrey, Guildford, United Kingdom, ³Dept. of Clinical and Experimental Med., Univ. of Surrey, Guildford, United Kingdom, ⁴Université de Lille, Lille, France, ⁵Dept. of Metabolism, Imperial Coll. London, London, United Kingdom

Abstract Body:

Aims/hypothesis: The loss of two or more sequential pregnancies before 24 weeks of gestation is defined as recurrent pregnancy loss (RPL). It is a frequently occurring human infertility-related condition affecting ~0.8% to 1.4% of women in the general population. The causal factor for RPL leading to treatment is usually identified in ~50% of patients only. Given the lack of GWAS studies and the possibility of various, currently unknown RPL risk factors, we tested a hypothesis about metabolic pathways contributing to RPL susceptibility. We investigated whether RPL shares biological pathways with type 2 and type 1 diabetes (T2D & T1D).

Methods: We investigated the predictive ability of T2D/T1D polygenic risk scores (PRS) for RPL risk using the LUCAR study (Lviv Ukrainian Cohort for Advancing Reproductive Health) from the Western Ukraine. LUCAR includes 273 idiopathic RPL cases and 535 controls with at least one healthy child, all having their genome-wide data imputed to TopMED panel. For PRS, we used 552 SNPs previously validated for T2D (PMID: 32541925) and 61 SNPs established for T1D (PMID: 32770166) in genome-wide association studies and assessed their associations with RPL via logistic regression.

Results: We showed that neither of the PRSs for T1D or T2D was associated with the risk of RPL (T1D: odds ratio 1.02, 95%CI[0.88-1.18]), (T2D: odds ratio 0.97, 95%CI[0.84-1.12]), although we did identify six individuals with a high polygenic susceptibility to T1D, of which one individual was also at high risk of developing T2D. The receiver operator characteristic (ROC) analysis demonstrated no predictive ability of T1D PRS with the area under the curve (AUC) of 0.49, 95%CI(0.45-0.53), and for T2D AUC=0.50, 95%CI(0.45-0.54).

Conclusions: Despite previous epidemiological studies suggesting associations between RPL and diabetes types, we did not detect such a relationship in the LUCAR study from Western Ukraine.

Funding: Institute of Advanced studies Fellowship, University of Surrey, UK; Faculty Research Support Fund, Faculty of Health and Medical Sciences, University of Surrey, UK; People-Centred AI Institute, University of Surrey, UK; Fellowship from the French Embassy, Ukraine; LONGITOOLS, H2020-SC1-2019-874739; PreciDIAB, ANR-18-IBHU-0001.

Complex Traits Posters - Wednesday

PB1341. Dissecting the role of pleiotropy in the genetic adaptation of Tibetans to high altitudes.

Authors:

A. Thornburg, I. Aneas, S-Y. Park, D. Sobreira, O. Gray, A. Di Rienzo, M. Nobrega; Univ. of Chicago, Chicago, IL

Abstract Body:

Several loci in the human genome have undergone positive selection, allowing local adaptation of a population to a specific selective pressure. Typically, positive selection in human populations is not necessarily driven by selective sweeps where variants at one locus are selected for and reach a high frequency in a given population. However, the Tibetan population has undergone strong positive selection in response to hypoxic stress at high altitudes in the *EPAS1* gene locus, which encodes a transcription factor subunit of the HIF pathway. Understanding why this locus is under strong positive selection will help understand the forces shaping genetic adaptation in humans. We identified an enhancer of *EPAS1* within the positively selected genomic region and harboring variants strongly associated with high altitude in Tibetans. We showed that this enhancer, ENH5, is active in several human cell types important for hypoxic response, such as heart, kidney, and endothelial cells, as well as having higher activity in response to hypoxia but with the Tibetan allele having a blunted response compared to the Han Chinese (lowlanders) allele. Generating a mouse ENH5 knockout line to model the effects of ENH5 *in vivo* also showed differences in gene expression in several tissue types. This suggests that adaptive pleiotropy could be the driver behind the strong selection signals in the *EPAS1* gene locus, meaning that differences in gene expression between tissues due to ENH5 can be conferring different, yet adaptive phenotypes. Thus, ENH5 could have been positively selected for due to its pleiotropic activity that results in multiple adaptive phenotypes to hypoxia. Interestingly, ENH5 was also shown to be active in adipose tissue, suggesting a yet to be characterized physiological role. To dissect the role of ENH5 in adipocytes, we place wildtype and ENH5 knockout mice on a high fat diet to induce adiposity. Prior to being placed on the high fat diet, wildtype and ENH5 knockout mice were subjected to an intraperitoneal glucose tolerance test (IPGTT). Male ENH5 knockout mice were found to be more glucose intolerant than wildtype mice, suggesting that ENH5 has a role in metabolism in adipose tissue. This finding as well as additional insight that will be gained from histological analysis and RNA-seq from these mice after being placed on a high fat diet will provide yet another phenotype this enhancer effects and a better understanding of the role of pleiotropy in genetic adaptation of Tibetans.

Complex Traits Posters - Thursday

PB1342. Dose-responsive mRNA biomarkers of alcohol consumption and placebo response: A transcriptome-wide gene expression analysis.

Authors:

A. Shetty¹, J. Cornell², X-Q. Wang³, L. Sadzewicz¹, L. Tallon¹, C. Seneviratne^{1,2}; ¹Inst. for Genome Sci. (Univ. of Maryland), Baltimore, MD, ²Dept. of Psychiatry (Univ. of Maryland), Baltimore, MD, ³Dept.s of Publ. Hlth.Sci. (Univ. of Virginia), Charlottesville, VA

Abstract Body:

Purpose: Gene expression is complex, and highly sensitive to environmental factors that temporospatially interact with an individual's biology. Gene expression studies delineating alcohol's pharmacologic effects from placebo and other factors in living individuals are scarce. This study aimed to screen for mRNA biomarkers of varying drinking patterns utilizing a longitudinal study conducted in a highly controlled environment minimizing environmental effects. **Methods:** Total RNA from peripheral white blood cells (WBC) were extracted from whole blood collected from 8 healthy binge drinkers who were enrolled in a 12-day randomized, placebo-controlled, double-blind human laboratory trial conducted for testing a series of biomarkers of alcohol consumption (NCT04363424). Three alcohol doses - placebo, moderate, and high were administered in 3 separate 4-day long experiments each with 3 drinking sessions and 7-day washout period. The high dose corresponded to binge level drinking. Bulk RNAseq was performed using Illumina HiSeq4000 platform. Gene expression was normalized for library size and baseline expression at each dosing experiment. We assessed dose-responsive alcohol effects on gene expression in WBC using a generalized linear mixed-effects model. The model included individuals as a random effect and was adjusted for dose administration visit, alcohol dose and their interaction. Significantly differentially expressed genes between conditions (i.e., alcohol doses) were determined using a false discovery rate (FDR) of 10% and a minimum $\log_2(\text{fold-change})$ of 0.25. **Results:** 18 genes (e.g., CD74) were associated with response to placebo although EtG and BrAC measures did not change from baseline throughout placebo administered days. Compared to placebo-alcohol, the high-dose was associated with 36 genes (e.g., RRM2; CRHBP; SYNPO) that were enriched in 34 pathways, including pathways for cellular transporter activities. The middle-dose dose was associated with 64 genes (e.g., ACACA; BAX; CCNB1) that were almost exclusively enriched in pathways governing cellular component organization. Validation of findings with *NanoString* assays and associated DNA variation are on-going. **Conclusion:** By testing gene expression patterns in response to placebo, moderate and binge-level drinking within an environment that minimized confounding effects, we identified potential mRNA biomarkers representing pathophysiology underlying binge drinking that warrants further validation for clinical application. Furthermore, to our knowledge, comprehensive statistical modeling that we have tested here has not been applied at transcriptome-wide level.

Complex Traits Posters - Wednesday

PB1343. Drug repurposing for osteoporosis in men by integrating multi-omics and pharmacogenomics

Authors:

J. Greenbaum, C. Qiu, K-J. Su, D-Y. Liu, Q. Tian, H. Shen, H-W. Deng; Tulane Ctr. of BioMed. Informatics and Genomics, New Orleans, LA

Abstract Body:

Current pharmacological interventions for osteoporosis are categorized as either antiresorptive (e.g., bisphosphonates), which decrease the rate of bone resorption by osteoclasts, or anabolic (e.g., parathyroid hormone analogues), which stimulate bone formation by osteoblasts. Despite the remarkable advancements in our understanding of the disease pathogenesis, the current osteoporosis medications are not satisfactory and can have major side effects. We integrated the paired transcriptome and methylome profiles in peripheral blood monocytes (PBMs) from a multi-ethnic sample of Caucasian (n=558) and African American (n=359) men sampled from the Louisiana Osteoporosis Study (LOS) with interactome and pharmacogenomics information to identify novel drug candidates for osteoporosis treatment. Beginning from initial seed genes that are differentially expressed/methylated between low/normal BMD groups, we used nonparametric bootstrapping-based simulated annealing to derive driver signaling networks by growing subnetworks until a predefined network score was optimized. Next, we used drug similarity information from STITCH to reconstruct the drug-drug functional similarity network for >1,000 compounds that have drug induced gene expression profiles for skeletal muscle, which is the best available disease relevant tissue, in the LINCS L1000 dataset. The known targets of drugs were collected from various databases, and a recommendation algorithm was applied to predict the off-targets. Drugs were then ranked by evaluating the strength of their targeting effects on the driver signaling networks identified using the multi-omics profiles. The findings revealed several plausible repurposing candidates including lovastatin (statin), thalidomide (immunomodulator), carvedilol (beta blocker), flavopiridol (CDK inhibitor), and several polyphenol compounds such as quercetin and apigenin. Although still requiring functional validation experiments, the findings from this study could potentially reduce the cost and increase the speed for the development of novel therapies to treat osteoporosis and prevent fragility fractures.

Complex Traits Posters - Thursday

PB1344. Effect of artificial light on metabolic disorders and circadian gene expression

Authors:

N. Fatima; King George's Med. Univ., Lucknow, India

Abstract Body:

The large no. of world population daily exposed to the blue light, from few minutes to several hours. Blue light/ screen light provide negative effect on circadian rhythm which causes sleep deprivation and develops many diseases. Total 12 Wistar rats were enrolled and divided into two groups. Control group and Blue Light (BL) treated group which consist of six rats in each group. BL model was developed by placing the rats in 12hr blue light and 12 hr in dark till three months. Half of the rats were sacrificed and remaining rats were shifted to Normal Light (NL) for three months to see the legacy effect. These rats were sacrificed and blood samples were collected. Body weight was measured monthly, with blood glucose, Insulin, melatonin, lipid profile were estimated and mRNA were expressed by RTPCR. The percentage gain of body weight of BL treated group was **27.9%** as compared of control group (**20.2%**). Blood glucose levels were increased and circulatory level of insulin, melatonin, total cholesterol, TG, HDL and LDL were decreased in blue light treated rats. BL treated group were shifted on NL they showed increase in **2.3%** body weight, melatonin and total cholesterol and HDL were significantly increased (**p= 0.0258**) and (**p= 0.037**) respectively. Level of glucose were significantly decreased (**p>0.05**) and no change in insulin level. Per1 and Bmal1 gene were up regulated in BL group and the expression of these gene shows the significantly down regulated in NL group **p= 0.0394** and **p= 0.0403** respectively. We found that the exposure of artificial light increases the prevalence of obesity and metabolic disorders.

Complex Traits Posters - Wednesday

PB1345*. Effect of cortisol on cortical organoids: Building a "stress in a dish" model system.

Authors:

C. Purmann, K. Farrise, Y. Huang, R. Pattni, M. Ho, V. G. Carrion, A. E. Urban; Stanford Univ., Palo Alto, CA

Abstract Body:

Research shows that exposure to chronic stress and traumatic experience can impact brain health and development. For some individuals, it may lead to Post Traumatic Stress Disorder (PTSD) and other forms of psychopathology. However, less is known about the factors that make an individual more or less susceptible to developing severe psychopathology after stress exposure. Especially understanding of cellular and genomic mechanisms is lacking. Here, we present a model system to study chronic and acute stress in an *in vitro* brain organoid system that is accessible to genomic analyses. We differentiated cortical organoids from human iPSCs and confirmed the expression of the glucocorticoid receptor gene *NR3C1* and the response of the receptor to cortisol. Mature organoids were exposed to physiological stress levels using 1 μ M and 500 nM cortisol. The reaction to cortisol was analyzed using single-cell RNA-seq. We found differentially expressed genes that are known to respond to cortisol in particular and stress in general. We identified several genes that are linked to PTSD. We also discovered new differentially expressed genes that have not been linked to stress and PTSD previously. In summary, we have built a model system to study stress in a dish.

Complex Traits Posters - Thursday

PB1346. Effects of epigenetic age acceleration on kidney function: a mendelian randomization study.

Authors:

Y. Pan¹, X. Sun¹, Z. Huang¹, R. Zhang¹, C. Li², A. Anderson¹, L. James³, T. Kelly¹; ¹Tulane Univ., New Orleans, LA, ²Tulane Univ. Sch. of Publ. Hlth.and Tropical Med., New Orleans, LA, ³Univ. of Illinois Chicago, Chicago, IL

Abstract Body:

Previous studies have reported cross-sectional associations between measures of epigenetic age acceleration (EAA) and kidney function phenotypes. However, the temporal and potentially causal relationships between these variables remain unclear. We conducted a bidirectional two-sample Mendelian randomization (MR) study of EAA and kidney function. Genetic instruments for EAA and estimate glomerular filtration rate (eGFR) were identified from previous genome-wide association study (GWAS) meta-analyses. Causal effects of EAA on kidney function and kidney function on EAA were assessed through summary-based MR analyses utilizing data from the CKDGen GWAS meta-analysis of log-transformed estimated glomerular filtration rate (log-eGFR; n=567,460) and GWAS meta-analyses of EAA (n=34,710). A powerful, allele score-based MR leveraging individual-level data from UK Biobank participants (n=433,462) further examined the effects of EAA on kidney function. Using summary-based MR, we found that each 5-year increase in intrinsic EAA (IEAA) and Grim age acceleration (GrimAA) was associated with -0.01 and -0.02 unit decreases in log-eGFR, respectively (P=0.02 and P=0.09, respectively), findings which were strongly supported by allele-based MR study (both P<0.001). Summary-based MR identified 24% increased odds of CKD with each 5-unit increase in IEAA (P=0.05), with consistent findings observed in allele-score based MR (P=0.07). Reverse-direction MR identified potentially causal effects of decreased kidney function on Hannum Age acceleration (HannumAA), GrimAA, and PhenoAge acceleration (PhenoAA), conferring 3.14, 1.99, and 2.88 year decreases in HanumAA, GrimAA, and PhenoAA, respectively (P=0.003, 0.05, and 0.002, respectively) with each 1-unit increase in log-eGFR. This study supports bidirectional causal relationships between EAA and kidney function.

Complex Traits Posters - Wednesday

PB1347. Effects of the Tau Haplotype on Cognition in Childhood

Authors:

D. Dokuru¹, C. Reynolds², M. Stallings³; ¹Univ. of Colorado, Boulder, Boulder, CO, ²Univ. of California - Riverside, Riverside, CA, ³Univ Colorado, Boulder, CO

Abstract Body:

Progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) are rare neurodegenerative diseases characterized by aggregates of tau. Compared to diseases with mixed pathology like Alzheimer's, these disorders do not show presence of other protein aggregates. To date, the largest genetic risk factor identified in both diseases is the tau haplotype, an inversion of a ~900kb block on the long arm of chromosome 17 and includes MAPT, the gene encoding the tau protein. While the alternative haplotype with the inversion confers protection against the development of tauopathies, the impact of the haplotype on cognition is largely unknown. In this study, we are assessing the effects of the tau haplotype on cognitive outcomes in children ages 7-9 (prior to the development of sufficient tau aggregates). For the initial exploratory analysis we used data from 9,457 9-year old individuals in the Adolescent Brain Cognitive Development (ABCD) study. We used PLINK to extract the haplotype tagging variant rs8070723 (G) to determine H2 haplotype dosage. The phenotypes we tested in the study included composite scores from the NIH Toolbox® Cognition Battery (<http://www.nihtoolbox.org>) (fluid, crystallized, and total), the WISC matrix reasoning task (estimate for IQ), and total whole brain cortical area (mm²). We generated general linear models for each phenotype. We created the models using the H2 haplotype tagging variant as a dosage with values of 0,1,2. We also controlled for the covariates age, sex, and race/ethnicity (PCs). All statistical models were constructed in the R software package (<http://www.r-project.org/>). We used a stringent Bonferonni corrected threshold of 0.01 to determine significance. Aside from crystallized composite score, all other measure showed significant decrement in cognitive outcomes associated with the H2 haplotype (IQ was marginally significant). This study shows that while the H2 haplotype is protective for developing tauopathies in the aging population (post-reproductive age), it is associated with more negative cognitive outcomes in early age (pre-reproductive age). To confirm our findings we replicated them in the Colorado Adoption/Twin Study of Lifespan behavioral development (CATSLife) dataset (N=953) using the cognitive measures available when the participants were aged 7: performance WISC, verbal WISC, and full scale WISC. While tests of mean differences did not reach significance for any of the measures in this small replication, the direction of effects was consistent with the initial findings.

Complex Traits Posters - Thursday

PB1348. Elevated levels of CXCL16 in severe COVID-19 patients: Effects on mortality.

Authors:

C. Johnson, J. Brzezynski, A. Kaden, S. Smieszek, V. Polymeropoulos, B. Przychodzen, C. Polymeropoulos, G. Birznieks, M. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC

Abstract Body:

Genome-wide association studies (GWAS) have recently identified chromosome 3p21.31, with a lead variant pointing to the *CXCR6* gene, as the strongest reported susceptibility locus for severe COVID-19 thus far. CXCL16 is synthesized as a transmembrane molecule that is expressed as a cell surface-bound molecule and as a soluble chemokine. The CXCR6/CXCL16 axis mediates homing of T cells to the lungs in a diseased state, and hyper-expression is associated with localized cellular injury. The aim of this GWAS was to characterize the CXCR6/CXCL16 axis in the pathogenesis of severe COVID-19. Plasma concentrations of CXCL16 collected at baseline from 115 hospitalized COVID-19 patients participating in the ODYSSEY clinical study were assessed together with a set of controls. Another cohort of samples (n = 79) from COVID-19 patients participating in the CALYPSO clinical study were evaluated to see if the effect could be replicated. CXCL16 levels in plasma were determined using enzyme-linked immunosorbent assay (ELISA). Furthermore, whole genome sequencing was conducted on all samples.

We report elevated levels of CXCL16 in this cohort of COVID-19 hospitalized patients, specifically in severe hospitalized COVID-19 patients (p -value < 0.02). Importantly, we report a significant effect of elevated CXCL16 on the mortality of this cohort (p -value < 0.03) and show that the effect is replicated in our second cohort (p -value < 0.0004). Clinically, at 700 pg/mL, the odds ratio is 20.6 (p -value = 0.04), suggesting that a COVID-19 patient has approximately a 25% mortality rate when CXCL16 levels are above 700 pg/mL. We further characterize the role of the *CXCR6* expression on CD8 T cells.

These latest findings further support the significant role of the CXCR6/CXCL16 axis in the immunopathogenesis of severe COVID-19 and warrant additional studies to understand which patients would benefit most from targeted treatments.

Complex Traits Posters - Thursday

PB1349*. Enriching genetic hypotheses of schizophrenia through neuroimaging transcriptomics.

Authors:

X. Bledsoe¹, E. Gamazon²; ¹Vanderbilt Univ., Nashville, TN, ²VUMC Clare Hall, Univ. of Cambridge, Nashville, TN

Abstract Body:

The pathophysiology of schizophrenia is highly complex, involving associations with hundreds of SNPs and genes. The effort to synthesize these disparate genetic elements into pathophysiological risk pathways is far from complete. Biochemical studies are complemented by neuroimaging data which highlight altered brain regions in schizophrenic patients. Hippocampal variation has emerged as a highly replicable neurologic markers of schizophrenia risk and severity.

We use novel methods to prioritize a high-confidence set of genes and cellular functions as candidate pathophysiological intermediates in schizophrenia. Integrating data from the United Kingdom Biobank neuroimaging study and 17 neuro-relevant in silico gene expression models from the Genome Tissue-Expression consortium, we conduct transcriptome-wide association studies for 18 quantitative measures of the hippocampus in healthy patients. We then query pre-existing data to identify the subset of these hippocampus-associated genes that colocalize to topologically associated domains (TADs) that are disrupted in neurologic iPSCs derived from schizophrenic patients.

We show that 123 of the genes impacted by TAD disruptions in cells from schizophrenic patients are significantly associated with morphological change in the hippocampus according to a study-wide Bonferroni corrected p-value threshold. We identify 115 genes that both localize to genomic regions disrupted in schizophrenogenic neuronal iPSCs and demonstrate significant associations with hippocampal morphology via genetically regulated expression. We identify 47 hippocampus-associated genes in disrupted TADs from schizophrenic glial iPSCs and 108 hippocampus genes in disrupted TADs from schizophrenic neural progenitor cell iPSCs. Through gene set enrichment analyses and systematic chart reviews of the pathways and cell functions impacted by these genes, we identify themes of neuronal morphogenesis, axonal development, cell adhesion, neurotransmitter release, and prior association with neurodegenerative disorders.

Through the integration of SNP data, gene expression, 3D genomic organization, neuroanatomical variation and clinical phenotyping, this work traces the impact of schizophrenia-associated variation through multiple levels of biological organization. We highlight a key set of candidate genetic contributors to schizophrenic risk along with putative cellular mechanisms of action. Through this novel integration of neuroimaging transcriptomic data with 3D genomic studies we demonstrate the utility of this study design to inform mechanisms of complex disease.

Complex Traits Posters - Wednesday

PB1350. Enrichment of Patients with Ehlers Danlos Syndrome in Idiopathic Gastroparesis - A Gene Set Enrichment Analysis.

Authors:

J. Carlin, S. Smieszek, M. Fisher, D. Madonick, C. Kupersmith, P. Moszczynski, C. Xiao, G. Birznieks, M. H. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC

Abstract Body:

Ehlers Danlos Syndrome (EDS) is a heritable disorder of the connective tissue. It is usually inherited as an autosomal dominant trait. EDS comprises a genetically heterogeneous group of connective tissue disorders of which the main clinical features are joint hypermobility, skin hyperextensibility, delayed wound healing with atrophic scarring, and generalized connective tissue fragility. The genetic etiology is not well explained. In the current gastroparesis study, we report 16 cases of EDS (confirmed by electronic medical records), specifically enriched in the cohort of idiopathic gastroparesis. Assuming incidence rate of 1 in 5000 the enrichment is 160-fold ($p=8.18743E-26$). In order to explore the genetic surrogates of EDS we have done gene set enrichment analysis on core EDS genes: COL1A1, COL1A2, COL3A1, COL5A1, and COL5A2 and ADAMTS2, FKBP14, PLOD1, and TNXB, as reported in LOVD. All EDS cases are of Caucasian ancestry, females with idiopathic gastroparesis. We report all coding/splicing variants in the core genes with a minor allele frequency (MAF) <0.05 . Most of our EDS cases with exception of 1, have at least one such nonsynonymous variant/or splicing variant within the core gene set. We report 2 duplicate variants across this set. The first variant is COL5A1:NM_000093:exon13:c.G1588A:p.G530S (rs61735045). It has a global MAF in gnomAD of 0.03 (ranging from 0.04 NFE to 0.0009 amongst East Asian). The variant is reported in ClinVar as pathogenic and associated with EDS, classic type and has a high CADD score of 24.9. The past findings suggest that heterozygosity for the Gly530Ser substitution is associated with mild ultrastructural abnormalities, while homozygosity for this mutation is associated with classical EDS. Furthermore, the variant is significant in association analysis of idiopathic versus diabetic gastroparesis patients. We report 5/128 diabetic versus 27/238 idiopathic, $p\text{-value}<0.007$ (EDS gene set). The other variant duplicated in our cohort is the TNXB:NM_019105:exon26:c.A9044G:p.K3015R, with a high CADD score of 23.3. Additionally, a gene enrichment analysis was performed on core genes, $MAF<5$, idiopathic versus diabetic gastroparesis, and led to a significant result, specifically enrichment of variants in COL5A1 and TNXB. These results suggest a statistically significant (corrected $p\text{-value}<0.001$) enrichment of EDS patients in this gastroparesis study, with confirmed known and novel variants in the EDS gene set. Furthermore, these results suggest an enrichment of such gene variants in the entire idiopathic cohort of patients with idiopathic gastroparesis.

Complex Traits Posters - Thursday

PB1351. Epstein-Barr virus- and genotype- dependent transcriptional regulation in B cells from patients with multiple sclerosis

Authors:

M. Granitto¹, S. Parameswaran¹, C. Forney¹, O. Donmez¹, K. Dunn¹, A. Diouf¹, C. Yin¹, A. Zabeti², L. Kottyan¹, M. Weirauch¹; ¹Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, ²Univ. of Cincinnati Med. Ctr., Cincinnati, OH

Abstract Body:

Multiple sclerosis (MS) is caused by genetic and environmental risk factors. To date, over 200 robust genetic associations have been identified through GWAS. Epstein-Barr virus (EBV) is the most consistently replicated environmental factor. We aim to globally identify the contribution of EBV and genotype-dependent transcriptional regulation in B cells. We discovered that 44 of the 109 MS-associated genetic loci contain an MS-associated variant directly located within a human genomic region occupied by the EBNA2 protein, with our latest unpublished data implicating EBV-encoded EBNA2, EBNA3C, and EBNA-LP proteins at 93 of the now 201 MS risk loci. A subsequent functional genomics study by our group revealed more than 400 EBNA2-dependent human genes, over 2,000 regions in the human genome with EBNA2-dependent chromatin accessibility, and more than 1,700 regions where EBNA2 alters chromatin looping interactions. We isolated peripheral blood B cells from patients with MS and controls. We created EBV-transformed B cell lines (LCLs) from each subject. We measured gene expression (RNA-seq), chromatin accessibility (ATAC-seq), and binding of EBNA2 co-factors RBPJ and PU.1 from primary B cells and LCLs using ChIP-seq. Genome-wide EBNA2 binding was also measured in LCLs. Strikingly, no genes were differentially expressed in primary B cells from patients with MS vs controls. In contrast, several hundred genes were differentially expressed in EBV B cell lines derived from MS patient primary B cells compared to EBV B cell lines derived from primary B cells of control subjects. These differences were not due to differences in the expression of EBV-encoded genes. We observed extensive genotype-dependent binding of EBNA2, RBPJ, and PU.1 at MS risk loci. Ongoing analyses focus on identifying MS-dependent binding of EBNA2, RBPJ, and PU.1 and integrating these data with differential gene expression data to comprehensively catalog EBV and MS-dependent gene regulation at MS risk loci in B cells.

Complex Traits Posters - Wednesday

PB1352. ERAP1, ERAP2, and two copies of HLA-Aw19 alleles increase the risk for Birdshot Chorioretinopathy in HLA-A29 carriers

Authors:

S. Gelfman¹, D. Monnet², A. Ligocki³, T. TABARY⁴, A. Moscati³, X. Bai¹, J. Freudenberg¹, B. Cooper³, J. Kosmicki¹, S. Wolf¹, Regeneron Genetics Center, GHS-RGC DiscovEHR Collaboration, M. Ferreira¹, J. Overton¹, J. Weyne³, E. Stahl¹, A. Baras¹, C. Romano³, J. Cohen⁴, G. Coppola¹, A. Brézin²; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Université de Paris, Hôpital Cochin, service d'ophtalmologie, Paris, France, ³Regeneron Pharmaceuticals, Tarrytown, NY, ⁴Univ. of Reims Champagne Ardennes, LRN EA4682, Reims, France

Abstract Body:

Purpose Birdshot Chorioretinopathy (BSCR) is strongly associated with HLA-A29. This study was designed to elucidate the genetic modifiers of BSCR in HLA-A29 carriers. **Methods** We sequenced the largest BSCR cohort to date, including 286 cases and 108 HLA-A29 positive controls to perform genome wide common and rare variant associations. We further typed the HLA alleles of cases and 45,386 HLA-A29 controls of European ancestry to identify HLA alleles that associate with BSCR risk. **Results** Carrying a second allele that belongs to the HLA-Aw19 broad antigen family (including HLA-A29, A30, A31, and A33) increases the risk for BSCR (OR=4.44, p=2.2e-03). This result was validated by comparing allele frequencies to large HLA-A29-controlled cohorts (n=45,386, OR>2.5, p<1.3e-06). We also confirm that ERAP1 and ERAP2 haplotypes modulate the risk for disease within our HLA-A29 controlled cohort. A meta-analysis with an independent dataset confirmed that ERAP1 and ERAP2 haplotypes modulate the risk for disease at a genome-wide significant level: ERAP1-rs27432 (OR 2.46; 95% CI 1.85-3.26; p=4.07e-10), an eQTL decreasing ERAP1 expression, and ERAP2-rs10044354 (OR 1.95; 95% CI 1.55-2.44; p=6.2e-09), an eQTL increasing ERAP2 expression. Furthermore, ERAP2-rs2248374 that disrupts ERAP2 expression is protective (OR 0.56; 95% CI [0.45-0.70]; p=2.39e-07). BSCR risk is additively increased when combining ERAP1/ERAP2 risk genotypes with two copies of HLA-Aw19 alleles (OR 13.53; 95% CI 3.79-54.77, p=1.17e-05). **Conclusions** The genetic factors increasing BSCR risk on the background of HLA-Aw19 alleles include gene variants associated with increased processing and presentation of ERAP2 specific peptides. This finding suggests that exceeding a peptide presentation threshold may predispose to BSCR by evoking an aberrant immune response in choroids of A29 carriers

Complex Traits Posters - Thursday

PB1353. Essential hypertension genomic regions from linkage analysis studies

Authors:

V. Magalhaes Borges¹, A. V. R. Horimoto², R. C. Mingroni-Netto³, A. Nato¹; ¹Dept. of BioMed. Sci., Joan C. Edwards Sch. of Med., Marshall Univ., Huntington, WV, ²Univ. of Washington, Seattle, WA, ³Univ de Sao Paulo, Sao Paulo, Brazil

Abstract Body:

Essential hypertension (EH) is a major risk factor for cardiovascular diseases, and it causes ~9.4 million deaths per year worldwide. Hypertension is a complex disease characterized by high systolic blood pressure (SBP) and/or high diastolic blood pressure (DBP). Reference values used by the International Society of Hypertension (2020 Hypertension Practice Guidelines) are SBP \geq 140 and/or DBP \geq 90 mmHg. The 2015 World Health Organization global data indicated that 22.1% of the worldwide population has hypertension (the prevalence was 24.6% in 2000). Family/twin studies revealed an estimated global population heritability of 15-60%. In the last few decades, a vast scientific production revealed several genome-wide genomic regions for the EH phenotype, many of which were later corroborated by other approaches. Our goal is to identify, cross examine, and compare these genomic regions implicated by independent linkage analysis studies for EH, then rank them based on LOD score and the impact of genetic factors (genes and variants). Identification of genomic regions implicated in EH from various studies allows researchers to focus on a smaller sample space instead of the entire genome. We evaluate 25 linkage analysis studies from 1999 until 2021, all of them using LOD score as the metric for linkage studies. The genomic regions were standardized for the GRCh37/hg19 genome build. Using R scripts, marker names and genetic locations from these linkage studies, and from Rutgers Combined Linkage-Physical Map (NCBI34/hg16, NCBI35/hg17 and NCBI36/hg18) were used to define the boundaries of each ROI and 26 ROIs were identified. The genomic regions found with the highest LOD score are: 2p25-p24 (LOD score 7.57) and 17q12 (LOD score 4.7). The genomic regions containing the most relevant genetic factors known to EH, are: 1q42-43 (AGT gene), 2p25.1-p22.3 (HYT3 locus), 7q36-q22 (CYP3A5 and NOS3 genes), 12p12.2-p13 (HYT4 locus and GNB3 gene), 12q21-q23 (ATP2B1 gene) and 17q21-q25 (HYT1 locus). Narrowing down these regions by investigating variants has the potential to elucidate EH etiology.

Complex Traits Posters - Wednesday

PB1354*. Estimating the causal influence of body mass index on gut microbiome variation

Authors:

N. Timpson¹, C. Hatcher¹, L. Corbin¹, K. Wade², J. Raes¹, **D. Hughes¹**; ¹Bristol Univ., Bristol, United Kingdom, ²Univ. of Bristol, Bristol, United Kingdom

Abstract Body:

Variation in the accumulation of body fat, body composition and obesity can all be proxied by body mass index (BMI) which itself is a recognized risk factor for numerous health outcomes including life expectancy, various cancers and cardiometabolic diseases. What remains unclear is the complete picture of factors influencing BMI. One possible factor is gut microbiome variation, which previous research has demonstrated to be strongly correlated with BMI. Critically, whilst it may be the case that human gut flora variation influences BMI, it remains at least equally - if not more - likely that the opposite effect is true. Here, we use a combination of genetic contributions to both BMI and faecal 16S rRNA sequencing data from the Flemish Gut Flora Project (n=2257) to estimate causal associations between BMI and gut flora - microbiome profiles being a likely downstream effect of BMI and the factors related to variation in this risk factor at a population level. Mendelian randomization (MR) analysis, using a polygenic score reliably associated with BMI, yielded evidence from total and sex-specific sub-analysis suggesting the presence of causal effects of BMI on gut flora. In total, 215 microbiome traits (MTs), including three alpha-diversity, one beta-diversity, 115 presence (vs. absence), 40 zero-truncated abundance, and 51 abundance traits, were included in all analyses. After correcting for multiple testing, 67 MTs were associated with standard deviation change in BMI ($\sim 4.5\text{kg/m}^2$) in (generalized) linear modelling. MR-derived effect estimates were strongly correlated with estimates from (generalized) linear models. After correcting for multiple testing, 14 MTs retained evidence for a strong causal association with BMI in the total population. These include genera *Sporobacter* and *Barnesiella*, family Porphyromonadaceae, beta-diversity MDS1, two alpha diversity metrics and enterotype Bact2. Sex-specific MR analyses showed that 10 MTs were persistently associated with BMI in females, but 0 MTs were associated with BMI in males specifically. All but one of these associations overlapped with those seen in the total population, the phytoestrogen metabolizing genera *Adlercreutzia*, which appeared for females only. Overall, results support a conclusion that gut microbiome variation can be causally influenced by variation in BMI and that observational studies of the microbiome as a risk factor need to account for likely reverse causality as one source of association when analysing BMI.

Complex Traits Posters - Thursday

PB1355. Ethnicity- and sex-specific genome-wide association study on Parkinson's disease

Authors:

S. Chung¹, K. Park¹, H-S. Ryu², S. Jeon¹, S. Kim¹, J. Kim¹; ¹Asan Med. Ctr., Seoul, Korea, Republic of, ²Kyungpook Natl. Univ. Hosp., Daegu, Korea, Republic of

Abstract Body:

Background: Most previous genome-wide association studies (GWASs) on Parkinson's disease (PD) focus on the European population. In addition to ethnicity, there are several sex-specific clinical differences in PD, but little is known about its genetic background. **Objectives:** We aimed to investigate ethnicity-specific, and sex-specific GWAS on PD in the Korean population. **Methods:** A total of 1,050 Korean PD patients and 5,000 controls were included. For primary analysis between, we performed a GWAS using a logistic additive model adjusted for the age and sex using a Korean Chip. Same statistical models were also applied to sex-specific analyses, between 554 female PD patients and 2,610 female controls and 496 male PD patients and 2,390 male controls. **Results:** Eight single nucleotide polymorphisms (SNPs) including four in the *SNCA* locus and three from the *PARK16* locus were associated with PD in Koreans. The rs34778348 in the *LRRK2* locus showed a strong association but failed to pass cluster quality control. There were no notable genome-wide significant markers near the *MAPT* or *GBA* loci. In the female-only analysis, rs34778348 in *LRRK2* and the four other SNPs in the *SNCA* showed strong association with PD. In the male-only analysis, no SNP surpassed the genome-wide significance threshold under Bonferroni correction; however, the most significant signal was rs708726 in the *PARK16* locus. **Conclusion:** This ethnicity- and sex-specific GWAS on PD implicates the pan-ethnic effect of *SNCA*, the East Asian-specific role of the *PARK16* and *LRRK2* G2385R variants, and the possible disproportionate effect of *SNCA* and *PARK16* between sexes for PD susceptibility. These findings suggest the different genetic contributions to sporadic PD in terms of ethnicity and sex.

Complex Traits Posters - Wednesday

PB1356. Evaluating the causal effect of tobacco smoking on whitematter brain aging: a Mendelian randomization analysis in UK Biobank

Authors:

C. Mo¹, J. Wang², Z. Ye³, H. Ke⁴, S. Liu⁵, K. Hatch⁶, S. Gao⁶, J. Magidson⁷, C. Chen⁶, B. Mitchell⁸, P. Kochunov⁶, L. Hong⁶, T. Ma⁷, C. Shuo⁶; ¹Harvard Med. Sch., Boston, MA, ²Qilu Hosp. of Shandong Univ., Jinan, China, ³Univ. of Maryland, Rockville, MD, ⁴Univ. of Maryland, Coll. Park, College Park, MD, ⁵Qilu Univ. of Technology (Shandong Academy of Sci.), Jinan, China, ⁶Univ. of Maryland Sch. of Med., Baltimore, MD, ⁷Univ. of Maryland, College Park, MD, ⁸Univ Maryland, Baltimore, Baltimore, MD

Abstract Body:

Background and aims: Tobacco smoking is a known risk factor associated with the accelerated decline of brain structures and functions during aging. However, the causal effect of smoking on brain aging remains unclear. To close this knowledge gap, we performed Mendelian randomization (MR) analysis to estimate the causal effect of smoking on white matter brain aging using genetic and neuroimaging data. **Design:** We used the generalized weighted regression model, extended from the inverse-variance weighted model, for causal effect estimation based on multiple correlated genetic instrumental variables (IVs) in an MR analysis. **Measurements:** The primary exposures were smoking status (SS) and cigarette per day (CPD), and the IVs were the genetic variants that met the MR analysis assumptions. The outcome variable was the “Brain Age Gap” (BAG), a neuroimaging-based metric that characterizes the progression level of age-related neural decline. We constructed the optimal model for BAG estimation based on a training set independent of a testing set used in the MR analysis. **Setting and participants:** The study included N = 23,624 subjects who have genotype, neuroimaging, and smoking data available from the UK Biobank. Among them, 10,717 non-smokers were used to train the BAG predictive model. The remaining 10,746 non-smokers and 2,161 smokers constituted the testing set for genetic-outcome association tests in MR analysis. **Findings and conclusions:** We found adverse causal effects of smoking behaviors on the brain aging process. The brain age of smokers was 0.37 years (causal effect = 0.37, 95% confidence interval (CI) = 0.10, 0.64; p-value = 6.4×10^{-3}) older than comparable non-smokers. Also, an extra cigarette per day increased brain age by 0.16 years (causal effect = 0.16, 95% CI = 0.07, 0.25; p-value = 5×10^{-4}). These suggested that smoking caused accelerated brain aging and supported previous findings of smoking-related cognitive decline.

Complex Traits Posters - Thursday

PB1357. Evaluating the frequency and impact of structural variation in amyotrophic lateral sclerosis.

Authors:

A. Dillio^{1,2}, S. Griese², G. A. Rouleau^{1,2}, The International Consortium on Amyotrophic Lateral Sclerosis Genetics (ALSGEN), The FALS Sequencing Consortium, S. M. K. Farhan^{1,2}; ¹Montreal Neurological Inst.-Hosp., Montreal, QC, Canada, ²McGill Univ., Montreal, QC, Canada

Abstract Body:

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease. In our recent study of 3,864 ALS cases and 7,839 ancestry matched controls, we observed a significant excess of rare protein-truncating variants among ALS cases, which was concentrated in constrained genes (Farhan et al., 2019). We also observed multiple distinct protein-truncating variants in a highly constrained gene, *DNAJC7*. While the study advanced our understanding of the genetic architecture of ALS, one major limitation was it did not consider larger variation, including structural or copy number variants (CNVs). Structural variants within the genome are known to be important risk factors in neurodegenerative disease risk, such as the hexanucleotide repeat expansion in *C9orf72* (DeJesus-Hernandez et al., 2011; Renton et al., 2011) implicated in ALS; however, there is little information on other structural variation, particularly CNVs, underlying ALS as they have been under studied given their complexity, and consequently, detection difficulties using conventional variant calling approaches. Here, we are applying an optimized dosage sensitive genomic framework on a sequencing dataset of 4,828 ALS cases and 14,091 controls to identify structural variant signatures unique to ALS cases. Following the generation of a joint called sequencing dataset, we used an optimized workflow, GATK-gCNV, and appropriate quality control to generate a CNV callset outlining genomic gains and losses. We are now applying logistic regression models incorporating covariates such as sample sex and population structure to detect signals of enrichment of genomic gains (duplications) and/or losses (deletions) in the ALS cases. Preliminarily, gene-set enrichment analyses suggest there is a significant enrichment of CNVs in ALS associated genes in the ALS cases versus controls, primarily resulting from deletions. Further, in addition to several potential significant associations between CNVs in novel genes and ALS, we have identified an association between the known genomic disorder associated CNV, 16p11.2 duplication, and ALS case status. As we continue to explore our preliminary findings, our analyses are undergoing replication by our collaborators at King's College London (Project MinE ALS) using the same optimized approach on 4,322 ALS cases and 1,832 age-matched controls, with the goal of harmonizing our data and performing a meta-analysis from the two independent datasets. Our study is among the first to advance our knowledge of the frequency and impact of structural variation in neurodegeneration — specifically ALS — using a large, powerful sample size.

Complex Traits Posters - Wednesday

PB1358. Evaluation of genetic support for targeting the IL-23 pathway: current and future therapeutic opportunities

Authors:

A. Hart¹, K. Sieber¹, J. Molineros¹, FinnGen, A. Fourie², D. Cua¹, D. Waterworth¹; ¹Janssen R&D, Spring House, PA, ²Janssen R&D, La Jolla, CA

Abstract Body:

Interleukin 23 (IL-23) is a key pro-inflammatory cytokine implicated in the pathogenesis of immune-mediated disease. To date, therapies targeting IL-23 through the *IL23A*-encoded p19 subunit, or the *IL12B*-encoded p40 subunit common to IL-12 and -23, have been approved for the treatment of psoriasis (PsO), inflammatory bowel disease (IBD), and psoriatic arthritis (PsA); however, there may be opportunities for expansion into new indications. Given that genetic evidence provides links to causal human biology, genetics can be used to evaluate target-indication pairs and provide rationale for new indications. Here, using a combination of fine-mapping and colocalization analyses, we comprehensively evaluated genetic evidence for ten key genes within the IL-23 pathway with the goal of identifying additional indications that may be targeted by IL-23 blockade.

We conducted meta-analyses of 24 immune-mediated diseases utilizing summary statistics from published GWAS, FinnGen, and UK Biobank and performed statistical fine-mapping and colocalization (coloc) with eQTLs from GTEx and the EBI eQTL Catalogue to nominate potential causal genes at disease-associated loci. We evaluated variant-to-gene mapping evidence (V2G) for loci at *IL23A*, *IL23R*, *IL12A*, *IL12B*, *IL12RB1*, *STAT3*, *JAK2*, *TYK2*, *IL17A*, and *IL17F*. Overall, we found genetic support consistent with the indications approved for therapies targeting *IL23A* and *IL12B*. For example, we found coding evidence for *IL23R* (R381Q) and coloc evidence for *IL12B* for IBD, PsO, and PsA. We observed very different patterns of V2G evidence for *IL12A* (IL-12 p35 subunit which forms heterodimer with p40 to make IL-12), where significant associations were identified with multiple sclerosis (MS), SLE, Sjogren's, Celiac disease, rheumatoid arthritis, and primary biliary cholangitis; colocalization demonstrated that risk was associated with decreased *IL12A* expression in these diseases. Interestingly, blockade of IL-12 and -23 (through the shared p40 subunit) failed to show efficacy in several of these indications (MS, SLE, RA).

In this study, we used data from well-powered meta-analyses to nominate target genes for immune-mediated diseases, and subsequently evaluated the genetic evidence for IL-23 pathway genes to guide indication expansion efforts. We largely identified associations with known indications such as IBD, PsO, and PsA, but also generated evidence for where p40 inhibition has not been or is not likely to be successful. Overall, our results demonstrate the utility of genetic data to support target and indication selection with the aim of bringing new medicines to patients.

Complex Traits Posters - Thursday

PB1359. Evaluation of mitochondrial DNA variation in autism and neurodevelopmental disorders

Authors:

Y. Wang¹, X. Guo¹, X. Hong², G. Wang², A. G. Clark^{3,4}, K. O. O'Brien¹, X. Wang^{2,5}, Z. Gu¹; ¹Div. of Nutritional Sci., Cornell Univ., Ithaca, NY, ²Ctr. on Early Life Origins of Disease, Dept. of Population, Family and Reproductive Hlth., Johns Hopkins Univ. Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ³Dept. of Molecular Biology and Genetics, Cornell Univ., Ithaca, NY, ⁴Dept. of Computational Biology, Cornell Univ., Ithaca, NY, ⁵Dept. of Pediatrics, Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Abstract Body:

Mitochondrial oxidative phosphorylation is essential for brain development. Its function is partially encoded by the mitochondrial genome (mtDNA). Defects in the quality and quantity of mtDNA have been previously suggested to be associated with neurological and developmental deficits in children. Here, we investigated the associations of mtDNA heteroplasmies (co-existence of mutated and unmutated mtDNA) and content, markers for mtDNA quality and quantity, with neurodevelopmental disorders in two independent samples: a family-based study (using peripheral blood samples from 1,938 families with parents, probands with autism and sibling controls in the Simons Simplex Collection [SSC]) and a prospective birth cohort (using peripheral/cord blood samples from 997 mother-child pairs, including 621 children with a future diagnosis of neurodevelopmental disorder and 376 children with neurotypical development in the Boston Birth Cohort [BBC]). We evaluated mtDNA heteroplasmies and content using data from whole-genome sequencing in the SSC and mtDNA-targeted sequencing (STAMP) in the BBC. In both samples, we obtained an average coverage of ~4,000-fold depth of unique reads on mtDNA. We found that predicted pathogenic (PP) heteroplasmies in children were associated with an increased risk of autism (Meta-OR=1.56, $P=0.00068$; $P<0.05$ in both samples), which contributed to ~3% (population attributable risk) of autism cases in the current study. In children diagnosed with autism, those carrying PP heteroplasmies had worse cognition (verbal and non-verbal IQ), adaption (VABS composite score), and motor functions (VABS motor skills). Furthermore, we detected decreased mtDNA content in children with autism and/or other neurodevelopmental disorders in both SSC and BBC. Interestingly, among children with autism and PP heteroplasmies in the SSC, increased mtDNA content showed benefits for cognition, communication, and behaviors, suggesting that increased mtDNA quantity may complement defective mtDNA quality in autism.

Collectively, our results from two independent and complementary study cohorts found a decline in the quality and quantity of mtDNA in children with autism and potentially other neurodevelopmental disorders. These alterations may be detectable years before the diagnosis of diseases. Comprehensive evaluation of mtDNA variation using sequencing approaches may offer predictive value for neurodevelopmental diseases in childhood.

Complex Traits Posters - Wednesday

PB1360. Evaluation of polygenic risk scores to differentiate between type 1 and type 2 diabetes.

Authors:

M. Shoaib¹, Q. Ye¹, M. Boehnke², C. Burant², S. Soleimanpour², S. Gagliano Taliun¹; ¹Montreal Heart Inst., Univ. of Montreal, Montreal, QC, Canada, ²Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Polygenic risk scores (PRS) quantify the genetic liability to disease and are calculated using an individual's genotype profile and disease-specific genome-wide association summary statistics. Type 1 (T1D) and type 2 (T2D) diabetes both are determined in part by genetic loci across the genome. Correctly differentiating diabetes subtypes is crucial for accurate diagnosis and treatment. PRS have the potential to address possible misclassification of T1D and T2D for eventual use in clinical practice through genetic testing. Here we evaluated PRS models for T1D and T2D in UK Biobank (UKB) participants. Specifically, we investigated the utility of both T1D and T2D PRS models to discriminate between diabetes subtypes (T1D and T2D) and controls in unrelated UKB individuals of European ancestry; T1D (n=2383, n=494 with age diabetes diagnosed ≤ 25 years), T2D (n = 15,284, n=7669 with age of diabetes diagnosed ≥ 45 years), and non-diabetic controls (n = 297,700). We split the data into a training and test sets for model training (via the clumping and thresholding approach) and subsequent testing of predictive accuracy. We derived PRS models from disease risk SNPs from external non-UKB genome-wide association studies (GWAS). The T1D PRS model with the best discrimination between T1D cases and controls (AUC=0.805) also yielded the best discrimination of T1D from T2D cases of European ancestry in the UKB (AUC=0.792) and provided modest separation in an independent validation cohort (AUC=0.686). The same model did not perform well when we applied it to individuals of another UKB ancestry group, Indian British (AUC=0.574). In contrast to the T1D model, the best T2D model did not discriminate between the two diabetes subtypes (AUC=0.527). Our analysis suggests that a T1D PRS model based on independent SNPs may help differentiate between T1D, T2D, and controls in European ancestry individuals.

Complex Traits Posters - Thursday

PB1361. Evidence of allelic series with fine-mapped protein quantitative trait loci across 1,470 protein abundance measurements and 30 biochemistry measurements in UK Biobank participants.

Authors:

A. Cortes¹, N. Bowker¹, Q. Wang¹, E. Arciero¹, C. Robins², P. Surendran¹, J. Sandhuria¹, J. Davitte³, E. Ingelsson⁴, M. Dermitzakis¹, T. Johnson⁵, R. Scott¹; ¹GSK, Stevenage, United Kingdom, ²GSK, Atlanta, GA, ³GlaxoSmithKline, Exton, PA, ⁴GlaxoSmithKline Pharmaceuticals, South San Francisco, CA, ⁵GlaxoSmithKline, Stevenage, United Kingdom

Abstract Body:

The UK Biobank Pharma Proteomics Project (UKB-PPP Consortium) has quantified 1,470 proteins in plasma samples from over 53,000 UK Biobank participants, greatly enabling identification of pQTLs and subsequent causal inference on the relationship between proteins and disease. Here we present fine-mapping results from genome-wide analysis of protein quantitative trait loci in individuals of European ancestry. We quantify the number of conditional independent associations for each protein in the cis region of the corresponding gene and perform colocalization analysis with fine-mapped GWAS statistics for 30 biochemistry traits from the UK Biobank. We investigate and identify evidence of allelic series where multiple independent signals observed for a protein colocalise with multiple independent signals in a biochemistry trait. For PCSK9 abundance, we found evidence of allelic series with LDL and total cholesterol, where multiple conditionally independent signals (each with p -value $< 5 \times 10^{-6}$ in a multivariate analysis) showed evidence of shared causal variant ($PP_{\text{coloc}} > 0.90$) with a conditionally distinct cholesterol association. With this example we illustrate how perturbation of PCSK9 can be instrumented by more than one genetic variant and how we can investigate a dose dependent effect of PCSK9 modulation on LDL levels. This example corroborates genetic and observational studies on the causal role of PCSK9 function and lipid levels and the therapeutic benefit to treat cardiovascular disease.

Complex Traits Posters - Wednesday

PB1362. Examination of PTSD symptoms in older veterans as a function of AD risk genes and combat implicates *CLU* as a stress-response gene

Authors:

M. Logue^{1,2}, E. J. Wolf¹, M. W. Miller¹, R. Zang¹, R. Sherva², K. M. Harrington^{3,2}, F. R. Jennifer^{4,2,5}, Z. Neale^{1,2}, the Million Veteran Program; ¹Natl. Ctr. for PTSD at VA Boston, Boston, MA, ²Boston Univ. Sch. of Med., Boston, MA, ³MAVERIC, VA Boston Hlth.care System, Boston, MA, ⁴TRACTS, VA Boston Hlth.care, Boston, MA, ⁵Harvard Med. Sch., Boston, MA

Abstract Body:

Studies have suggested that late onset Alzheimer's disease (AD) risk genes, and in particular the high-risk *APOE* $\epsilon 4$ variant, may interact with combat exposure to increase posttraumatic stress symptoms (PTS) particularly in older veterans. However, results have been inconsistent, and large-scale posttraumatic stress disorder (PTSD) genome-wide association studies (GWASs), which include younger participants and non-veterans, have not identified PTSD risk variants in the *APOE* region. This study examined associations between survey-based current PTS (PTSD Checklist - Civilian version (PCL-C), G x E effects involving AD risk variants, and self-report combat exposure (Deployment Risk and Resilience Inventory) using cohorts of combat-deployed European ancestry (EA) and African ancestry (AA) veterans from the Million Veteran Program (MVP). As we were interested in PTS as a function of aging, and based on differential risk of AD, we included veterans age 45 and older split into a middle age group (ages 45 to 64; n=23,167 and 4,912 in EA and AA cohorts respectively) and an older age group (65+; n=68,707 and 5,353 in the EA and AA cohorts). Regression models assessed the association of PTS with interactions between combat and $\epsilon 4$ in the EUR and AA cohorts. In the EUR cohort, we also examined an AD polygenic risk score (PRS) excluding *APOE* and candidate risk loci from a recent EUR AD GWAS (Bellenguez et al. 2022). Our results did not reveal either $\epsilon 4$ or AD PRS x combat interactions. However, in the middle age EA group, we observed a combat x SNP interaction with rs11787077 in the gene *CLU*, such that the AD risk allele was associated with an increased impact of combat on PTS (beta=0.75, p=0.00016, p_{FDR}=0.014). An examination of results from our prior study of PTSD and gene expression (Logue et al. 2021) found higher *CLU* expression in the dorsolateral prefrontal cortex of decedents with PTSD (p=0.017) and that rs11787077 was an expression quantitative trait locus. In conclusion, in the largest sample employed to date to address this question, we did not replicate previously reported interactions between *APOE* $\epsilon 4$ (or AD PRS) and combat in predicting PTS among aging veterans. However, we did find that an AD risk locus in the gene *CLU*, also known as *APOJ*, interacted with combat in its association with PTSD severity, which was supported by an examination of *CLU* expression in the PFC. *CLU* variants have been associated with reduced hippocampal volume and default mode network connectivity, both of which are associated with PTSD. This veteran-relevant *CLU* interaction may represent a target for PTSD treatment and a possible genetic link underlying the increased risk for dementia in veterans with PTSD.

Complex Traits Posters - Thursday

PB1363. Exome analysis reveals plausible glaucoma causing mutations in novel genes in dominant JOAG families of Pakistani origin.

Authors:

H. Ayub¹, F. Akhtar², M. Azam³, S. Ali³, H. Khan¹, R. Qamar³; ¹Pak Austria Fachhochschule: Inst. of Applied Sci. and Technology, Haripur, Pakistan, ²Pakistan Inst. of Ophthalmology, Al-Shifa Trust Eye Hosp., Rawalpindi, Pakistan, ³COMSATS Univ. Islamabad, Islamabad, Pakistan

Abstract Body:

Glaucoma is a complex disease with varying mode of inheritance, age of onset & intricate genetic heterogeneity. The current study was aimed to identify the genetic basis of 32 conscripted glaucoma families of Pakistani origin. The families were recruited based on raised intra ocular pressure, visual field damage & gonioscopic findings. Candidate genes exclusion mapping was done by direct sequencing of the probands of each family. Sequence analysis revealed no pathogenic mutation in myocilin and optineurin. Further exome sequencing was done for 2 families with juvenile glaucoma onset & dominant inheritance of the disease that assisted in identification of a novel variant in *PHKG1* (c.125 C>T; p.Thr42Met) to be segregating with the disease in one family. A unique pattern of trigenic inheritance was observed in the second family where three variants segregated with the phenotype (*MYO18A* c.2071G>A; p.Arg691Cys, *COL9A2* c.1061C>T; p.Pro354Leu, *ENOX1* c.171C>T; p.Met57Ile). We postulate the third variant in *COL9A2* to be disease causative variant in combination with other two variants in a family thus has a modifier effect. The study contributes to the current genetic etiology of glaucoma & adds to the understanding of the molecular mechanisms.

Complex Traits Posters - Wednesday

PB1364. Exome sequencing of six singlet Qatar families identified nine neurodevelopmental candidate genes

Authors:

V. Gupta, A. Ben-Mahmoud, A. Kubbar, F. Alshaban, L. W. Stanton, H. G. Kim; Qatar BioMed. Res. Inst., Doha, Qatar

Abstract Body:

Phenotypically and genetically heterogeneous, the genetic etiology of 20-60% of autistic individuals remains unknown. Exome sequencing on six syndromic autism trios from Qatar resulted in the identification of seven novel candidate genes (*SLC25A6*, *MFSD6*, *GRK6*, *KLRC3*, *SCFD2*, *PTGRI*, *LRRCL8*), pathogenicity of which has been substantiated by the CADD scores, frequency in gnomAD, and reported sporadic variants. In three consanguineous and two non-consanguineous families, we identified five candidate homozygous, one *de novo*, and one X-linked recessive variants including one splicing and six missense variants. Additionally, two variants (one novel and one known) were found in known autism genes, *KMT2C* and *ZBTB11*. Common phenotypes found in all probands were autism, developmental disorder, and language/speech delay. One out of six individuals had vision problem, and two out of six individuals had loss of acquired language as additional phenotypes along with autistic features. Extensive data sharing through GeneMatcher and international collaboration will be pivotal to recruit additional deleterious variants, which will result in the disease gene identification and delineation of genotype/phenotype relationship.

Complex Traits Posters - Thursday

PB1365. Exome-wide association study of DSM-5 antisocial personality disorder in a nationally representative sample

Authors:

H. Zhang; NIH/NIAAA, Bethesda, MD

Abstract Body:

Antisocial personality disorder (ASPD) is a prevalent, severe mental health condition with prevalence of 4.3% in the general population. Twin and adoption studies show that genetic factors account for half of the risk for antisocial behavior. Previous association studies typically focus on common variants. To test the effects of rare and low-frequency coding variants on ASPD, we performed an exome-wide association study on an Affymetrix exome and custom SNP array (n=363,496 SNPs), which included 603 ASPD cases and 3,983 super controls who had neither a SUD nor a psychiatric disorder in the European ancestry (EA) and 258 ASPD cases and 1,899 super controls in the African American (AA). All samples were collected from the National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III) and diagnosed by DSM-5 following a direct, structured psychiatric interview. Case-control analysis unraveled 4 SNPs ($P < 5 \times 10^{-5}$) and 5 gene-level loci ($P < 5 \times 10^{-5}$) with rare and low-frequency coding variants in EA. Among these loci, 2 gene-level signals (*ZADH2*, *AKAP5*) were replicated in African Americans within NESARC-III. For SNPs that yielded $P < 10^{-3}$ in the EA discovery cohort, three common variants (*FAM111A*, *RP13-348B13.2*, *LRRC63*) passed a significant threshold ($P < 0.05$) in the AA replication cohort and generated significant outcomes in a meta-analysis ($P < 10^{-3}$). Our analysis indicated both common and uncommon variants within susceptibility genes are involved in the pathogenesis of ASPD. Our results also suggest that gene-based analysis on uncommon variants may be powerful in identifying and prioritizing candidate loci for association studies.

Complex Traits Posters - Wednesday

PB1366. Exploring disease-course trajectories using genomic scores in the Finnish SUPER study - a cohort of 10,407 psychotic individuals.

Authors:

A. Kämpe¹, A. Ahola-Olli¹, O. Pietilainen², L. Urpa¹, T. Singh³, J. Suvisaari⁴, M. Lahteenvuo¹, S-F. Researchers¹, M. Daly¹, A. Palotie¹; ¹Inst. for Molecular Med. Finland, Helsinki University, Helsinki, Finland, ²Dept. of Publ. Hlth., Helsinki Univ., Helsinki, Finland, ³Massachusetts Gen. Hosp., Boston, MA, ⁴Natl. Inst. for Hlth. and Welfare, Helsinki, Finland

Abstract Body:

Psychotic disorders are phenotypically heterogeneous, and specific diagnoses often change during an individual's lifetime. In registry-based research, hierarchical models for lifetime psychotic diagnoses have been proposed to address the interchangeability of diagnoses over time. Here, we provide clinical and genetic support for the hierarchical model previously published within the SUPER study, ranking the four major psychotic disorders based on clinical perception of severity as: 1) Schizophrenia; 2) Schizoaffective disorder; 3) Bipolar disorder and 4) Major depressive disorder with psychotic features.

The Finnish SUPER study comprises 10,407 individuals with a psychotic disorder. The SUPER study's inclusive recruitment of all psychotic disorders and its extensive retrospective follow-up time from Finnish healthcare records (median 48 yrs, IQR 36-49 yrs) provides a unique opportunity to study the interchangeability between, and disease severity within, psychotic disorders. We show that the order of severity, measured as hospitalisation burden over time, well reflect the hierarchical ranking of the four major psychotic groups. High polygenic score for schizophrenia was associated with increased hospitalisation burden ($p=6.51e-05$) and the transitioning from an initial lower ranked psychotic diagnosis to a more severely ranked psychotic diagnosis. The strongest genetic support for the hierarchical model was seen for individuals with bipolar disorder ($n=2438$) where the schizophrenia score was highly associated with the later conversion to schizophrenia or schizoaffective disorder (HR: 1.20, $p=6.29e-05$). Major depressive disorder with psychotic features ($n=1554$) had the widest range of onset age and the shortest survival until conversion to a more severely ranked disorder, suggesting the diagnosis to be more transitory.

Next, we investigated the association of genomic scores commonly used to measure psychotic susceptibility with diagnosis specific disease severity. We show that apart for the schizophrenia score, the educational attainment score had the largest discriminatory power on disease severity. For individuals with schizophrenia the educational attainment score was positively associated to a broad range of functional outcomes such as hospitalisation burden ($p=0.00026$), cognitive measures (PAL adjusted total errors, $p=9.9e-04$), and job-status (OR=1.19, $p=0.0011$). We provide support that the effect of the genomic score for educational attainment on hospitalisation burden is at least partly mediated by its large effect on risk of drug abuse (HR=0.66, $p=1.80e-24$), which is a major risk factor for needing hospital care.

Complex Traits Posters - Thursday

PB1367. Exploring genome-wide association results for neuroticism as a proxy for mental health disorders in individuals of African ancestry

Authors:

M. Kaka, P. Jain, A. Topaloudi, P. Paschou; Purdue Univ., West Lafayette, IN

Abstract Body:

To date, most of the genome-wide association studies (GWAS) are performed in populations of European descent. Given the differences in population genetic structure between Europeans and African populations, it is of urgent need to investigate whether non-European populations can already start benefiting from the results uncovered in large-scale European studies. Neuroticism is a relatively stable personality trait characterized by negative emotionality and is assessed using the 12 items of the neuroticism scale from the Eysenck Personality Questionnaire-Revised Short Form (EPQ-R-S). It has been found to be genetically correlated with some neuropsychiatric disorders such as Schizophrenia, Major Depressive Disorder and Bipolar Disorder. In exploring the transferability studies, GWAS in individuals of African ancestry from UKBB dataset is performed, using quantitative and qualitative phenotypes. Imputation of the variants was performed using an African reference panel and standard quality control measures were used in filtering the dataset. Principal component analysis as well as admixture mapping was also carried out to identify and confirm individuals of African ancestry in the dataset. Afterwards, we performed association testing for each of the 12 Neuroticism items- as qualitative phenotypes, as well as Neuroticism- as a quantitative phenotype. Significant SNPs were identified, and these were then subsequently explored in relevant downstream analyses. In addition, we explored the genetic relationship between Africans and Europeans to confirm the transferability of the GWAS results for Neuroticism, using Popcorn. Hence, as a preliminary study, this study bears the potential to establish Neuroticism as a proxy for mental health disorders in populations of African ancestry as published in Europeans.

Complex Traits Posters - Wednesday

PB1368. Exploring the complexity of scleroderma etiology by trio whole genome sequencing

Authors:

S. Ketkar¹, H. Dai², T. Tan¹, E. Atkinson¹, L. Burrage³, B. Dawson¹, K. Worley³, M. Lyons⁴, S. Assassi⁴, M. Mayes⁴, B. Lee¹; ¹Baylor Coll. of Med., Houston, TX, ²Baylor Coll. of Med. / Baylor Genetics, Houston, TX, ³Baylor Coll. Med., Houston, TX, ⁴Univ. of Texas Hlth.Sci. Ctr., Houston, TX

Abstract Body:

Scleroderma is a heterogeneous rare autoimmune fibrosing disorder affecting connective tissue, which can be grouped into localized scleroderma (LSc) and systemic scleroderma (SSc) depending on if skin is the only affected organ or not. It has been estimated that around 20 cases in every 100,000 individuals worldwide and one in every 4,000 adults in United States. Etiology of scleroderma is still largely unknown, while many genes have been suggested as candidates. Multiple factors, including those environmental, can contribute to the pathological process of the disease, which make it more difficult to identify possible disease-causing genetic alterations. In this study, we have applied whole genome sequencing (WGS) in 101 indexed family trio, supplemented with transcriptome sequencing on six cultured fibroblast cells of patients and five family controls where available. Small nucleotides variations (SNV) and copy number variations (CNV) have been examined, emphasizing on *de novo* calls. We also tested enrichments of rare variants from candidate genes previously proposed in literature by comparing with GnomAD dataset and also conducting simulation test. We identified 42 exonic and 34 ncRNA *de novo* SNV changes in the 101 trios, from total of over 6000 *de novo* calls genome wide. While the clinical significances of those *de novo* calls are still unknown, higher than expected *de novo* calls along with increased expression level of *NEK7* gene in patients could suggest further investigation on possible roles in the pathogenesis of this gene. We also observed significant enrichment of rare variants in candidate genes in the patient cohort, further supporting the complexity/multi-factorial etiology of scleroderma. We didn't find clinically diagnostic genetic alterations in the 101 trios through whole genome sequencing while further investigations should be continued to dissect more on our current findings. For those proposed candidate genes, we have observed enrichment only for rare variants but not any disease-causing *de novo* calls, which suggest more efforts should be warranted to further clarify the candidate genes and possible pathogenetic mechanism.

Complex Traits Posters - Thursday

PB1369. Expression of adipokine genes associated with gestational diabetes mellitus

Authors:

R. Issa; Inst. of Genetic Engineering, Cairo, Egypt

Abstract Body:

Diabetes mellitus is a group of metabolic diseases in which there are higher sugar levels over a prolonged period .Diabetes is due to either the pancreas not producing enough insulin or the cells of the body not responding properly to insulin produced .Aim of this study to assess the expression profile of genes with potential role in the development of insulin resistance .(Adipokines).Adipokines have over 50 different proteins.They are very different from each other in their structure and function which gives them a unique ability to regular process such as lipid metabolism,vascular function,insulin resistance and glucose intolerance..RT -PCR was used for expression analysis of the studied genes which are leptin,adiponectin receptor 1and 2and resistin.Materials and methods:This case control study was performed on 50 pregnant women (35 gestational diabetes GDM patients and 15 control subjects without GDM)All women aged between 26-+5 years,are screened for family history for GDM.Excluded criteria ;patients who have other disorders in their liver and renal function analysis or have high blood pressure Diagnosis;Fasting blood sugar and 2hr after oral glucose 75 gm The results were interpreted according to National Diabetes data group criteria (NDDC)Results Leptin gene expression is higher in patients with GDM ,than in normal pregnant women.Expression of adiponectin genes in pregnant women with GDM decreases in comparison with their matched healthy pregnant group.Adiponectin receptor 1is in the same range in both groups with slight high in GDM compared with controls While adiponectin receptor 2decreases in patients with GDM Expression of Resistin gene is significantly elevated in GDM women compared to matched group

Complex Traits Posters - Wednesday

PB1370. Expression profiling of small bowel tissue from Crohn's disease patients reveals alterations of immune functions and dampening of epithelial response.

Authors:

Y. Lee¹, J. Baek¹, Y. Kim¹, S. Park¹, S. Jung¹, D. Park¹, S. Hwang², J. Lee³, S. Park², S-K. Yang², K. Song¹, Y. Yoon³, B. Ye², H-S. Lee¹; ¹Dept. of Biochemistry and Molecular Biology, Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, ²Dept. of Gastroenterology, Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, ³Div. of Colon and Rectal Surgery, Dept. of Surgery, Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of

Abstract Body:

Crohn's disease (CD) is a chronic inflammatory disease associated with dysregulated immune responses in the gut. Previous studies on inflamed bowel tissues from CD patients showed significant transcriptional changes. We examined the gene expression profiles of uninflamed samples of small bowel to identify alterations in intestinal cells associated with the chronic and progressive course of CD. Macroscopically uninflamed small bowel tissues were obtained from 70 patients with CD and 9 patients with colon cancer during bowel resection, and the latter were used as non-inflammatory bowel disease (IBD) controls. Paired-end RNA sequencing was performed using Illumina HiSeq4000. We performed differential gene expression analysis using DESeq2 and pathway analysis using Gene Set Enrichment Analysis (GSEA). Correlation networks analysis (WGCNA) and cell deconvolution methods (CIBERSORTx) were applied to identify modules and functionally enriched transcriptional signatures of CD. The identified genes were verified using the publically available single-cell RNA sequencing data set (GSE134809). A comparison of CD patients (63% male, median age: 34 years) and non-IBD patients (44% male, median age: 67 years) showed 372 differentially expressed genes (false discovery rate $P < 0.01$, $|\log_2$ fold change > 1.2) comprising 305 protein-coding genes and 67 lncRNAs. Among the 372 genes, 49 protein-coding genes and 5 lncRNAs overlapped with IBD susceptibility loci. Pathways related to immune and inflammatory reactions including TNF- α signaling via NF- κ B ($P = 1.76 \times 10^{-3}$) and IFN- γ response ($P = 1.80 \times 10^{-3}$) were up-regulated in CD, while metabolic pathways including oxidative phosphorylation ($P = 2.24 \times 10^{-3}$) were down-regulated. Compared with non-IBD controls, CD patients had significantly higher proportions of immune cells including plasma cells ($P = 1.26 \times 10^{-10}$) and T-cells ($P = 1.49 \times 10^{-7}$), but a lower proportion of epithelial cells ($P = 1.57 \times 10^{-6}$). We also identified co-upregulated genes (CD module) and co-downregulated genes (anti-CD module) in CD; the hub genes in the CD module were mainly expressed in plasma cells and T-cells, whereas the hub genes in the anti-CD module were enriched in epithelial cells (GSE134809). In conclusion, significant differences in gene expression and cellular composition between CD patients and non-IBD controls implicate molecular alterations associated with immune functions in the uninflamed margins of bowel tissues in CD patients.

Complex Traits Posters - Thursday

PB1371. Extensive Transcriptional Complexity of Autism Associated *SHANK* Family Genes by Capture Based and Long Read sequencing Method

Authors:

X. Lu¹, P. Meade², Y. Ma³, G. Peng⁴, A. J. Forte⁵, H. Zhao⁴, Y. Wang⁶, G. Wang⁷, A. Quiñones-Hinojosa², Y-H. Jiang¹; ¹Dept. of Genetics, Yale Sch. of Med., New Haven, CT, ²Dept. of Neurosurgery, Mayo Clinic, Jacksonville, FL, ³3. Dept. of Neurology, Children's Hosp. of Fudan Univ., Shanghai, China, ⁴Dept. of Biostatistics, Yale Sch. of Publ. Hlth., New Haven, CT, ⁵Div. of Plastic Surgery, Mayo Clinic, Jacksonville, FL, ⁶Dept. of Neurology, Children's Hosp. of Fudan Univ., Shanghai, China, ⁷Yale Ctr. for Genome Analysis, Yale Sch. of Med., New Haven, CT

Abstract Body:

Transcriptional regulations of alternative promoter and splicing contribute to the neural protein diversity and are critical for the development and functions of brain. Transcriptome-wide isoform-level dysregulation has been implicated in neuropsychiatric disorders including autism. However, the transcript structures of most neuronal genes remain to be fully characterized. Assembled RNA-seq could uncover transcriptome regulation but are suboptimal for identifying low abundant and long mRNA isoforms and new exons. *SHANK* family genes (*SHANK1-3*) are strong ASD causing genes and isoform diverse. Isoform specific functions are supported by isoform specific mutant mice. However, the full structures of *SHANK1-3* isoforms remain uncharacterized because of their longer than 10kb-sized mRNAs. We applied single molecule long-reading RNA sequence (standard Iso-Seq), with target enrichment (captured Iso-Seq) using specific designed probe panel, assessed isoform diversity of *SHANK1-3* without transcript assembly error. We applied standard and captured Iso-Seq to normal children and adult human cortex, and striatum and cortex from wild-type, *Shank3*^{Δe4-22}, *Shank3*^{Δe4-9} and *Shank3*^{Δe21} mutant mice. We uncovered 345 and 545 *Shank3* isoforms in mouse cerebral cortex and striatum respectively. Initial analysis has shown that mouse *Shank1* and *Shank2* have ~350 isoforms in cerebral cortex and ~200 isoforms in striatum. In normal human, *SHANK1-3* genes presented ~500 isoforms. In silico analysis revealed that 76% of transcripts have coding capability. The combination of SAM, PRO and SH3 domains of all *SHANK* family proteins exist in cortex, but not in striatum of mouse. The combination of SAM, PRO, ANK and ubiquitin-like domains exist in striatum, but not in cortex. The alliance of SAM, SH3 and ANK domains is only observed in *SHANK1* in cortex but not for *SHANK2* and *SHANK3*. We observed novel exons, extensive fusion transcripts between *SHANK* family and adjacent genes in both human and mouse. Analyses of exonic deletions in *Shank3* mutant mice validated the methods and revealed an unusual pattern of residual isoforms that is specific to targeted mutation of *Shank3*. We also observed extensive new and alternative promoters with tissue specific usage. Further, quantified by single cell RNA-seq, above isoform diversity presented cell type specificity. Our novel technical platform using capture based and long reading sequence has revealed an extreme complexity of transcriptional regulation of *SHANK* family genes. This new knowledge will guide the future functional studies of dissecting the pathophysiology of *SHANK* family causing autism and neuropsychiatric disorders.

Complex Traits Posters - Wednesday

PB1372. Findings at both known and novel loci in genome wide linkage and association study of a depression endophenotype, major depressive disorder with chronic pain.

Authors:

D. Nolan^{1,2}, W. Valentine-Cooper¹, J. Burian¹, V. Vieland³; ¹Nationwide Children's Hosp., Columbus, OH, ²The Ohio State Univ., Columbus, OH, ³The Res. Inst. at Nationwide Children's Hosp., Columbus, OH

Abstract Body:

Although multiple lines of evidence suggest a strong genetic component to major depressive disorder, the susceptibility loci identified thus far account for an extremely small proportion of the genetic variance. We have performed a genome wide linkage and association scan using a biologically relevant endophenotype, major depressive disorder with chronic pain, to identify susceptibility loci within this patient subset. We used a family-based sample from the National Institute of Mental Health Repository and Genomic Resource (NRGR). The samples were originally collected in the GenRed study of early onset recurrent depression and individuals were assessed by DSM IV criteria in addition to completing the diagnostic interview for genetic studies (DIGS), a structured interview of over 2,000 items. Using these data, we were able to define an affection status that incorporated the DSM IV diagnosis of major depressive disorder along with a classification of chronic pain, defined by medical history and medication usage. These data were analyzed under the posterior probability of linkage (PPL) and/or trait-marker linkage disequilibrium (PPLD) framework, as implemented in the software package KELVIN, which measures the evidence of linkage or association on the probability (0,..1) scale. Linkage analysis yielded PPL = 0.35 at chromosome 8p23 and 0.51 at chromosome 16q12. The linked region on 8p23 contains a single gene, *CSMD1*, that has been associated with depression and bipolar disorder as well as quantitative traits such as intelligence in multiple studies. The linkage peak on 16q12 appears to be a novel finding, containing approximately 15 genes, none of which have been associated with similar traits thus far. Our genome wide association results identified some SNP associations within our linked regions. However, our strongest PPL association value was with a SNP on chromosome 5p12 with PPLD= 0.36 in an intronic region of the gene *HCNI*. Multiple studies, including a mouse knockdown model of *HCNI*, have implicated this gene in the therapeutic response to ketamine as well as depression and this gene is being explored as a therapeutic target for depression. Applying the PPL method to this endophenotype of depression, we have identified two genomic regions with well-established roles in the pathogenesis of depression, supporting the utility of the method as well as selection of this endophenotype. Furthermore, we have identified a novel region linked to major depressive disorder with chronic pain. This region warrants further study, as it may contain one or more genes associated with major depression and chronic pain.

Complex Traits Posters - Thursday

PB1373. Fine-mapping 110 migraine risk loci using 98,375 migraine cases

Authors:

H. Hautakangas¹, FinnGen, A. Palotie^{1,2,3}, M. Pirinen^{1,4,5}; ¹Inst. for Molecular Med. Finland FIMM, Helsinki Inst. of Life Sci. (HiLIFE), Helsinki, Finland, ²Analytic and Translational Genetics Unit, Dept. of Med., Dept. of Neurology and Dept. of Psychiatry, Massachusetts Gen. Hosp., Boston, MA, ³The Stanley Ctr. for Psychiatric Res. and Program in Med. and Population Genetics, The Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland, ⁵Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland

Abstract Body:

Migraine is a common and heterogeneous brain disorder with a strong genetic component. Genome-wide association studies (GWAS) have identified over hundred migraine risk loci but causal variants and genes remain largely unresolved. Here, we aimed to narrow down potential causal variants via fine-mapping the autosomal risk loci identified in the largest migraine GWAS to date using FINEMAP software.

Accurate linkage disequilibrium (LD) information is crucial for reliable fine-mapping as mismatches between GWAS population and LD reference can lead to spurious results. Further, common fine-mapping methods assume that same samples have been used for all variants, but this is rarely true in large meta-analyses. An ability to quantify how these issues affect a fine-mapping study is an important, although often neglected, topic.

Our meta-analysis combined GWAS from 23andMe, Inc., UK Biobank and FinnGen for a total sample size of 98,375 migraine cases and 869,160 controls. To avoid methodological problems due to varying sample sizes across variants, we included only variants that were available in all three cohorts. For all loci, we had access to LD from UK Biobank+FinnGen (UKB+FG). Additionally, for 28/110 loci, we had access to complete in-sample LD including 23andMe. Across these 28 loci, we observed that fine-mapping was accurate in 89% of the loci when UKB+FG was the LD reference and in 68% when only UKB was used as LD reference. Here fine-mapping is called accurate if for all variants posterior inclusion probability (PIP) was within 0.10 of its in-sample result. We conclude that UKB+FG reference leads to reliable results in a large majority of loci, clearly improving over the common method of using only UKB.

For a majority of the migraine risk loci, FINEMAP suggested only one (53%) or two (36%) causal variants. Sizes of 95% credible sets ranged from 1 to 5,362 variants, and ten variants had very high PIP (>0.9). We searched the VEP database for the credible set variants that affected protein function. In total, 310 unique missense variants were found of which four had high PIP: rs1133400 (PIP=0.93) in *INPP5A*, rs28929474 (PIP=0.64) in *SERPINA1*, rs6330 (PIP=0.61) in *NGF*, and rs6339 (PIP=0.56) in *NTRK1*. Further, 20 high impact variants (stop gained, splice acceptor, or splice donor) were among the credible sets with modest PIPs ranging from 4.3×10^{-5} to 0.06.

To conclude, prioritizing putative causal variants in migraine risk loci is essential for future translational research on migraine. Here, we presented fine-mapping of 110 migraine risk loci. The high concordance with the results from in-sample LD suggests that our results are reliable in a large majority of the loci.

Complex Traits Posters - Wednesday

PB1374. Fine-mapping 56 lipidomics loci identified by univariate and multivariate genome-wide association analyses

Authors:

L. Ottensmann¹, R. Tabassum¹, S. Ruotsalainen¹, M. Gerl², E. Widen¹, K. Simons^{2,3}, S. Ripatti^{1,4,5}, M. Pirinen^{1,5,6}, ¹Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, ²Lipotype GmbH, Dresden, Germany, ³Max Planck Inst. of Molecular Cell Biology and Genetics, Dresden, Germany, ⁴The Broad Inst. of MIT and Harvard, Cambridge, MA, ⁵Dept. of Publ. Hlth., Clinicum, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, ⁶Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland

Abstract Body:

Aims: The human plasma lipidome captures information beyond traditional lipids and is strongly associated with the risk of cardiovascular diseases (CVD) and beyond. Genome-wide association studies (GWAS) of individual lipid species (univariate analysis) have identified many lipid-associated loci, however the influence of genetic variants on lipid metabolism and CVD risk is still mostly not understood. Multivariate analysis of correlated lipid species improves statistical power and therefore has potential to identify additional lipid-associated loci and identify the causal variants underlying the associations.

Methods: We performed univariate GWAS of 179 lipid species for 7177 participants from the Finnish GeneRISK cohort and multivariate analyses for 11 clusters of correlated lipid species using metaCCA software. The associations were fine-mapped with FINEMAP software to identify the likely causal variant sets and for gene prioritization. We also compared the fine-mapping results to those for the traditional lipids HDL-C, LDL-C, TC and TG in the UKBB.

Results: Univariate and multivariate analysis identified 495 trait-region pairs reaching the Bonferroni-corrected P-value threshold and distributed across 56 loci. Fine-mapping analysis detected 98 independent credible set signals. We identified 11 new loci, where the lead variant was in or near genes *DTL*, *STK39*, *CDS1*, *AGPAT2*, *SGPL1*, *SOAT2*, *KCNJ12*, *YPEL2*, *SPHK2*, *NINL* and *AGPAT3* in multivariate analysis, of which only *AGPAT2*, *SGPL1* and *SOAT2* reached the threshold in univariate analysis. For 26 loci, fine-mapping identified variants with high probability of being causal (>0.9) including 11 deleterious missense variants in 7 known and 2 new loci. Comparison with fine-mapping of traditional lipids in UKBB showed that 5 of the 11 variants did not reach a probability > 0.001 in UKBB results. We prioritized 20 genes, for which fine-mapping analysis identified functional variants with probability > 0.5. These genes are most enriched for the lipid metabolic process gene set of Gene Ontology (adjusted P=3e-17). The prioritized genes were taken to screening in cell lines.

Conclusions: We identified 11 novel loci by multivariate analysis of which only 3 were identified by univariate analysis. Fine-mapping identified 11 deleterious missense variants with high probability of being causal and comparison with results in the UKBB showed that 5 of these variants reached high probability only in our GWAS.

Complex Traits Posters - Thursday

PB1375. Fine-mapping and genomic analyses identify causal variants and genes for hypertension

Authors:

C. Bell¹, S. Van Duijvenboden¹, J. Ramírez¹, A. P. Morris², P. B. Munroe³; ¹Queen Mary Univ. of London, London, United Kingdom, ²Ctr. for Genetics and Genomics Versus Arthritis, The Univ. of Manchester, Manchester, United Kingdom, ³Queen Mary Univ. of London, London, United Kingdom

Abstract Body:

Genome-wide association studies (GWAS) of blood pressure (BP) have successfully identified >1000 loci. However, the genes and biological pathways through which these impact are mostly unknown. We applied a fine-mapping pipeline to identify causal variants at systolic BP (SBP), diastolic BP (DBP) and pulse pressure (PP) loci through integration of published European ancestry GWAS (PMID:30429575), tissue-specific epigenomic annotations (Roadmap/Epilogos), colocalization with transcriptomics (GTEx), and Promoter Capture Hi-C (PMID:31501517). We partitioned BP associations at 606 loci into 1,850 distinct association signals for ≥ 1 BP trait at genome-wide significance ($P < 5 \times 10^{-8}$) using conditional analyses. Of these signals, 532 were associated with at least two BP traits (333 SBP & DBP, 267 SBP & PP, 100 DBP & PP), and 84 with all three. All shared signals were directionally concordant, except for 17 discordant DBP & PP shared signals (*i.e.*, DBP increasing allele was decreasing PP). For 208 (24%), 224 (24.8%) and 159 (22.9%) of SBP, DBP and PP signals, respectively, a single SNV accounted for >75% of the posterior probability of driving the BP association under an annotation-informed prior, and, therefore, was “high-confidence” for causality. We identified 65 high-confidence missense variants, including 20 across two BP traits, and *RGL3* (p.Pro162His) for all three. Three distinct signals reside in the hereditary hemochromatosis *HFE* gene (p.His63Asp, p.Ser65Cys & p.Cys282Tyr) with portal hypertension a recognised phenotype of this condition. Several include genes implicated in kidney-related disorders: *NCOA7*, *LAMB2*, *NPHS2* & *PLXNB2*. Fifteen missense variants had a posterior probability of >99.9%, including at known BP genes, *e.g.*, *SLC39A8*, *ADRB2* & *DBH*, as well as candidates *NR1P1* (regulator of the mineralocorticoid receptor, *MMP14*) and *PLCB3* (intracellular signal transduction enzyme increased in mouse hypertension/trophy models). We integrated our fine-mapped results with *cis*-eQTL in disease relevant tissues from GTEx and observed pairwise colocalization for 96 (SBP), 107 (DBP) and 84 (PP) distinct signals. We also overlapped our signals with the *cis*-regulatory elements identified to target protein-coding gene promoters (capture Hi-C) and for 34 signals the indicated gene was the same as the eQTL analyses, and for 23/34 the same tissue (*RERE*, *ASAP2*, *FGD5*, *MAP1B*, *MXD3*, *RNF130*, *GOPC*, *TNS3*, *IRF5*, *AKR1B1*, *CLN8*, *SLC20A2*, *COL27A1*, *RAD52*, *DPF3*, *PRKCA*, *USP36*, *TGFBI*, *OSER1*, *RGS19*, *MRPS6*, *SLC5A3*, *TRIOBP*). Our analyses provide supporting data for 100s of potential candidate genes at BP loci providing new mechanistic hypotheses for functional validation.

Complex Traits Posters - Wednesday

PB1376. Fine-mapping and signal co-localization of asthma and white blood cell traits within the chr17q12-21 locus across diverse ancestries.

Authors:

C. Ben-Eghan^{1,2}, M. Munter², C. TERAQ³, G. Lathrop^{1,2}, A. Grant^{1,4}; ¹Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, ²McGill Univ. and Genome Quebec Innovation Ctr., Montreal, QC, Canada, ³RIKEN IMS, Tokyo, Japan, ⁴Faculty of Dental Med. and Oral Hlth.Sci., Dept. of Anesthesia, Faculty of Med., Alan Edwards Ctr. for Res. on Pain, McGill Univ., Montreal, QC, Canada

Abstract Body:

Asthma is a heterogeneous disease, with certain forms related to age-of-onset (AO), characterized by the activation and expansion of white blood cell (WBC) subsets, including eosinophils and neutrophils, in the lower airways. Across genome-wide association studies of complex diseases in large study populations, chr17q12-21 stands out as one of the highest peaks, first identified for childhood asthma in a European-descent population. While specific biological mechanisms explaining the top signal have been explored, the role of WBC subset distributions have not been investigated in a comprehensive evaluation across all independent asthma signals within the locus.

We analysed chr17q12-21 (37.2-38.8Mb, hg19) across 5 populations: European (EUR); African (AFR); East-Asian (EAS); South-Asian (SAS) from the UK Biobank and East-Asian from the Biobank Japan (BBJ). We ran 2 association scans for asthma: (1) A population-based & meta-analysis scan.(2) AO-stratified analysis in EUR: onset <18years (13,419); 18-40 years (12,526); <40years (25,945) and >40years (18,052) using 361,244 controls. We also ran association testing for basophil (BASO); eosinophil (EOS); lymphocyte (LYM); monocyte (MON); neutrophil (NEU) and total WBC counts (WBC#). We identified independent signals using GCTA-COJO and tested for co-localization in all asthma & WBC trait combinations via COLOC. Finally, we ran statistical fine-mapping (SuSiE & FINEMAP) and screened for evidence of mediation through WBC traits (Regmedint).

A union of 8 independent signals (*S1-S8*)($R^2 \geq 0.5$) were identified through conditional association analysis across AO-related strata among EUR. The top signal (*S4*) was also identified in a meta-analysis of non-EUR study populations. Three signals including *S4* were found to co-localize with a EUR asthma phenotype and EOS, LYM, MON, NEU, or WBC# (posterior probability ≥ 0.6). For 2 of these signals, *S4* & *S8*, multi-causal variant co-localization with LYM was identified. Moreover, signal *S4* was found to co-localize with LYM for asthma phenotype AO-<18years and among SAS and BBJ. Evidence of mediation through EOS with AO-<40years was found for 30% of the total effect of *S8* on asthma.

Our results of co-localization and mediation with evidence from diverse ethnicities demonstrate that despite the distance of WBC trait measures from the site (lung) or from the time of an asthma exacerbation, resting WBC levels may be causally related to certain asthma association signals. A causal pathway with a specific impact on a specific WBC trait as a potential mediator is likely to point to a relevant candidate molecular pathway for pharmacological intervention.

Complex Traits Posters - Thursday

PB1377*. Fine-mapping the association of *CYP2A6* with nicotine metabolism in African ancestry smokers.

Authors:

J. Pouget¹, A. El-Boraie¹, A. Langlois¹, C. Lerman², J. Knight³, L. S. Cox⁴, N. Nollen⁴, J. Ahluwalia⁵, M. Chenoweth⁶, R. Tyndale¹; ¹CAMH and Univ. of Toronto, Toronto, ON, Canada, ²Univ Pennsylvania Hlth Sys, Philadelphia, PA, ³Univ. of Toronto and Lancaster Univ., Bailrigg, United Kingdom, ⁴Univ. of Kansas, Lawrence, KS, ⁵Brown Univ., Providence, RI, ⁶Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Background: Tobacco use is the leading preventable cause of death and disease in North America. Nicotine is the primary addictive agent in tobacco, and individual differences in metabolism of nicotine by the liver enzyme *CYP2A6* alter smoking behaviours as well as cessation and health outcomes. The nicotine metabolite ratio (NMR; 3'hydroxycotinine/cotinine) is a stable biomarker for nicotine clearance with demonstrated clinical utility in personalizing cessation treatment. We previously found that genetic variation in the *CYP2A6* region on chromosome 19 was strongly associated with the NMR in African ancestry smokers. In this study, we investigated this regional association in more detail.

Methods: Our sample comprised African ancestry smokers from two clinical trials, Pharmacogenetics of Nicotine Addiction Treatment (PNAT)-2 (n=504) and Kick-it-at-Swope (KIS)-3 (n=450). The NMR was measured at intake, with levels of cotinine and 3'hydroxycotinine determined from blood samples using mass spectrometry. Genome-wide genotyping was conducted using the Illumina OmniExpressExome array, with a custom add-on for 2,688 variants previously associated with nicotine metabolism/smoking for richer coverage of regions of interest (including *CYP2A6*). Known *CYP2A6* structural variants were directly genotyped. Chromosome 19 variants were imputed using the TOPMed Imputation Server, with the cosmopolitan Version R2 reference panel. Association testing was completed using SNPTEST. Stepwise conditional analyses were completed to identify independent associations in regions reaching genome-wide significance. Bayesian fine-mapping was undertaken using FINEMAP to identify putatively causal variants.

Results: Our analyses confirmed genome-wide association of the *CYP2A6* region of chromosome 19 with the NMR among African ancestry smokers. Stepwise conditional analyses identified four independent associations in the *CYP2A6* region (5.5kb upstream of *CYP2A6*, 16kb and 9kb downstream of *CYP2A6*, and further downstream of *CYP2A6* within a long non-coding RNA transcript). Bayesian fine-mapping identified four putatively causal variants, including a SNP in linkage disequilibrium with *CYP2A6*17* (a known functional allele conferring decreased *CYP2A6* activity).

Conclusions: Our study highlights multiple independent genetic associations with the NMR in the *CYP2A6* region in African ancestry smokers. These variants may causally influence nicotine clearance and thus smoking behaviours, and tobacco-related disease risk. Future work includes replication and evaluation of the potential functional impact of robustly associated variants.

Complex Traits Posters - Wednesday

PB1378. Fine-mapping to find differences in effect sizes across ancestries for skeletal phenotypes

Authors:

E. Bartell^{1,2,3}, K. M. Tsuo^{1,2,4}, K. Lin⁵, W. Gan⁵, J. B. Cole^{6,1,2,3,7}, S. Vedantam^{3,2}, R. M. Salem⁸, R. G. Walters^{5,9}, J. N. Hirschhorn^{1,2,3}; ¹Dept. of Genetics, Harvard Med. Sch., Boston, MA, ²Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, ³Div. of Endocrinology and Ctr. for Basic and Translational Obesity Res., Boston Children's Hosp., Boston, MA, ⁴Massachusetts Gen. Hosp. Res. Inst., Boston, MA, ⁵Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, ⁶Program in Metabolism, Broad Inst. of MIT and Harvard, Cambridge, MA, ⁷Diabetes Unit and Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁸UC San Diego, La Jolla, CA, ⁹MRC Population Hlth.Res. Unit, Univ. of Oxford, Oxford, United Kingdom

Abstract Body:

Background: Humans share the vast majority of their DNA across ancestral groups; shared heritability suggests genetic effects are mostly shared. However, identifying genetic signals with true differences in effect sizes across ancestries can prove difficult, as differences in linkage disequilibrium and allele frequencies can create differences in association despite no differences in variant effects. Genome-wide association studies (GWAS) and variant level fine-mapping give us tools to identify loci with potential differences in genetic effects between ancestries even in the presence of differing LD.

Goal: Use statistical fine-mapping to detect loci with variants having different genetic effects across ancestries.

Design: We used data from the UK Biobank (UKB) and China Kadoorie Biobank (CKB) to study height and sitting height ratio (SHR) in ~475k individuals from 2 ancestries. We performed GWAS using BOLT-LMM in ~400k European and ~76k East Asian individuals, and performed fine-mapping in UKB using SuSiE. For all SNPs identified in credible sets (CSs) by SuSiE, we evaluated effect size heterogeneity between ancestries using Cochran's Q. To identify heterogeneous association signals, we defined CSs as significantly heterogeneous if all effect alleles in the CS had significantly heterogeneous effect sizes across ancestries. SNPs were checked for allele flips as an artifactual cause of heterogeneity. For heterogeneous CSs, we then tested for enrichment of functional annotations overlapping each CS's highest PIP SNP.

Results: We identify 40 height and 28 SHR associated CSs that are significantly heterogeneous between UKB and CKB (2.8% of height and 2.5% of SHR CSs). All CSs identified had SNPs with opposite directions of effect, likely because this scenario yields greatest power given these stringent criteria. Similarly, these SNPs had smaller minor allele frequencies differences across ancestries than SNPs in non-heterogeneous CSs, and the median number of SNPs in heterogeneous CSs were lower (1 vs 6 for height and 10 for SHR in all CSs), likely also reflecting greater power for detection. Heterogeneous CSs for height and SHR never overlapped. Height heterogeneous CSs were most enriched for H3K9ac peaks (t-test $p=0.0048$ vs CS SNPs), and those for SHR were most enriched for H3K27ac ($p=0.0076$).

Conclusions: Using GWAS of height and SHR and statistical fine-mapping across 2 major ancestries, we identify 68 CSs that significantly differ in effect size between ancestries. Our stringent criteria identify a small subset of signals; these loci with evidence of differential effect sizes across ancestries are enriched for multiple functional annotations.

Complex Traits Posters - Thursday

PB1379*. Functional analysis of rare copy number variations across psychiatric disorders

Authors:

W. Engchuan¹, O. Shanta², M. Klein², J. MacDonald¹, B. Thiruvahindrapuram¹, A. Maihofer², G. HUGUET³, K. Kimberley⁴, I. Sønderby⁵, Z. Wang¹, M. Zarrei¹, G. Pellecchia¹, D. Merico⁶, S. Jacquemont³, S. Scherer¹, J. Sebat², PGC BD, SCZ, PTSD, ADHD, MDD, ASD and CNV Workgroups; ¹The Hosp. for Sick Children, Toronto, ON, Canada, ²Univ. of California, San Diego, CA, ³Univ. of Montreal, Montreal, QC, Canada, ⁴Cardiff Univ., Cardiff, United Kingdom, ⁵Oslo Univ. Hosp., Oslo, Norway, ⁶Deep Genomics, Inc., Toronto, ON, Canada

Abstract Body:

Psychiatric disorders have been long known to be highly heritable, however, their genetic etiology is still under-explained due to their complexity and heterogeneity. Besides the known implicated rare copy number variations (CNVs) loci, many studies have attempted to identify additional associations. However, with the limited sample size, many of the findings failed to replicate. In this study, we combined and harmonized multiple microarray datasets, resulting in one of the largest psychiatric disorders data with a total of 542,274 samples of European ancestry (126,936 cases diagnosed with one or more psychiatric conditions and 415,338 controls). The major psychiatric disorders presented in the dataset are major depressive disorder (MDD, n~38,000), schizophrenia (SCZ, n~31,000), bipolar disorder (BD, n~24,000), autism spectrum disorder (ASD, n~15,000), and posttraumatic stress disorder (PTSD, n~13,000). Using a centralized pipeline, we called rare CNVs (case-control combined frequency <1%, size >10kb and >10 underlying probes) and performed multiple levels of CNV association analysis in order to understand the genetic factors that are common and different between major psychiatric disorders, as well as to additionally identify novel disorder-specific genetic factors. A case-control analysis of unrelated samples was done on each individual disorder of MDD, SCZ, BD, PTSD, as well as the cross-disorder (excluding ASD families), while a family-based analysis was performed for ASD. Meta-analysis was performed to combine the test statistics of the cross-disorder analysis and the ASD analysis as a final cross-disorder test statistics. A gene-set and pathway analysis showed an enrichment of deletions impacting synaptic genes, and genes related to nervous system development and neurotrophin signalling pathway to be common in cases of different psychiatric disorders, especially in SCZ and ASD. We also found an enrichment of deletions in genes related to forebrain development in MDD, while duplications impacting ribosomal RNA processing genes were found to be enriched in BD. Analyzing single-cell RNA-seq data of the human neocortex showed that top associated genes in SCZ, ASD and MDD were expressed significantly higher in excitatory neurons. Finally, genome-wide gene-based analysis identified strong signal in previously known SCZ and ASD loci, e.g. 22q11.2 deletion, 16p11.2 duplication, 15q11.2 duplication, 15q13.3 deletion, 3q29 deletion, and deletion of NRXN1 as common risk of psychiatric disorders. Future plans for the analysis on samples of non-European ancestry and cross-ancestry meta-analysis will be also discussed.

Complex Traits Posters - Wednesday

PB1380. Functional validation of rs7132908 as the causal variant at the childhood obesity locus on chr12q13.

Authors:

S. Littleton^{1,2}, J. Maguire², J. Bradfield^{3,2}, J. A. Pippin², C. Su², A. Chesi¹, A. D. Wells^{2,1}, R. I. Berkowitz^{1,2}, M. C. Pahl², S. F. A. Grant^{2,1}; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Children's Hosp. of Philadelphia, Philadelphia, PA, ³Quantinuum Res. LLC, San Diego, CA

Abstract Body:

The *FAIM2* locus is reproducibly associated with childhood obesity by GWAS but is understudied given the association is less pronounced in adults. This locus is tagged by rs7138803 but based on our previously published variant-to-gene mapping efforts plus related eQTL observations, we hypothesize that rs7132908 is the causal variant at this locus. We also hypothesize that rs7132908 resides within an enhancer that regulates *FAIM2* expression and that the rs7132908 risk allele decreases *FAIM2* expression. Enrichment analyses of childhood obesity GWAS signals nominate central nervous system tissue and neuronal cell types. The arcuate nucleus of the hypothalamus brain region is responsible for regulating appetite. We were therefore motivated to study the effect of genotype at rs7132908 on the function of hypothalamic arcuate nucleus neurons. We generated isogenic human embryonic stem cell (hESC) lines homozygous for either the rs7132908 non-risk or risk allele which were validated with karyotyping, *de novo* CNV analysis and Sanger sequencing at the 10 most likely off-target sites. Each hESC line was then differentiated to a heterogeneous *in vitro* model of the hypothalamic arcuate nucleus. We visually observed a smaller proportion of cells demonstrating neuron morphology with the rs7132908 risk genotype. This observation was confirmed at the transcript level using snRNA-seq and at the protein level with immunofluorescence. Using paired snRNA-seq and snATAC-seq, we observed that chromatin accessibility at rs7132908 is significantly correlated with expression of 21 genes within 1Mb, including *FAIM2* (correlation 0.13, adjusted $P=2.2 \times 10^{-6}$). *FAIM2* had decreased expression in the differentiated cells with the rs7132908 risk genotype (fold change 0.51, $P=1.3 \times 10^{-5}$). We also observed decreased chromatin accessibility at rs7132908 (fold change 0.74, adjusted $P=3.4 \times 10^{-9}$) and the *FAIM2* promoter (fold change 0.73, adjusted $P=1.4 \times 10^{-10}$) in the differentiated cells with the rs7132908 risk genotype. We have separately validated the allele-specific enhancer activity of rs7132908 for *FAIM2* with a luciferase assay in human primary astrocytes. The rs7132908 non-risk allele region increased luciferase reporter expression relative to the *FAIM2* promoter alone by a fold change of 1.87, while the risk allele decreased expression by a fold change of 0.68 (non-risk allele+*FAIM2* promoter vs. risk allele+*FAIM2* promoter adjusted $P<0.0001$). These results warrant further replication efforts and investigation into the mechanism underlying the effects of the rs7132908 risk allele on gene expression and neuron differentiation.

Complex Traits Posters - Wednesday

PB1381. Gene-based association studies of rare variants identify novel genes for asthma-associated quantitative traits in African-American children.

Authors:

S. Clay¹, M. Dapas¹, C. Visness², R. Wood³, G. T. O'Connor⁴, R. Cohen⁴, G. K. Hershey⁵, C. Kercsmar⁶, R. S. Gruchalla⁷, M. Gill⁷, A. H. Liu⁸, H. Kim⁹, M. Kattan¹⁰, L. Bacharier¹¹, D. Rastogi¹², K. Rivera-Spoliaric¹³, R. Robison¹⁴, P. Gergen¹⁵, W. Busse¹⁶, J. Gern¹⁶, D. Jackson¹⁶, C. Ober¹; ¹Univ. of Chicago, Chicago, IL, ²Rho Inc, Chapel Hill, NC, ³Johns Hopkins Univ. Med. Ctr., Baltimore, MD, ⁴Boston Univ. Sch. of Med., Boston, MA, ⁵Univ. of Cincinnati Coll. of Med., Cincinnati, OH, ⁶Cincinnati Children's Hosp., Cincinnati, OH, ⁷Univ. of Texas Southwestern Med. Ctr., Dallas, TX, ⁸Children's Hosp. Colorado, Aurora, CO, ⁹Henry Ford Hlth.System, Detroit, MI, ¹⁰Columbia Univ. Coll. of Physicians and Surgeons, New York, NY, ¹¹Monroe Carell Jr. Children's Hosp. at Vanderbilt Univ. Med. Ctr., Nashville, TN, ¹²Children's Natl. Hlth.System, Washington, DC, ¹³Washington Univ., St. Louis, MO, ¹⁴Ann and Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, ¹⁵Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, ¹⁶Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI

Abstract Body:

Genome-wide association studies (GWAS) have successfully identified thousands of associations between common variants and hundreds of complex traits. However, rare variants, which comprise the bulk of genetic variation, have been largely ignored because they are not well represented by the variants on commonly used genotyping arrays. Moreover, GWAS have focused primarily on European-ancestry cohorts. Thus, little is known about risk variants in other populations. In this study, we explored the contributions of rare variants (MAF <0.01) to asthma-associated quantitative traits using whole-genome sequences in two primarily African-American (67%; 25% Hispanic) cohorts: the Asthma Phenotypes in the Inner City (APIC) study (n=508 children with asthma) and the URban Environment and Childhood Asthma (URECA) prospective birth cohort (n=173 children with asthma and 225 children without asthma). We performed gene-based association tests with STAAR (variant-set test for association using annotation information) and examined associations with traits reflecting the allergic (total IgE, sensitization to specific allergens), pulmonary (FEV1 % predicted, FEV1/FVC, bronchodilator response), and inflammatory/immune (FeNO, blood eosinophil count, blood neutrophil count) components of asthma. One gene (*USF1*; $p=2.18 \times 10^{-7}$) was associated with blood neutrophil count after multiple test correction. Two other gene-trait associations were notable: *TNFRSF21* with total IgE ($p=6.47 \times 10^{-6}$) and *PIK3R6* and blood eosinophil count ($p=4.10 \times 10^{-5}$). These three associations were supported by studies in phenomeXcan or mouse knockouts (KO): predicted expression of *USF1* and *PIK3R6* were associated with neutrophil ($p=4.66 \times 10^{-4}$) and eosinophil ($p=3.96 \times 10^{-6}$) percentages, respectively, in phenomeXcan. *Tnfrsf21* KO mice had increased IgE levels and *Usf1* KO mice had increased blood neutrophil count. These studies identified three novel gene associations with atopic and inflammatory phenotypes, highlighting the potential importance of rare variants in the development of these traits and of including non-European ancestry populations in genetic studies.

Complex Traits Posters - Thursday

PB1382. Gene-based tests for early and late onset Alzheimer Disease: common and non-overlapping factors

Authors:

E. Lucio da Fonseca¹, M. N. Jean-Francois¹, J. T. Kurup², C. Yang³, S. H. Slifer¹, E. R. Martin¹, B. W. Kunkle¹, A. Naj⁴, L. A. Farrer⁵, J. L. Haines⁶, R. P. Mayeux⁷, G. D. Schellenberg⁸, M. A. Pericak-Vance¹, M. V. Fernandez³, C. Cruchaga³, C. Reitz², G. W. Beecham¹, A. D. Genetics Consortium (ADGC)⁹; ¹Univ. of Miami, Miller Sch. of Med., John P. Hussman Inst. for Human Genomics, Miami, FL, ²Columbia Univ., The Taub Inst. for Res. on Alzheimer's Disease and the Aging Brain, The Gertrude H. Sergievsky Ctr. and Dept.s of Neurology and Epidemiology, New York, NY, ³Washington Univ. Sch. of Med., St. Louis, MO, ⁴Univ. of Pennsylvania, Perelman Sch. of Med., Dept. of Biostatistics and Epidemiology/Ctr. for Clinical Epidemiology and Biostatistics, Philadelphia, PA, ⁵Boston Univ. Sch. of Med., Dept. of Med., BioMed. Genetics, Boston, MA, ⁶Case Western Reserve Univ. Sch. of Med., Dept. of Population & Quantitative Hlth.Sci., Cleveland Inst. for Computational Biology, Cleveland, OH, ⁷Columbia Univ., Dept.s of Neurology, Psychiatry, and Epidemiology, Gertrude H. Sergievsky Ctr., Taub Inst. for Res. on the Aging Brain, Coll. of Physicians and Surgeons, New York, NY, ⁸Univ. of Pennsylvania Perelman Sch. of Med., Path & Lab Med, Stellar Chance, Philadelphia, PA, ⁹Univ. of Pennsylvania, Perelman Sch. of Med., Philadelphia, PA

Abstract Body:

Background: Early Onset Alzheimer Disease (EOAD, age at onset [AAO] ≤ 65) is a severe form of AD. Most genetic studies have focused on autosomal dominant forms of EOAD, and there is little understanding of similarities and differences between early and late onset forms of AD (LOAD, [AAO] >65). We present here a genome-wide association study of non-Mendelian EOAD and compare it to LOAD GWAS.

Methods: Single-variant analyses were performed using additive logistic regression for case-control models (full: SNPs, PCs, sex and APOE- $\epsilon 4$ dosage as covariates; reduced: SNPs and PCs); secondary models included PCs+sex and PCs+APOE. Models were applied on data of unrelated individuals derived from the Alzheimer Disease Genetics Consortium (ADGC). Samples varied from 1293-1459 cases and 8894-9366 controls for EOAD, and 8795-9508 cases and 9702-10273 controls for LOAD, depending on the considered model. The SNP heritability (h^2) and genetic correlation (r_g) were estimated using LD score regression. Gene-based and pathway analysis (gene sets from Msigdb-v7.0) were performed using FUMA/MAGMA-v1.6.

Results: We identified two novel loci associated with EOAD: Chromosome 4 (full model: chr4:102027610, $P=2.98 \times 10^{-08}$, near *PPP3CA*) and Chromosome 12 (PCs+sex model: rs117001070, $P=2.11 \times 10^{-08}$, between *LINC02444* and *LINC02882*). Heritability and genetic correlation showed moderate but incomplete genetic correlation between EOAD and LOAD. Other than *APOE* ($p=3.89 \times 10^{-12}$), gene-based tests showed nominal association for full model EOAD: *C8orf44-SGK3* ($p=9.95 \times 10^{-06}$), *SGK3* ($p=2.01 \times 10^{-05}$) and *HIST1H2AC* ($p=3.04 \times 10^{-05}$). For full model LOAD, significant genes were *APOE* ($p=7.11 \times 10^{-15}$), *MS4A6A* ($p=2.27 \times 10^{-06}$), *MS4A2* ($p=2.31 \times 10^{-06}$), *APOC1* ($p=2.53 \times 10^{-06}$), *MS4A4E* ($p=2.77 \times 10^{-06}$), *TOMM40* ($p=3.22 \times 10^{-06}$) and *NCF4* ($p=1.80 \times 10^{-05}$). Plausible pathways such as HDL remodeling ($p=5.50 \times 10^{-05}$) and positive regulation of cholesterol efflux ($p=9.79 \times 10^{-05}$) were observed for full model EOAD and were nominally associated with full model LOAD ($p=0.04$ and $p=4.05 \times 10^{-03}$).

Conclusions: Gene-based tests confirmed that *APOE* is a common factor for EOAD and LOAD. Distinct nominally significant genes were detected for EOAD and LOAD, suggesting that the genetic etiology of EOAD has an incomplete genetic overlap with LOAD.

Complex Traits Posters - Wednesday

PB1383. Gene-by-stress interaction in Alzheimer's disease risks.

Authors:

A. Singh, K. Arbeev, A. Yashkin, I. Akushevich, A. Yashin; Duke Univ., Durham, NC

Abstract Body:

The WHO estimates that currently more than 55 million people live with Alzheimer's disease related-dementia (ADRD) worldwide and that with nearly 10 million new cases every year, this number is estimated to increase to 75 million by 2030 and 132 million by 2050, resulting in a significant rise in ADRD-related deaths. Psychosocial stress is considered a risk factor that may contribute in the development of age-related neuro- and psychopathologies. Therefore, we hypothesized that a part of ADRD-risk could be due to complex interplay of genes and psychosocial stress. We tested this hypothesis using the Health and Retirement Study (HRS) surveys and Medicare-linked data. We created a synthetic measure for psychosocial stress using the algorithm based on stress proxy-indicators, as described in Singh et al., 2015. Guided by correlation with stress, we selected age-at-onset of ADRD as an outcome variable and evaluated race-stratified gene-by-stress interaction (i.e., GxE) GWAS using linear regression that included SNP, SEX, STRESS, and ancestry principle components (PCs) as other predictors in the model. We observed GxE associations at p-value of order 10^{-8} in both Black and White samples. Out of the top 25 most significant GWAS SNPs in both races, almost all SNPs were mapped to non-coding regions, indicating that GxE interactions may mostly influence ADRD through gene-regulatory roles. The genic-region SNPs were mapped to four characterized genes (*BCR*, *MCC*, *KCNQ5*, and *DCP2*) in Whites and three genes (*EMCN*, *DPYD*, and *INSYN2A*) in Blacks, which included known associations with neurodevelopmental disorders including autism, mental retardation, intellectual disability, and also for cancers including colorectal carcinoma, gastric carcinoma, chronic myeloid leukemia, and lymphoma. Our findings suggest that a part of ADRD may result from a complex interplay of genes and daily-life stress and that the interplay may possibly influence through shared genes a mechanistic feature shared by both cancer tumors and Alzheimer's disease, i.e., growth of new blood vessels in the brain.

Complex Traits Posters - Thursday

PB1384. Gene-environment interaction of coffee with body mass index in multiple populations

Authors:

M. Shivakumar¹, D. Kim¹, E. Choe²; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract Body:

Obesity is a global pandemic disease, and its prevalence has tripled in the last four decades. A flurry of genome-wide association studies (GWAS) has found many variants associated with body mass index (BMI) and subsequently, polygenic risk scores (PRS-BMI) derived from them have shown the genetic basis and risk of obesity. Additionally, interactions between genetic factors and environmental risk factors like lifestyle and diet, such as coffee consumption, also play an important role in obesity. In this study, we sought to investigate interactions between PRS-BMI and coffee consumption in the European population (UK Biobank, N = 218,777) and the Korean population (GENIE, N = 5,501).

GAINT consortium and Biobank Japan GWAS summary statistics were used to calculate BMI PRS for UKBB and GENIE populations, respectively. Subjects who consumed less than or equal to 1 cup of coffee per day were considered as control and greater than 1 cup of consumption as cases. Log-rank test was used to check the significance of the interaction between PRS-BMI and coffee consumption, adjusting for age, sex, IPAQ physical activity score, and principal components. Further, the subjects were stratified into BMI categories - BMI \leq 18.5 (underweight), $18.5 < \text{BMI} \leq 25$ (normal) and BMI > 25 (overweight & obese) the PRS BMI-coffee interaction was tested in the subgroups.

The interaction between coffee and PRS BMI was significant (UKBB P = 2.65×10^{-7} and GENIE P = 0.03) in both populations. The interaction was significant even after adjusting for coffee metabolism score. In the subgroup analysis, the interaction was significant in BMI > 25 (P = 0.003) and BMI > 30 (P = 0.027) groups in UKBB.

The findings in this study suggest that there is a significant interaction between BMI-PRS and coffee in different populations. In the case of the European population, there is a significant interaction in overweight and obese subjects. The results from this study may be applicable for personalized medicine and preventive treatment through risk assessments and interventions.

Complex Traits Posters - Wednesday

PB1385. Gene-level GWAS fine-mapping combined with single cell RNA-seq yields insight into neurodegenerative disease biology

Authors:

B. van de Geijn, J. Blischak, M. McCarthy, T. Bhangale; Genentech, South San Francisco, CA

Abstract Body:

Identifying causal genes from GWAS remains challenging since associations are established at the variant and not gene level. Moreover, most causal variants lie in non-coding regions that don't immediately connect to single genes. Methods such as MAGMA condense GWAS associations onto the gene level using windows around genes, but fail to nominate single causal genes due to gene density and linkage disequilibrium. We combined fine-mapping with various recently developed variant-to-gene (V2G) linking strategies to better pinpoint causal genes across the entire genome. We then used our gene-level scores to interrogate single cell expression patterns in Alzheimer's disease cases and controls (Morabito et al 2021).

To create gene-level GWAS scores for a trait, we first QCed summary statistics to avoid spurious causal signals due to reference LD mismatches and performed functional fine-mapping (Weissbrod et al 2020). We used the causal effect sizes to estimate the heritability contribution for each variant, then calculated gene level heritability scores by linking variants to genes using two V2G strategies (cS2G (Gazal et al 2022) and expression modifier scores (Wang et al 2021)) weighted by connection strength. We compared our approach to popular MAGMA by analyzing 44 traits - selected because they have matched silver-standard associated genesets (Zhang et al. medRxiv). We found that the fine-mapped gene scoring with cS2G gene connections produced stronger associations than MAGMA ($p=0.02$).

We then used our gene scores to investigate causal cell-types for disease. We scored cells based on their expression (single nucleus RNA-seq) profiles and gene scores using data from Morabito et al., a study of 191,890 nuclei from postmortem brain tissue that includes both individuals with Alzheimer's disease (AD) and healthy controls. We did this by taking the inner product of the vector of standardized expression for each cell and the vector of gene scores for AD. We examined score differences across cell-types and sub-cell-type clusters. We found that both microglia and astrocytes showed enrichment for cell scores above other frontal cortex cell types (4.8x and 3.9x enrichment respectively; both $p<1e-15$) as well as one subtype of oligodendrocytes (2.14x enrichment; $p<1e-15$). The role of microglia in AD has been described in the past, but our methods highlight additional cell-populations which could point to distinct mechanisms. We also found differences in scoring distributions between cases and controls in sub-cell-type clusters of astrocytes and microglia. We believe fine-mapping gene and cell scoring will aid the understanding of causal cell-types for many diseases.

Complex Traits Posters - Thursday

PB1386. Genes at asthma-associated GWAS loci are induced in activated CD4 tissue resident memory T cell

Authors:

N. Schoettler; Univ. of Chicago, Chicago, IL

Abstract Body:

Asthma is a complex genetic disease with >200 independent loci identified in genome-wide association studies (GWAS). However, the causal variants and causal genes have not been characterized for the majority of loci, possibly as a result of the paucity of relevant cell models of gene expression. Lung CD4 tissue resident memory T (T_{RM}) cells are the most abundant T cell subset in the lung, where they can generate potent memory responses; these cells are rare in other tissues, including blood. I sought to determine whether human lung CD4 T_{RM} cell stimulation modulates the expression of genes at asthma GWAS loci. Human lung leukocytes from 10 donors were cultured for 20 hours with anti-CD3/anti-CD28 or media followed by FACS sorting of CD4 T_{RM} cell and CD4 effector memory T (T_{EM}) cell subsets and RNA sequencing. For two donors, I also conducted single-cell RNA sequencing (scRNAseq) on sorted CD4 T_{RM} cells that were expanded for 14 days as well as unexpanded CD4 T_{RM} cells treated for 20 hours with anti-CD3/anti-CD28. A total of 329 genes were differentially expressed between CD4 T_{RM} cells and CD4 T_{EM} cells after anti-CD3/CD28 treatment (FDR < 5%), with 77 genes having higher expression and 252 genes having lower expression in CD4 T_{RM} cells compared with T_{EM} cells. Of these, 14 genes localized to an asthma GWAS locus. As examples of GWAS response genes, *IL5*, *IL13*, *IL4* (5q31.1), *IL2* (4q27), and *IL4R* (16p12.1) were more highly expressed in activated CD4 T_{RM} cells. The scRNAseq revealed patterns of co-expression of *IL5* and *IL13* in T_{EM} after activation ($R^2 = 0.37$), but not in CD4 T_{RM} cells. While co-expression of the other cytokine genes (*IL4* and *IL2*) was only rarely observed in resting CD4 T_{RM} cells, increased co-expression of *IL4/IL5* and *IL4/IL13* was observed more frequently in activated CD4 T_{RM} cells. These results associate asthma-risk loci with asthma-relevant genes during lung CD4 T_{RM} cell activation suggesting that studies linking GWAS variants to gene function in blood or resting cells are likely missing important features of the gene regulatory landscape in asthma.

Complex Traits Posters - Wednesday

PB1387. Genes causing significant effects on the dynamics of peroxisomes

Authors:

N. Roy, S. Jangam, J. Andrews, M. Wangler; Baylor Coll. of Med., Houston, TX

Abstract Body:

Peroxisomes are single membrane-bounded eukaryotic ubiquitous subcellular organelles, which vary in size, shape and number depending on the cellular environment. Peroxisomes mediate crucial biological functions including beta-oxidation of fatty acids, biosynthesis of plasmalogen lipids, synthesis of bile acids for fat digestion and detoxification of alcohol. Reactive oxidative species (ROS) produced by peroxisomes is one of the major contributors to several multisystem complex diseases of the brain, liver, bone and kidney. Besides environment, several genes are known to play vital role in the dynamics of peroxisomes, causing its enhanced division and fission. Prolonged and continuous consumption of alcohol (>80gm of alcohol for >10 years) stimulates the catalase pathway of alcohol metabolism, mediated by peroxisomes, causing increased production of ROS, resulting in alcoholic chronic liver disease (ALD). Hepatic encephalopathy or loss of brain function is a typical feature of decompensated ALD patients. A list of significantly associated genes (n=150) (OR>1; p-value=<1E-05) from a GWA study was used as the foundation for our study. The list of genes was subjected to DRSC Integrative Ortholog Prediction Tool (DIOPT) Scoring for shortlisting based on their score. RNAi lines of the respective fly (*Drosophila Melanogaster*) genes (orthologs) were obtained from Bloomington Drosophila Stock Center (BDSC) and crossed with ActinGAL4-UASGFPSKL fly lines. UAS-LacZ-RNAi line was used as control. Evaluation of effect of knocked down genes on the dynamics of peroxisomes were done by confocal microscopy imaging (Leica STED) and analysis of size and number of peroxisomes were done by ImageJ Software. Adult flies of each cross were tallied to observe the extent of lethality of knock-down genes. Out of our original list of 69 genes with high DIOPT Score, 52 genes were shown to have a direct neuro-pathological effect. Out of 10 genes analyzed so far, knock-down of two genes, MTPAP and Rabex-5, showed significant effects on size and number of peroxisomes in the cells of salivary glands and fat body as compared to control. The average area of a peroxisome was higher (1.488 and 1.887 micron² for MTPAP and Rabex-5 respectively) as compared to that of 1.021 micron² of LacZ-RNAi line. The role of peroxisomes in degenerative pathways causing devastating complex diseases has not been extensively explored. Therefore, finding genes effecting dynamics of peroxisomes may significantly contribute to finding biological markers and therapeutic interventions for curing or preventing a disease. MTPAP and Rabex-5 genes seem to be potential candidate genes for this purpose.

Complex Traits Posters - Wednesday

PB1388. Genetic analysis of bruxism and its associations to sleep, psychiatric and behavioural traits.

Authors:

T. Strausz¹, S. Strausz^{1,2}, T. Palotie^{3,4}, J. Ahlberg^{3,4}, FinnGen, H. M. Ollila^{1,5,6,7}; ¹Inst. for Molecular Med., HiLIFE, Univ. of Helsinki, Helsinki, Finland, ²Dept. of Genetics, Stanford Univ. Sch. of Med., Stanford, CA, ³Dept. of Oral and Maxillofacial Diseases, Head and Neck Ctr., Helsinki Univ. Hosp., Helsinki, Finland, ⁴Dept. of Oral and Maxillofacial Diseases, Clinicum, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, ⁵Broad Inst. of MIT and Harvard, Cambridge, MA, ⁶Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁷Anesthesia, Critical Care, and Pain Med., Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA

Abstract Body:

Sleep bruxism (SB) is a sleep related complex reflex characterized by a cascade of events: tachycardia, deep breathing, blood pressure rise, forceful tooth grinding and jaw clenching and swallowing. Severe and treatment requiring bruxism occurs in 5-10% of the population, with tooth wear or damage and painful conditions such as headaches burdening not only individuals but also health providers. SB reportedly has several underlying factors: psychological, behavioural, pharmacological and physiological. These factors, with varying degrees of evidence for their role at present, are likely to act on an underlying genetic predisposition that is probably polygenic in nature given the lack of evidence for major genes, not to mention the lack overall of sufficiently powered genetic studies even though half of the variation across populations has been suggested to be genetic in origin.

We conducted the first register-based study with 377,277 individuals including 12,297 SB patients (ICD-10 code G47.8 or F45.8) using data from the FinnGen cohort in combination with data from primary care and hospital registries. First, our epidemiological analysis revealed that age (OR=0.97, $P<2.0\times 10^{-16}$), sex (OR=2.79, $P<2.0\times 10^{-16}$), sleep apnoea (OR=1.89, $P<2.0\times 10^{-16}$), gastroesophageal reflux disease (OR=2.06, $P<2.0\times 10^{-16}$), myalgia (OR=4.42, $P<2.0\times 10^{-16}$) and depression (OR=1.80, $P<2.0\times 10^{-16}$) associated with the risk of bruxism as previously reported and hence this analysis also validates our data. Second, in the genome-wide association analysis (GWAS) for SB, we identified one associated genetic locus ($P<5.0\times 10^{-8}$): rs10193179, an intron variant in *MYO3B*. Third, we calculated genetic correlations between bruxism and its known epidemiological correlates and sleep traits revealing significant correlations: sleep apnoea ($rg=0.32$, $P=9.59\times 10^{-6}$), gastroesophageal reflux disease ($rg=0.63$, $P=1.30\times 10^{-8}$), myalgia ($rg=0.76$, $P=1.52\times 10^{-6}$), depression ($rg=0.50$, $P=3.22\times 10^{-8}$), temporomandibular disorder ($rg=0.73$, $P=7.53\times 10^{-9}$), use of hypnotics and sedatives ($rg=0.46$, $P=4.60\times 10^{-9}$), insomnia ($rg=0.31$, $P=0.031$), sleepiness ($rg=0.14$, $P=0.030$) and sleep efficiency ($rg=-0.17$, $P=0.038$).

This study shows that bruxism can be studied through large biobank-based cohorts opening new possibilities for genetic elucidation. It is particularly interesting that we found strong genetic correlations between bruxism and its comorbidities and bruxism and sleep traits, suggesting that these diseases and traits may share a common genetic background. This connection may also bring new insights into the diagnosis and treatment of bruxism.

Complex Traits Posters - Thursday

PB1389. Genetic analysis using large biobank controls validates known genes and discover novel genes in Hirschsprung disease.

Authors:

M. Fu, S. Chatterjee, M. Erazo, H. Berk-Rauch, A. Chakravarti; New York Univ. Sch. of Med., New York, NY

Abstract Body:

Background

Hirschsprung disease (HSCR) is a congenital neurodevelopmental disorder of the enteric nervous system (ENS) and the most common cause of intestinal obstruction in newborns. Recent (Tilghman et al., 2019) molecular analyses of this multifactorial disorder has identified a total of 33 genes and loci explaining 63% of its population attributable risk. The remainder are likely rare pathogenic segregating and de novo variants at many ENS genes. Here we aimed to discover such genes and variants using the largest collection of HSCR patients of European ancestry as cases and the UK Biobank (UKB) data as controls.

Methods

We ascertained 345 HSCR affected individuals as cases and selected 50 controls per case from ~200K UKB exomes based on the nearest Euclidean distance of common ($\geq 1\%$) variants by principal components. Statistical analyses were performed using GATK4 and DeepVariant for cases and controls, respectively. Extensive quality control analyses at the variant level were performed to confirm the comparability of case and control data. We focused our genetic analyses on rare variants ($MAF \leq 0.1\%$ in gnomAD v3) and performed genome-wide burden enrichment analysis by variant types (i.e., loss-of-function (LoF), insertions/deletions, missense & synonymous). Individual genes were assessed using the odds ratio followed by database annotations and pathway enrichment analysis.

Results

We found HSCR was enriched with LoFs as compared to all other variant types and identified 105 genes associated with HSCR. Among them, 22, inclusive of the HSCR major genes *RET* and *EDNRB*, were enriched for gene expression in the human embryonic gut. These genes were enriched for transmembrane receptor protein tyrosine kinase activity with potential interactions with the SOX10, TFAP2C and NKX2-1 transcription factors, all of which are known to be influential in neural cell development.

Conclusion

We demonstrated how very large external biobank data can be used as effective controls to increase the statistical power of gene discovery, adding an additional 20 new genes to the known set of 33 HSCR genes and loci.

Complex Traits Posters - Wednesday

PB1390. Genetic and environmental regulation of caudate nucleus transcriptome in schizophrenia

Authors:

K. Benjamin^{1,2}, A. S. Feltrin¹, B. R. André¹, A. E. Jaffe¹, J. M. Stolz¹, L. Collado-Torres¹, L. A. Huuki¹, Q. Chen¹, E. E. Burke¹, R. Arora¹, J. Shin¹, W. S. Ulrich¹, A. Deep-Soboslay¹, R. Tao¹, BrainSeq Consortium, T. M. Hyde¹, J. E. Kleinman¹, J. A. Erwin¹, D. R. Weinberger¹, A. C. M. Paquola¹; ¹Lieber Inst. for Brain Dev., Baltimore, MD, ²Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Abstract Body:

Recent studies of gene expression in the human brain have provided biological insights about the genetic origins of psychiatric disorders, such as schizophrenia. Most of these studies, however, have focused exclusively on cortical regions though subcortical nuclei, such as striatum, have figured prominently in the circuitry implicated in schizophrenia and its dense dopaminergic innervation is targeted by current antipsychotic drugs. To gain novel insight into risk mechanisms underlying schizophrenia, we performed a comprehensive analysis of the genetic and transcriptional landscape of schizophrenia in the postmortem caudate nucleus of 443 individuals. Integrating expression quantitative trait loci (eQTLs) analysis, Mendelian Randomization with the latest schizophrenia GWAS, transcriptome wide association study (TWAS), and differential expression analysis, we identified many new genes associated with schizophrenia risk including possibly the dopamine D2 receptor short isoform. We examined the effect of antipsychotics in the caudate and show extensive influence on gene expression. Using a new approach based on deep neural networks, we construct caudate nucleus gene expression networks that highlight interactions involving schizophrenia risk. Altogether, these analyses provide a new resource for the study of schizophrenia that can bring insight into risk mechanisms and potential novel therapeutic targets.

Complex Traits Posters - Thursday

PB1391. Genetic architecture of asthma in African American Patients

Authors:

X. Chang, M. March, F. Mentch, H. Qu, Y. Liu, J. Glessner, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Asthma is a chronic inflammatory disorder with a strong genetic inheritance. Although over one hundred loci were reported through the genome-wide association study of European populations, the genetic underpinning of asthma in African Americans remains largely elusive. In this study, we aimed to identify new genetic loci associated with asthma in African Americans. Three cohorts were genotyped using the Illumina SNP array platform, totaling 6,975 cases and 4,429 controls. Genotype imputation was performed using the TOPMed reference panel including whole genome sequencing data from over 100,000 individuals. Meta-analysis was conducted to identify genetic loci associated with asthma in African Americans. Our study identified 12 loci surpassing the classical genome-wide significant threshold (5×10^{-8}) including 8 Loci which reached stricter significant threshold (3×10^{-8}). Among them, six loci are associated with enhancer activities, two loci are in DNase I hypersensitive regions, and all of them are associated with regulatory motifs. Moreover, locus 9p24.1 (rs10975467, $P = 1.63 \times 10^{-8}$) has been previously associated with asthma in Europeans. Locus 11q13.4 locus (rs7480008) is an eQTL (expression quantitative trait loci) of XRRA1 in lung ($P = 9.4 \times 10^{-10}$), and locus 13q14.3 (rs1543525) is a sQTL (splicing quantitative trait loci) of DHRS12 in lung ($P = 1.1 \times 10^{-13}$). Our findings demonstrate that some genetic risk factors in asthma are shared among populations of African and European descent and provide candidate genetic loci for therapeutic target identification and prioritization.

Complex Traits Posters - Wednesday

PB1392*. Genetic architecture of insulin-stimulated glucose uptake reveals novel regulators of post-prandial glucose metabolism

Authors:

A. Williamson^{1,2}, D. M. Norris², X. Yin³, K. Broadaway⁴, A. H. Moxley⁴, S. Vadlamudi⁴, E. P. Wilson⁴, N. Grarup⁵, M. Boehnke³, N. J. Wareham¹, K. L. Mohlke⁴, E. Wheeler¹, S. O'Rahilly², D. J. Fazakerley², C. Langenberg^{1,6}; ¹MRC Epidemiology Unit, Wellcome-MRC Inst. of Metabolic Sci., Univ. of Cambridge, Cambridge, United Kingdom, ²Wellcome-MRC Inst. of Metabolic Sci. Metabolic Res. Lab., Univ. of Cambridge, Cambridge, United Kingdom, ³Dept. of Biostatistics and Ctr. for Statistical Genetics, Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, ⁴Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, ⁵Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark, ⁶Computational Med., Berlin Inst. of Hlth.at Charité–Univ.smedizin, Berlin, Germany

Abstract Body:

Insulin resistance and pancreatic beta cell dysfunction are the two key pathophysiological mechanisms predisposing to type 2 diabetes (T2D). The molecular mechanisms through which insulin controls glucose metabolism in fasting and post-prandial states are distinct, with the former primarily involving the control of expression and phosphorylation of enzymes for gluconeogenesis and glycogenolysis in the liver, while the latter involves the rapid translocation of an intracellular pool of stored glucose transporters to the plasma membrane of muscle cells and adipocytes. Despite this, large-scale genetic studies of insulin resistance have predominantly focused on the fasting state and therefore primarily capture insulin action in the liver.

We conducted a genetic discovery of post-prandial insulin sensitivity using two indices which integrate fasting and glucose-stimulated insulin levels during an oral glucose tolerance test (OGTT): insulin fold change and modified Stumvoll Insulin Sensitivity Index. We meta-analysed 27 genome-wide association studies (GWAS) for each of these traits in up to ~54,000 participants representing European, Hispanic American, and East Asian ancestries.

We identified 12 loci ($p < 5 \times 10^{-8}$), of which 10 were novel, including a signal at *SLC2A4* encoding the canonical insulin sensitive glucose transporter GLUT4, associated with post-challenge insulin resistance and reduced expression of *SLC2A4* in skeletal muscle. All but one of the genetic signals were shared with risk of T2D (posterior probability of colocalisation > 80%). Using an integrative approach, incorporating information on glycaemic trait associations and skeletal muscle gene expression, we prioritised 8 additional loci of interest associated with post-challenge insulin resistance that did not meet the traditional genome wide significance threshold.

Knockdown of 41 candidate genes at these loci in a reporter assay in 3T3-L1 adipocytes identified 10 genes newly implicated in GLUT4 trafficking. Knockdown of these genes resulted in changes in regulation of GLUT4 trafficking (N genes = 8) or GLUT4 expression (N genes = 2).

In summary, through the integration of genetic data and *in vitro* experimental follow-up, we identified loci that affect T2D risk through impaired glucose uptake in peripheral tissues.

Complex Traits Posters - Thursday

PB1393. Genetic associations of rare variants in alcohol and tobacco use in up-to 526,400 individuals

Authors:

S-K. Jang¹, X. Wang², J. Otto¹; ¹Univ. of Minnesota, Minneapolis, MN, ²Penn State Coll. of Med., Hershey, PA

Abstract Body:

The use and misuse of alcohol and tobacco are major risk factors of common diseases and disability around the globe with about 50% of population risk attributed to genetic variation. While recent large-scale GWAS meta-analysis was successful at discovering common variants associated with substance use, our knowledge on causal genes and biological mechanisms of these complex behaviors are limited as common variant hits have overall small effect sizes, unclear individual significance, and are frequently in substantial LD. Research into rare variant associations and its implicated genes may identify variants with high impact on alcohol and tobacco use and provide an increasingly refined complete picture of genetic architecture of substance use. Here, we performed rare variant association test with up to 526,400 samples of predominantly European ancestry with Hispanic (N~5,458), African (N~12,813), and East Asian (N~4,306) subsamples. We meta-analyzed data across UK Biobank exome sequences (N~194K), TOPMed-imputed UK biobank microarray genotypes (N~276K), and deep whole-genome sequences from Trans-omics Precision Medicine (TOPMed) consortium (N~53K). We tested the association between alcohol/tobacco use and ~100 million variants with MAF < 1% and minor allele count greater than five. Using fixed-effects meta-analysis, we identified four rare variants that are significantly associated with smoking and three associated with alcohol use (p -value < 1×10^{-9}) and conditionally independent from common variants. These include an intronic variant of *GABRB2* (rs544298483; MAF < 0.01%, T allele associated with ~1SD increase in smoking risk, p -value= 1.62×10^{-10}) for smoking initiation which has higher allele frequency in East-Asian population. We performed gene-based test where rare protein-altering variants were aggregated based on an adaptive MAF threshold using two different masks, one with using only protein-truncating variants (i.e., stop, start, frameshift, splice acceptor/donor) and another using both protein-truncating and missense variants. We found six significant genes associated with tobacco and alcohol use, including the well-known alcohol metabolism gene *ADH1C* (p -value= 1.12×10^{-27}). We will further compare common and rare variant gene signals and test rare variant burden at the level of biological pathways and cell-types, followed by functional validation experiment.

Complex Traits Posters - Wednesday

PB1394. Genetic characterization of central serous chorioretinopathy and pleiotropic effects with age-related macular degeneration

Authors:

J. Rämö^{1,2,3}, E. Abner⁴, E. H. C. van Dijk⁵, X. Wang³, S. H. Choi³, T. Kiiskinen¹, R. Kajanne¹, J. Mehtonen¹, FinnGen, P. Palta^{6,4}, M. Kurki^{3,1}, L. Sobrin⁷, S. Yzer⁸, S. A. Lubitz^{9,3}, P. T. Ellinor^{9,3}, A. I. den Hollander^{10,11}, T. Esko¹², M. J. Daly^{9,3,1}, J. A. Turunen¹³, A. Palotie^{6,3,9}, C. J. Boon^{14,15}, E. J. Rossin^{16,2}; ¹Inst. for Molecular Med. Finland, Helsinki, Finland, ²Massachusetts Eye and Ear, Boston, MA, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Estonian Genome Ctr., Univ. of Tartu, Tartu, Estonia, ⁵Leiden Univ. Med. Ctr., Leiden, Netherlands, ⁶Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, ⁷Harvard Med Sch, Boston, MA, ⁸Dept. of Ophthalmology, Radboudumc, Nijmegen, Netherlands, ⁹Massachusetts Gen. Hosp., Boston, MA, ¹⁰Radboud Univ Nijmegen MC, Nijmegen, Netherlands, ¹¹AbbVie, Genomic Res. Ctr., Cambridge, MA, ¹²Estonian Genome Ctr., Univ. of Tartu, Tartu, Tartumaa, Estonia, ¹³Folkhälsan Res. Ctr., Helsinki, Finland, ¹⁴Dept. of Ophthalmology, Amsterdam Univ. Med. Ctr.s, Univ. of Amsterdam, Amsterdam, Netherlands, ¹⁵Dept. of Ophthalmology, Leiden Univ. Med. Ctr., Leiden, Netherlands, ¹⁶Harvard Med. Sch., Dept. of Ophthalmology, Boston, MA

Abstract Body:

Central serous chorioretinopathy (CSC) is a leakage maculopathy affecting vision in working-age individuals. Although ophthalmologists manage the disease routinely, the etiology of CSC is not known and a better genetic understanding could represent a breakthrough. The chronic form of CSC shares some overlapping features with age-related macular degeneration (AMD), yet other features are distinct. Previous CSC GWAS have identified three loci, including a known AMD locus near the gene coding for complement factor H (CFH) with discordant effects on AMD and CSC risk.

We performed a GWAS meta-analysis of CSC in the FinnGen study (552 cases; 343,461 controls), Estonian Biobank (104 cases, 181,063 controls) and 521 previously reported Dutch chronic CSC patients. We replicated two previously known CSC association loci (near *CFH* and *GATA5*) and identified four novel genetic loci marked by common variants near *NOTCH4*, *CD34/CD46* and *PREX1* and the rare intronic variant rs538404658 in *MSRA* (OR 24.6, $p = 4.1 \times 10^{-9}$, AF 0.1%). Rs538404658 was associated with CSC in FinnGen and Estonian Biobank with similar odds ratios (24.8 and 22.9).

Among 34 reported AMD loci, 5 of 6 loci near complement pathway genes (*CFH*, *CFI*, *C2/C4A/C4B/CFB*, *VTN* and *C9* loci) had discordant effect directions for AMD and CSC (effect estimate $r = -0.76$ across all 6 loci).

We then constructed a polygenic score (PGS) for AMD in FinnGen using PRS-CS based on previous GWAS of AMD and evaluated its association with CSC risk in FinnGen. The AMD-PGS was negatively associated with diagnosis of CSC (OR 0.76 per +1 SD in AMD-PGS, $p = 7.4 \times 10^{-10}$). This association was attenuated by removing the *CFH* region (OR 0.89, $p = 0.006$) and further by removing all 6 AMD loci near complement genes from the PGS (OR 0.93, $p = 0.09$).

Finally, we constructed a PGS for each of the concentrations of circulating CFH and five CFH-related proteins. CFH-PGS was not associated with AMD or CSC, but a PGS for circulating CHFR4 was predictive of AMD (OR 1.06, $p = 8.4 \times 10^{-7}$) and had a discordant but nonsignificant effect estimate for CSC (OR 0.93, $p = 0.073$).

Here, we double the number of genetic risk loci for CSC. We demonstrate that polygenic AMD risk is negatively correlated with CSC risk, and that this pleiotropic effect is mostly mediated by complement-related loci.

Complex Traits Posters - Thursday

PB1395. Genetic colocalization between Covid-19 and diseases of the immune system.

Authors:

M. Zechner, E. Pairo-Castineira, K. Baillie; Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract Body:

The Covid-19 pandemic has had a devastating global impact over the past three years. Despite rapid scientific progress, Covid-19 remains a devastating force of morbidity and mortality across the globe, with the need for better treatment options to emerge from a more thorough understanding of the disease. Here, we aim to further our understanding of the functional mechanisms behind severe Covid-19 by studying its genetic intersection with diseases that affect the immune system. Pleiotropy - whereby a single gene can affect multiple traits - is widespread throughout the human genome, and can be used to elucidate disease mechanisms by highlighting which diseases are likely to share common molecular architecture and therefore potential drug targets. Furthermore, analysis of combined genetic data from multiple diseases allows us to identify genetic variants that have so far been associated with other disease phenotypes but may impact the outcome of Covid-19 as well.

It has been shown that the inflammatory response to SARS-coronavirus-2 (SARS-CoV-2) infection has a major impact on Covid-19 pathogenesis. Both critical illness caused by Covid-19 and immune system-mediated diseases in general are known to have strong genetic components. We therefore chose to compare our genome-wide association study (GWAS) data of patients with severe Covid-19 with 56 GWAS of infectious and autoimmune diseases. In order to ascertain shared genetic loci we performed a genome-wide multi-trait colocalization analysis using the Bayesian algorithm HyPrColoc (Hypothesis Prioritisation for multi-trait Colocalization), which calculates whether the traits share the same causal variant in a genomic region. To date we have found multiple genetic connections between Covid-19 and diseases affecting the immune system, including idiopathic pulmonary fibrosis, asthma, Crohn's disease, systemic lupus erythematosus and allergic disease. Our next step is to identify affected genes in order to gain insight into the functional mechanisms and pinpoint targets for treatment.

Overall, our results shed light on which functional mechanisms severe Covid-19 shares with other illnesses of the immune system, opening up drug targets for already-approved medicines to be used for the treatment of Covid-19.

Complex Traits Posters - Wednesday

PB1396. Genetic epidemiology of carotid intima-media thickness in Sub-Saharan African populations contributes to atherosclerosis biology

Authors:

P. Boua¹, J-T. Brandenburg², H. Sorgho¹, A. Choudhury³, H. Tinto¹, C. G. Mathew⁴, M. Ramsay⁵, as members of AWI-Gen, and the H3Africa Consortium; ¹Clinical Res. Unit of Nanoro (CRUN), Ouagadougou, Burkina Faso, ²Sydney Brenner Inst. for Molecular BioSci., Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg,, South Africa, ³Sydney Brenner Inst. for Molecular BioSci., Faculty of Hlth.Sci., Univ. of Witwatersrand, Johannesburg, South Africa, ⁴Dept. of Med. and Molecular Genetics, Faculty of Life Sci. and Med., King's Coll. London, London, United Kingdom, ⁵Sydney Brenner Inst. for Molecular BioSci., Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, Gauteng, South Africa

Abstract Body:

Atherosclerosis is a complex, progressive disorder affecting large and medium-sized arteries, preceding the development of cardiovascular diseases (CVDs) such as ischemic heart disease, stroke, heart failure, peripheral arterial disease and rheumatic heart disease. Despite the epidemiological complexity of CVDs in sub-Saharan Africa (SSA), genetic susceptibility remains understudied. The evolutionary roots and genetic diversity of the continent represent an opportunity to shed new light on the genetic architecture of complex traits such as atherosclerosis. We used carotid intima-media thickness (cIMT) to investigate the genetic susceptibility to atherosclerosis in SSA. cIMT was measured by ultrasound and genotyping was performed using the H3Africa genotyping array (2.3M SNPs) followed by imputation using Sanger Imputation Server. GWAS was performed in a resident African population-based cohort of 7894 adults from the AWI-Gen study. The current work outlined how the benefit of using multiple genetic epidemiology approaches lead to discovery in underrepresented populations. Additionally to the previous findings from the AWI-Gen cohort (sex-stratification gwas and gene-smoking interaction in West African men), we present results from stratified GWAS of cIMT according to hypertension and obesity. The GWAS identified two new cIMT-associated African-specific loci in *SIRPA* ($p=4.7E-8$) and *FBXL17* ($2.5E-8$). Sex-stratified analysis revealed two male-specific loci in *SNX29* ($6.3E-9$) and *MAP3K7* ($5.3E-8$), and two female-specific loci in *ARNT2* ($2.4E-9$) and *PROK1* ($1.0E-8$). Gene-based analysis resulted with significant associations for *CALD1* ($P=5.9E-7 < 2.6E-6$) and *FLT4* ($P=4.3E-7 < 2.6E-6$). New variants were identified for the gene-smoking interactions for cIMT within the previously described *RCBTB1* region ($3.1E-8$), a newly identified locus in the *BCHE* ($2.2E-8$) region, and variants in the regulatory and open chromatin regions of *TBC1D8* ($5.9E-9$). We found loci associated with cIMT in non-overweight-obese participants for *COX7C* ($4.1E-8$), and with hypertension for *PRSS38* ($1.5E-8$). Overall, this study successfully identified 13 novel loci linked to atherosclerosis biology. These results highlight the role of sex differences, gene-smoking interactions, obesity and hypertension in the pathophysiology of atherosclerosis. Inclusion of under-represented populations in medical genomics research is essential for advancing an understanding of complex trait genetics.

Complex Traits Posters - Thursday

PB1397. Genetic Exploration of Military Sudden Cardiac Arrest (GEMini): Genome Sequencing Results from the First 24 Participants.

Authors:

L. Hellwig^{1,2}, M. D. Wilkerson¹, J. Villar^{1,2}, X. Zhang¹, A. Mains¹, M. J. De Castro³, C. Dalgard¹, M. Haigney¹; ¹Uniformed Services Univ., Bethesda, MD, ²Henry M. Jackson Fndn. for the Advancement of Military Med., Inc., Bethesda, MD, ³Air Force Med. Genetics Ctr., Keesler AFB, MS

Abstract Body:

Background: Heart disease is the most common cause of death in the United States and cardiac arrest is the most common cause of death among heart disease patients. A significant proportion of sudden cardiac death is attributed to underlying genetic risk factors, particularly in the young and the incidence of sudden cardiac death in military recruits is four-fold higher than civilian athletes. Identification of disease-causing genetic variants in individuals who have sudden cardiac arrest (SCA) is critical to their medical management and informs the care of at-risk family members.

Methods: The GEMini study seeks to explore the genetic causes of SCA in military members and beneficiaries. SCA survivors who experienced SCA <40 years old are recruited, consented, and enrolled. Participants have undergone extensive phenotypic evaluation and undergo genome sequencing (GS). Clinically validated primary and secondary genomic findings (ACMGv3) are returned. Samples can also be used to study the functional effects of genetic variants.

Results: At this time, 24 participants have been enrolled. Genomes of the first 22 participants were analyzed, including 20 SCA survivors and 2 first degree relatives. The 22 total SCA survivors also had clinical genetic testing. Only 6 had clinical genetic testing prior to meeting with the study team and the clinical cardiogenetics service. Prior to enrollment, no definitive disease-causing genetic changes were identified in any of the SCA survivors. Updated clinical testing and/or GS identified clinically actionable findings in 20% (4/20) of SCA survivors. One pathogenic variant in *PKP2* and one pathogenic variant in *KCNQ1* were identified via clinical testing upon study enrollment and were consistent with the participants' phenotypes. One pathogenic variant was identified in *SDHA* and one likely pathogenic variant was identified in *ALPK3* via GS. In the first 22 GEMini participants with GS, there were 143 total variants of uncertain significance (VUS) identified within a panel of 450 cardiovascular disease-related genes.

Discussion: Genomic testing identifies clinically-actionable findings in SCA survivors. The GEMini study also suggests that standard-of-care clinical genetic testing for SCA survivors may not be routine in the MHS. These results demonstrate the complexity of GS analysis such as differentiating secondary versus primary findings. So far, GEMini revealed a large number of VUSes in SCA survivors. Additional functional studies will be conducted on high priority VUSes in order to gain new variant-specific information, potentially improving variant classification and clinical management.

Complex Traits Posters - Wednesday

PB1398. Genetic Interactions Implicate Disruption in Ciliogenesis in the Etiology of Non-syndromic Orofacial Clefts

Authors:

A. Alade¹, N. Mukhopadhyay², E. Zeng¹, M. Withanage¹, W. Awotoye³, A. Oladayo⁴, T. Busch¹, V. Sule⁵, L. Gowans⁶, P. Mossey⁷, M. Eshete⁸, W. Adeyemo⁹, A. Adeyemo¹⁰, J. Murray¹¹, S. Weinberg¹², M. Marazita¹³, A. Butali⁴; ¹Iowa Inst. for Oral Hlth.Res., Iowa, IA, ²Univ Pittsburgh/Sch Dental Med, Pittsburgh, PA, ³Univ. of Iowa, Iowa, IA, ⁴Univ. of Iowa, Coralville, IA, ⁵Uiowa, coralville, IA, ⁶Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, ⁷Univ. of Dundee, Dundee, United Kingdom, ⁸Addis Ababa Univ., Addis Ababa, Ethiopia, ⁹Univ. of lagos, Lagos, Nigeria, ¹⁰NIH, Bethesda, MD, ¹¹U of Iowa, Iowa City, IA, ¹²Univ of Pittsburgh, Pittsburgh, PA, ¹³Univ Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA

Abstract Body:

Background: Non-syndromic orofacial clefts (NSOFCs) are complex traits caused by multiple genes and environmental factors. To date, over 60 risk loci have been implicated in the etiology of NSOFCs. However, the effect size of each locus is quite small, and their combined effect do not account for the total heritability of NSOFCs. SNP-SNP interaction has been shown to be a stronger mechanism for understanding the risk for NSOFCs and to explain some of the missing heritability. Therefore, we conducted a SNP-SNP interaction analysis using data from the African NSOFCs genome-wide association study (GWAS) and replicated our findings in the Pittsburgh Orofacial Clefts (POFC) multiethnic GWAS data. **Methods:** The African and POFC multiethnic GWAS study consisted of 1898 (1000 cases and 898 controls) and 11644 (2915 cases and 8729 controls) participants, respectively. First, we used the w-test to test for the association between SNPs pairs and NSOFCs allowing for interaction. Subsequently, we conducted a likelihood ratio test on the significantly associated SNP pairs comparing the full model (includes the interaction term) with the reduced model (excludes the interaction term), while adjusting for sex and the top principal components. These analyses were done for SNPs within genes in the folate pathway for both datasets, and genome-wide for the African dataset. **Results:** Among the folate pathway genes, we found significant pairwise interactions between SNPs in *MTHFD1* and *MTHFR*, and NSOFCs in both datasets. Genome-wide, we found significant associations with 16 SNP pairs. However, only one SNP pair (rs41269377 (1q43) and rs11558540 (*SAR1B*)) remained significant based on interaction effects alone. **Discussion:** *MTHFR* is a known cleft candidate gene while *MTHFD1* is a widely studied cleft candidate gene with inconsistent results across populations. The SNP rs41269377 lies within a non-coding region (1q43) and an analysis of the topologically associated domain (TAD) around the SNP identified *SDCCAG8* gene as a potential candidate locus. Mouse knock-out models of this gene showed a cleft palate phenotype. Additionally, *SDCCAG8* gene interacts with RAB effector proteins forming complexes important for primary cilia formation. The cilium is a structure essential for hedgehog signaling which is an established signaling pathway in the pathogenesis of NSOFCs. **Conclusions:** In our search for genetic risk factors for NSOFCs, we found evidence for gene-gene joint effects and interactions between folate pathway genes as well as for genes involved in ciliogenesis. Additional studies will be needed to explore the biological mechanisms mediated by these interactions.

Complex Traits Posters - Thursday

PB1399. Genetic Mechanisms of Supraventricular Tachycardias

Authors:

L-C. Weng^{1,2}, S. Khurshid^{1,2}, A. W. Hall³, V. N. Morrill¹, V. Nauffal^{2,4}, D. Fatkin^{5,6,7}, M. S. Olesen^{8,9}, S. H. Shah^{10,11}, D. Friedman¹², T. Esko¹³, C. M. Haggerty^{14,15}, M. J. Cutler¹⁶, S. Knight^{16,17}, G. Nadkarni¹⁸, R. Loos¹⁹, S. Damrauer²⁰, D. Rader²¹, G. M. Marcus²², M. Shoemaker²³, D. Roden²⁴, J. D. Roberts²⁵, L. Eckhardt²⁶, P. Wilson²⁷, K. Cho^{4,28,29}, Y. Sun³⁰, T. M. Treu³¹, S. Choi², C. Roselli², P. T. Ellinor^{1,2}, S. A. Lubitz^{1,2}; ¹Massachusetts Gen. Hosp., Boston, MA, ²Cardiovascular Disease Initiative, The Broad Inst. of MIT and Harvard, Cambridge, MA, ³The Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Brigham and Women's Hosp., Boston, MA, ⁵Victor Chang Cardiac Res. Inst., Darlinghurst, New South Wales, Australia, ⁶Sch. of Clinical Med., UNSW Med. and Hlth., UNSW Sydney, Kensington, New South Wales, Australia, ⁷St Vincent's Hosp., Darlinghurst, New South Wales, Australia, ⁸Dept. of BioMed. Sci., Univ. of Copenhagen, Copenhagen, Denmark, ⁹Lab. for Molecular Cardiology, Dept. of Cardiology, Copenhagen Univ. Hosp. - RigsHosp.et, Copenhagen, Denmark, ¹⁰Div. of Cardiology, Dept. of Med., Duke Univ. Sch. of Med., Durham, NC, ¹¹Duke Molecular Physiology Inst., Duke Univ. Sch. of Med., Durham, NC, ¹²Div. of Electrophysiology, Yale Univ. Sch. of Med., New Haven, CT, ¹³Univ. of Tartu, Tartu, Estonia, ¹⁴Dept. of Translational Data Sci. and Informatics, Geisinger, Danville, PA, ¹⁵Heart Inst., Geisinger, Danville, PA, ¹⁶Intermountain Heart Inst., Intermountain Med. Ctr., Murray, UT, ¹⁷Dept. of Med., Univ. of Utah, Salt Lake City, UT, ¹⁸Icahn Sch. of Med. at Mount Sinai, New York, NY, ¹⁹Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, USA; Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Dept. of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark, ²⁰Hosp. of the Univ. of Pennsylvania, Philadelphia, PA, ²¹Div. of Cardiovascular Med., Dept. of Med., Perelman Sch. of Med. at the Univ. of Pennsylvania, Philadelphia, PA, ²²Div. of Cardiology, Univ. of California, San Francisco, CA, ²³Dept. of Med., Div. of Cardiovascular Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, ²⁴VUMC, Nashville, TN, ²⁵Section of Cardiac Electrophysiology, Div. of Cardiology, Dept. of Med., Western Univ., London, ON, Canada, ²⁶Dept. of Med., Div. of Cardiovascular Med., Univ. of Wisconsin–Madison, Madison, WI, ²⁷Dept. of Epidemiology, Rollins Sch. of Publ. Hlth., Emory Univ., Atlanta, GA, ²⁸VA Boston Hlth.care System, Boston, MA, ²⁹Harvard Med. Sch., Boston, MA, ³⁰Emory Univ., Atlanta, GA, ³¹MAVERIC, VA Boston Hlth.care System, Boston, MA

Abstract Body:

Background: Supraventricular tachycardias such as atrioventricular nodal reentrant tachycardia (AVNRT), and Wolff-Parkinson-White syndrome / atrioventricular reentrant tachycardia (WPW/AVRT) are common heart rhythm abnormalities associated with substantial morbidity. Supraventricular tachycardias have been reported to be heritable, but the genetic determinants are poorly understood. **Methods:** We performed genome-wide association studies from seven studies and meta-analyzed results for each condition separately. For AVNRT, we included 1,996 AVNRT cases and more than 94,000 controls. For WPW/AVRT we included 2,494 cases and over 1 million controls. Significant variants were those with $p < 5 \times 10^{-8}$, and suggestive variants were those with $p < 1 \times 10^{-6}$. We identified expression quantitative trait loci from GTEx version 8 and performed transcriptome-wide analyses. We examined associations between identified variants and cardiovascular and electrocardiogram traits from prior studies. **Results:** For AVNRT (mean age 49-65 years, 30% female, and 98% white), we identified one genome-wide association locus near *NKX2-5* and three suggestive loci near *TTN*, *RAPGEF5*, *LOC105372518*. For WPW/AVRT (mean age 37-65 years, 58% female, and 80% white), we identified two genome-wide associated loci near *CCDC141* and *SCN5A*, and three suggestive loci near *VIT/STRN*, *NEDD9*, and *MYB*. Variants at *TTN* and *STRN* were associated

ASHG 2022 Annual Meeting Poster Abstracts

with *FKBP7* and *STRN* expression in heart tissues, respectively. Associated variants near *NKX2-5*, *CCDC141*, and *SCN5A* were previously associated with heart rate, resting heart rate, and QRS/QT/PR intervals, respectively. **Conclusions:** Our findings highlight genes associated with cardiac development, ion channels, and myocardium as important potential effectors for the risk of supraventricular tachycardias.

Complex Traits Posters - Wednesday

PB1400. Genetic Overlap among Alzheimer's Disease and Five Psychiatric Disorders: Genetic Convergence and Specificity.

Authors:

N. Enduru, B. S. Fernandes, Y. Dai, Z. Zhao; Univ Texas HSC Houston, Houston, TX

Abstract Body:

Background: Alzheimer's Disease (AD) has been one of the most significant health problems for the older population. Several psychiatric disorders have been associated with AD. Clinical and epidemiological evidence suggests that major depressive disorder (MDD), bipolar disorder (BD), schizophrenia (SZ), attention deficit hyperactive disorder (ADHD), and autism spectrum disorder (ASD) increase the risk of developing AD in the later stage of life. Yet very few studies investigated the genetic relationship between AD and these psychiatric disorders.

Methods: We obtained genome-wide association study (GWAS) summary statistics data on AD, BD, MDD, ADHD, and ASD from the Psychiatric Genomic Consortium (PGC) working group. We implemented two approaches to identify genetic overlap between any two traits. We applied conjunctive false discovery rate (cFDR) analysis through pleioFDR, and the case-case genome-wide association study (CC-GWAS) method. Finally, we tested causality between psychiatric disorders and AD using Mendelian randomization based on the 2SMR method.

Results: Overall there was no significant genetic correlation (r_g) between AD and MDD ($r_g = -0.06$; $P = 0.2$), AD and BD ($r_g = 0.006$; $P = 0.9$), AD and SZ ($r_g = 0.04$; $P = 0.3$), and AD and ASD ($r_g = 0.0008$; $P = 0.99$). Only AD and ADHD had significant genetic correlation ($r_g = -0.23$; $P = 0.01$). After examining pleiotropic enrichment of AD given BD (and other disorders), we then performed cFDR to identify the shared loci for a given pair of traits. We identified two SNPs with a $cFDR_{(AD\&BD)} < 0.05$. *WAC(rs332183)*, $cFDR_{(AD\&BD)} = 0.030$, $Z_{(AD)} = -3.95$, $Z_{(BD)} = 3.97$) and *MTSS1L(rs56264804)*, $cFDR_{(AD\&BD)} = 0.012$, $Z_{(AD)} = -4.36$, $Z_{(BD)} = 4.19$) was associated with AD and BD with opposite direction of effect. *MTCH2(rs10838738)*, $cFDR_{(AD\&MDD)} = 0.007$, $Z_{(AD)} = -5.76$, $Z_{(MDD)} = -4.31$) was associated with AD and MDD with same direction of effect. *DDB2(rs11039131)*, $cFDR_{(AD\&SZ)} = 0.03$, $Z_{(AD)} = -5.11$, $Z_{(SZ)} = -3.88$) and *ADAM10(rs2555356)*, $cFDR_{(AD\&SZ)} = 0.012$, $Z_{(AD)} = -4.55$, $Z_{(SZ)} = -3.93$) was associated with AD and SZ with same direction of effect. We did not find any overlapping SNPs based on the conjunctive analysis of AD with ADHD or ASD. CC-GWAS identified 86 independent loci among AD and the 5 psychiatric disorders, of which 19 are CC-GWAS specific loci; 11 of which were not reported previously. 2SMR results did not support psychiatric disorders as causal factors for AD (MR Egger > 0.05).

Discussion: Overall, there is a polygenic overlap between AD and psychiatric disorders, suggesting shared pathogenesis. However, we failed to find any causal relationship between psychiatric disorders and AD.

Complex Traits Posters - Thursday

PB1401. Genetic overlap of idiopathic pulmonary fibrosis and hypertension.

Authors:

G. Parcesepe¹, R. Allen¹, B. Guillen-Guio², R. G. Jenkins^{3,4}, L. V. Wain^{1,5}, c. DEMISTIFI³; ¹Univ. of Leicester, Leicester, United Kingdom, ²Hosp. Univ.rio Ntra. Sra. de Candelaria, Santa Cruz de Tenerife, Spain, ³Univ. of Nottingham, Nottingham, United Kingdom, ⁴Nottingham BioMed. Res. Ctr., Nottingham, United Kingdom, ⁵Leicester Respiratory BioMed. Res. Ctr., Glenfield Hosp., Leicester, United Kingdom

Abstract Body:

Fibrosis is a feature of multiple chronic conditions. Idiopathic Pulmonary Fibrosis (IPF) is the most well-known fibrotic disease of the lung, whereby excess deposition of extracellular matrix impairs gas exchange and leads to respiratory failure. Up to 50% of individuals with IPF have co-morbid hypertension, defined using systolic and diastolic blood pressure (SBP, DBP). High blood pressure can lead to organ fibrosis or can be a consequence of artery stiffening due to fibrosis. Studies have implicated common processes, such as TGF- β signalling, in both IPF and blood pressure regulation.

Genome-wide association studies (GWAS) of blood pressure have identified >900 independent signals of association. GWAS of IPF risk have identified 20 independent signals. We hypothesised that there are shared genetic risk factors for IPF and hypertension that could be informative about shared fibrotic pathways.

We analysed genome-wide genetic correlation using LD score regression (LDSC), and locus-specific overlap, using the largest publicly available GWAS of SBP, DBP and IPF. Genome-wide genetic correlation was low (IPF with; SBP = -0.077, $p=0.02$; DBP = -0.027, $p=0.43$). 7 variants previously reported for association with IPF were also significantly associated ($P<1\times 10^{-5}$) with SBP or DBP. Variants near *MAD1L1* and *MAPT* increased risk of both IPF and hypertension, whilst variants near *TERC*, *FAM13A*, *TERT*, and *DEPTOR*, and at 17q22.1 had opposite directions of effect on risk of IPF and hypertension.

These findings support that there are shared fibrotic mechanisms between IPF and hypertension. The opposite effects of variants at specific loci highlight the need for caution when considering therapeutic targeting of these shared pathways.

Complex Traits Posters - Wednesday

PB1402. Genetic pleiotropy effects on kidney function and soluble receptor for advanced glycation end-products: The Long Life Family Study (LLFS).

Authors:

M. Feitosa¹, B. Thyagarajan², M. Wojczynski¹, S. Lin¹, A. Kuipers³, S. Acharya¹, A. Kulminski⁴, G. J. Patti¹, M. Brent¹, K. Christensen⁵, J. Zmuda³, M. Province¹; ¹Washington Univ Sch Med., St Louis, MO, ²Univ. of Minnesota, Minneapolis, MN, ³Univ of Pittsburgh, Pittsburgh, PA, ⁴Duke Univ, Durham, NC, ⁵Univ. of Southern Denmark, Odense, Denmark

Abstract Body:

Patients with chronic kidney disease (CKD) have increased oxidative stress and chronic inflammation, which may escalate the production of advanced glycation end-products (AGE). The estimated glomerular filtration rate (eGFR) is used for detecting and managing kidney diseases. Low eGFR and high circulating level of soluble receptor for AGE (sRAGE) are associated with aging, CKD, and cardiovascular risk. We estimated the eGFR using serum creatinine (eGFRcre) and cystatin C (eGFRcys) from CKD Epidemiology Collaboration (2021). We evaluated whether eGFRcre (and/or eGFRcys) share pleiotropic genetic factors with sRAGE, assessing whole-genome sequencing variants with minor allele count ≥ 20 in 4,182 individuals (age range: 24-110) from exceptional healthy-aging families from the LLFS (Wojczynski et al., 2021). As eGFRcre and eGFRcys are correlated negatively with sRAGE ($r=-0.25$, $P=1.4 \times 10^{-37}$, and $r=-0.30$, $P=2.7 \times 10^{-53}$, respectively), we employed correlated meta-analyses (CMA) (Province et al., 2013) to test whether combining genome-wide association (GWAS) P -values enhances the ability to detect variant sharing effects on these correlated traits. We considered a variant pleiotropic if the univariate GWA $P \leq 0.01$ and CMA $P \leq 5.0 \times 10^{-08}$. We identified 15 novel loci for eGFRcre (and/or eGFRcys) demonstrating pleiotropy with sRAGE, including *FAFI-CDKN2C-MIR4421*, *SH3GLB1-SELENOF*, *STK32B*, *TMEM144-RXF1*, *CREB5*, *CNTNAP2*, *LINC00707*, *LOC101928272-LINC02670*, *LINC02737-CNTN5*, *RIMBP2*, *POLRID-GSX1*, *NPAS3*, *LINC02301-SNORD3P3*, *CASC17-ROCR*, and *GPRI42-GPRC5C-LINC01987-LINC01973*. Some loci were reported as GWAS suggestive ($P < 10^{-7}$) for eGFRcre, CKD, diabetic kidney disease, blood urea nitrogen levels, and serum uric acid levels. Additionally, we identified loci harboring genes with roles in pathways, including CREB (cAMP response element-binding protein), PI3K-Akt signaling, integrin, and apoptotic signaling. Our findings show pleiotropic genetic associations with eGFR and sRAGE levels, which may explain part of the correlated genetic architecture between kidney function, oxidative stress, low-grade chronic inflammation, and their shared pathways in accelerated biological aging.

Complex Traits Posters - Thursday

PB1403. Genetic prediction of weight and waist circumference regain after weight loss with Intensive Lifestyle Intervention: A secondary analysis of the Look AHEAD trial.

Authors:

M. Revsbech Christiansen^{1,2}, T. O. Kilpeläinen¹, J. M. McCaffery²; ¹Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Copenhagen, Denmark, ²Dept. of Allied Hlth.Sci., Univ. of Connecticut, Storrs, CT

Abstract Body:

Many people can achieve weight loss successfully. However, weight maintenance is a struggle, and most weight loss is regained within 5 years. Only few studies have examined whether genetic variants predict weight regain. Furthermore, studies tend to focus on weight and body mass index (BMI), but waist circumference (WC) also predicts cardiovascular morbidity and mortality. In this secondary data analysis of Look AHEAD we examine whether genetic risk for higher BMI or abdominal adiposity (WC or waist-to-hip ratio adjusted for BMI (WCadjBMI and WHRadjBMI)) predict BMI and waist circumference change after an initial 1-year weight loss. Look AHEAD is a randomized, controlled trial testing the impact of Intensive Lifestyle Intervention (ILI) compared to a control group on health outcomes among participants with overweight or obesity and type 2 diabetes. A subset of individuals randomized to ILI who lost at least 3% of their weight (White subset N= 709; Full sample N=1050) were included in this study. To assess the genetic risk, we used summary statistics for BMI, WHRadjBMI and WCadjBMI to construct genetic risk scores. Outcomes included the changes in BMI, WCadjBMI and WC from year 1 to years 2, 4 and 8. The models accounted for age, sex, baseline trait level, change in trait from baseline to year 1, and the 3 first genetic principal components for the full sample. Mean loss of BMI and WC in ILI were -3.22 and -7.70 cm at year 1; among ILI participants who lost 3% of their weight, mean changes in BMI and WC were 0.97 and 2.48cm at year 2, 1.92 and 4.88cm at year 4 and 2.04 and 6.67cm at year 8. In linear models, a genetic risk for increased WHRadjBMI predicted increased WCadjBMI in White ILI losers between years 1 and 2 ($\beta \pm SE = 1.02 \times 10^{-3} \pm 4.62 \times 10^{-4}$, $P = 0.027$), years 1 and 4 ($1.24 \times 10^{-3} \pm 5.16 \times 10^{-4}$, $P = 0.016$) and years 1 and 8 ($1.45 \times 10^{-3} \pm 5.33 \times 10^{-4}$, $P = 6.69 \times 10^{-3}$). The genetic risk score for WHRadjBMI also predicted WCadjBMI in the full sample for year 1 to 2 ($8.60 \times 10^{-4} \pm 3.75 \times 10^{-4}$, $P = 0.022$), year 1 to 4 ($1.10 \times 10^{-3} \pm 4.36 \times 10^{-4}$, $P = 0.012$) and year 1 to 8 ($1.50 \times 10^{-3} \pm 4.58 \times 10^{-4}$, $P = 1.09 \times 10^{-3}$). For WC alone, the WHRadjBMI genetic risk score was predictive of a WC change from year 1 to 2 in the White subset (0.0378 ± 0.0181 cm, $P = 0.037$). This translates to a 1.1cm difference in WC change from year 1 to 2 between groups stratified according to the 10th and 90th percentile of the WHRadjBMI GRS. These data show that a genetic risk for increased central adiposity is predictive for higher gain in waist circumference in years 1 to 8 after a 3% weight loss. The results imply that genetically predisposed subgroups may be at higher risk for WC increase and could benefit from additional intervention to maintain WC after a weight loss.

Complex Traits Posters - Wednesday

PB1404. Genetic Predictors of Memory Performance in Older Adults

Authors:

D. Archer^{1,2}, J. M. Eissman^{2,1}, S. Mukherjee³, M. Lee³, S-E. Choi³, P. Scollar³, E. H. Trittschuh³, J. B. Mez⁴, W. S. Bush⁵, C. D. Engelman⁶, Q. Lu⁷, D. W. Fardo⁸, K. F. Widaman⁹, R. F. Buckley¹⁰, E. C. Mormino¹¹, B. W. Kunkle¹², A. C. Naj¹³, L. R. Clark⁷, K. A. Gifford^{1,2}, Alzheimer's Disease Neuroimaging Initiative (ADNI), Alzheimer's Disease Genetics Consortium (ADGC), Alzheimer's Disease Sequencing Project (ADSP), M. Cuccaro¹⁴, C. Cruchaga¹⁵, M. A. Pericak-Vance¹², L. A. Farrer¹⁶, J. M. Vance¹², L-S. Wang¹³, G. D. Schellenberg¹⁷, J. L. Haines⁵, A. L. Jefferson^{1,2}, S. C. Johnson⁷, W. A. Kukull³, M. S. Albert¹⁸, C. D. Keene³, A. J. Saykin¹⁹, E. B. Larson³, R. A. Sperling¹⁰, R. Mayeux²⁰, P. M. Thompson²¹, E. R. Martin²², D. A. Bennett²³, L. L. Barnes²³, J. A. Schneider²³, P. K. Crane³, L. Dumitrescu^{1,2}, T. J. Hohman^{1,2}; ¹Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Vanderbilt Memory & Alzheimer's Ctr., Nashville, TN, ³Univ. of Washington, Seattle, WA, ⁴Boston Univ., Jamaica Plain, MA, ⁵Case Western Reserve Univ, Cleveland, OH, ⁶Univ. of Wisconsin Madison, Madison, WI, ⁷Univ. of Wisconsin-Madison, Madison, WI, ⁸Univ. of Kentucky, Lexington, KY, ⁹Univ. of California Riverside, Riverside, CA, ¹⁰Massachusetts Gen. Hosp., Boston, MA, ¹¹Stanford Univ., Stanford University, CA, ¹²Univ. of Miami, Miami, FL, ¹³Univ. of Pennsylvania, Philadelphia, PA, ¹⁴John P. Hussman Inst. for Human Genomics, Miami, FL, ¹⁵Washington Univ, St Louis, MO, ¹⁶Boston Univ Sch Med, Boston, MA, ¹⁷Univ Pennsylvania Sch Med, Philadelphia, PA, ¹⁸Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ¹⁹Indiana Univ Sch. of Med., Indianapolis, IN, ²⁰Columbia Univ, New York, NY, ²¹USC, Los Angeles, CA, ²²Univ. of Miami Miller Sch. of Med., Miami, FL, ²³Rush Univ. Med. Ctr., Chicago, IL

Abstract Body:

Longitudinal change in memory performance can serve as a powerful endophenotype of Alzheimer's disease (AD) that accurately reflects the continuum of cognitive decline that occurs over the course of aging and disease. In the present study, we investigate the genetic architecture of memory performance leveraging psychometric-based techniques to harmonize memory scores in 20,205 non-Hispanic White (NHW) and 2,612 non-Hispanic Black (NHB) participants across 4 longitudinal cohorts of aging and AD (ACT, ADNI, NACC, ROSMAP). Longitudinal memory scores were available in 14,877 and 1,853 in NHW and NHB participants, respectively. Genome-wide association studies (GWAS) of baseline memory performance and memory decline were performed using linear regression, covarying for age, sex, and the first five principal components to control for unmeasured population stratification. Analyses were run in each cohort/ancestry group separately, then subsequently meta-analyzed. For both memory outcomes, strong association signals were observed at 5 known AD loci, including *APOE* (rs429358, both $p < 1.19 \times 10^{-58}$), *BINI* (rs6733739, both $p < 2.18 \times 10^{-9}$), *PICALM* (rs3851179, both $p < 2.87 \times 10^{-7}$), *CRI* (rs4844610, $p = 3.63 \times 10^{-9}$) and *CD2AP* (rs9473117, $p = 3.98 \times 10^{-6}$). In addition to these known loci, we identified 2 novel genome-wide significant loci in the longitudinal memory decline analysis on chromosome 1 (rs2760073, intronic *C1orf50*, $p = 4.77 \times 10^{-8}$) and chromosome 12 (rs78378232, intergenic, $p = 3.47 \times 10^{-8}$). The chromosome 1 locus (rs2760073) is an eQTL for *CCDC23*, *ERMAP*, and *LEPRE1* in whole blood (eqtlgen.org) and these genes are involved in axon development, cell adhesion mediation, and collagen synthesis, respectively. This large analysis of memory performance in a population that includes the full clinical and preclinical spectrum of AD highlights the utility of sensitive longitudinal measures to confirm known loci and identify novel genomic signals. Expanding this approach to include more cohorts enriched for cognitively normal participants may provide the opportunity to deconvolve the genetic architecture of age-related cognitive decline from the genetic architecture of disease.

Complex Traits Posters - Thursday

PB1405. Genetic profiling and improved predictive capability with protein-based risk scores for IBD and disease progression using UK Biobank data.

Authors:

X. Zeng¹, P. Charuworn¹, V. Watson¹, D. Neasham², O. Archangelidi², D. Ellwanger¹, B. Pei³, Y-P. Lai³, I. Eres¹, A. Rampersaud³, S. Wang¹, G. Kafatos², B. Gibson¹, M. van Lookeren Campagne¹, Y-H. Hsu³; ¹Amgen, South San Francisco, CA, ²Amgen, Thousand Oaks, CA, ³Amgen, Cambridge, MA

Abstract Body:

Inflammatory bowel disease (IBD) is a broad term that describes conditions characterized by chronic inflammation of the GI tract. In 20 - 30% of patients, IBD progresses in extent and behavior over time; however, given the unpredictability of disease development at diagnosis, it is difficult to identify patients who will eventually progress to more extensive disease. Identifying patients at risk for progression would help personalize medical management, targeting earlier aggressive interventions. While recent large-scale genome-wide association studies (GWAS) have successfully identified risk loci for IBD, efforts to understand the genetic mechanism for IBD progression have been less fruitful due to the lack of proper data sources. With longitudinal clinical data in over 500K genotyped subjects, the UK Biobank (UKBB) presents a promising resource for revealing the genetic basis of IBD progression.

In this study, 8,109 clinically diagnosed IBD subjects were identified in the UKBB based on electronic health medical records. Occurrence of progression related outcomes (including IBD related surgeries, colon cancer, and primary sclerosing cholangitis) were used to define patients with IBD progression. We identified 2,451 IBD progressors and 5,658 non-progressors.

Two GWAS strategies with different underlying hypotheses were implemented - logistic and ordinal GWAS. Logistic GWAS comparing IBD progressors versus non-progressors identified 20 novel risk loci that have not been implicated in IBD manifestation. In parallel, to identify genetic variants associated with IBD severity, we stratified the cohort into healthy, non-progressors, and progressors and performed ordinal GWAS; this analysis identified 22 genome-wide significant risk loci, including 13 that were also found to be associated with IBD risk, suggesting that the pathophysiological mechanisms involved in IBD progression may be shared with that of IBD manifestation. Overall, our analysis suggests that IBD manifestation and progression have both shared and unique genetic underpinnings.

Further, protein-based risk scores were derived for both IBD and disease progression using proteomics data obtained from the UK Biobank Pharma Proteomics Project for over 54K participants. The scores derived from panel proteins selected by Lasso regression have the best discriminative capacity in the testing cohort for IBD and disease progression with an AUC of 0.76 and 0.64, respectively. For IBD, we show that a poly-omics risk score model incorporating both proteomics and genomics data (AUC = 0.78) outperforms models derived solely from either genomics data (AUC = 0.66) or proteomics data (AUC = 0.76).

Complex Traits Posters - Wednesday

PB1406. Genetic Regulation of Gene Expression in HIV+ T Cells Associated With Control of HIV spVL in African Populations

Authors:

R. Tough^{1,2}, S. N. Gingras^{1,2}, J. F. Tuff², C. M. Card^{3,1}, D. M. Tang², P. J. McLaren^{2,1}; ¹Dept. of Med. Microbiol. and Infectious Diseases, Univ. of Manitoba, Winnipeg, MB, Canada, ²Natl. Lab. for HIV Genetics at JC Wilt Infectious Diseases Res. Ctr., Natl. Microbiol. Lab., Publ. Hlth. Agency of Canada, Winnipeg, MB, Canada, ³Natl. Lab. for HIV Immunology at JC Wilt Infectious Diseases Res. Ctr., Natl. Microbiol. Lab., Publ. Hlth. Agency of Canada, Winnipeg, MB, Canada

Abstract Body:

HIV set-point viral load (spVL), the amount of virus in the blood during the chronic phase of infection, is a strong determinant of disease progression and transmission potential. Previous genome-wide association studies show that ~25% of the variance in HIV spVL in the infected population is attributable to common host genetic variants, mainly in the CCR5 and class I HLA genes. However, the majority of these studies only include individuals of European ancestry thus omitting a substantial proportion of global genetic diversity. To address this gap, we recently conducted a genome-wide association study of HIV spVL in 3,879 individuals of African Ancestry and observed a novel locus on chromosome 1 that is significantly associated with decreased HIV spVL (top SNP rs59784663, $p = 6.4E-10$, $\beta = -0.3$). Associated variants at this locus are unique to populations of African ancestry and the region has not been previously shown to impact HIV spVL in Europeans. The locus includes four protein-coding genes; PRKAB2, FMO5, CHD1L and BCL9, however, the precise causal variant(s), causal gene and/or causal pathway are unknown. To test the association between the top SNP and gene expression, we selected individuals that are homozygous reference (N=12) or heterozygous (N=12) at rs59784663 and performed transcriptional profiling of stimulated (anti-CD3/CD28) and unstimulated primary CD4+ T cells, the major cellular target of HIV infection. We did not observe a significant relationship between rs59784663 genotype and any of the genes in the region. However, we did observe a trend towards lower relative expression of CHD1L in stimulated CD4+ T cells and PRKAB2/FMO5 in unstimulated CD4+ T cells in individuals carrying the variant allele. In a transcriptome-wide analysis, we observed dysregulation of key signaling pathways; such as upregulation of AKT activation (FDR $P < 0.025$) and DNA double strand break sensing (FDR $P < 0.022$) in stimulated CD4+ T cells and upregulation of TNFR1-induced NF κ B signaling pathway (FDR $P < 0.005$) in unstimulated CD4+ T cells in heterozygous individuals. These data suggest that variants in the chromosome 1 region may be acting in trans to regulate expression of genes involved in HIV pathogenesis and have implications for design of novel anti-HIV therapeutics.

Complex Traits Posters - Thursday

PB1407. Genetic regulation of gene expression in the frontal cortex across the human lifespan

Authors:

H. Young¹, A. de Pins¹, R. Signer², C. Seah¹, A. Cote³, L. Huckins¹; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²Icahn Sch. of Med. at Mount Sinai, Fort Collins, CO, ³Icahn Sch. of Med., New York, NY

Abstract Body:

Background Genome-wide association studies have identified a large number of loci associated with psychiatric disorders by amassing huge sample sizes and compiling samples across a large number of contexts and environments. While great for discovery, this approach can obscure important context-specific impacts of genetic variation. Expression quantitative trait loci (eQTLs) provide an opportunity to identify genomic variants that regulate gene expression in tissues of interest, here the frontal cortex. Genetic regulation of gene expression can be highly context-dependent, and thus to obtain a more full picture of the regulatory landscape we seek to incorporate context into the search for eQTLs. Here, we extend this approach to aging in order to uncover associations of age-dynamic regulation and probe how that regulation is associated with neuropsychiatric disorders.

Methods Leveraging seven post mortem frontal cortex datasets that span a wide range of ages throughout adulthood into advanced aging, we collected matched genotypic and transcriptomic data for 1,811 individuals in order to identify age-specific and age-dynamic eQTLs. We conduct age-stratified eQTL searches (broken down into decade-sized age bins: 40-49, 50-59, ... 90+, N range = 195 - 425) as well as analyses to detect age*SNP interaction effects. Colocalization analyses and clustering values derived from cubic spline interpolation are used to identify age-dynamic eQTLs, and those eQTLs are tested for colocalization with neuropsychiatric traits.

Results We identify 7,734 eGenes with a significant age*SNP interaction effect. Age-stratified eQTL searches found 3,393 eGenes had a significant eQTL in only one of the decade-stratified analyses and many other sets of eGenes were identified with distinct set membership between all six age-stratified analyses (significant in 2, 3, etc. different age ranges), indicating age-specific regulation. K-means clustering of cubic spline interpolation allowed for the grouping of eQTLs based on their temporal trajectories into sets and identified two out of six clusters with age-dynamic eQTLs effect size trajectories across the lifespan.

Discussion Our findings show the benefit of considering age a useful investigative tool instead of treating it as a confounding variable that must be controlled for. The incorporation of environmental context can lead to discovery of dynamic genetic regulation that is key to understand the full genetic architecture of psychiatric disorders.

Complex Traits Posters - Wednesday

PB1408. Genetic susceptibility for autoimmune diseases and white blood cell count variability.

Authors:

N. Vaitinadin¹, C. M. Stein¹, J. Mosley², V. Kawai³; ¹Vanderbilt Univ Med. Ctr., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Vanderbilt Univ Med Ctr., Nashville, TN

Abstract Body:

Background: Fluctuations in white blood cell (WBC) counts is a common clinical characteristic in autoimmune (AI) conditions. Whether a genetic predisposition to an AI disorder associates with WBC counts in the general population is not known. We evaluated whether polygenic predictors for 7 common AI diseases are associated with WBC count variability. **Methods:** Publicly available genome-wide association study (GWAS) summary statistics were used to develop SNP-based genetic instruments for 7 common AI diseases. We tested the associations between each genetic instrument and measured WBC counts with a two-sample inverse variance weighted method (IVWM) and SNP weighted associations derived from a GWAS for WBC counts. Effect size in the IVWM analysis represents change in log[WBC counts] per change in log odds-ratio of the AI disease. For those AI diseases with significant associations in the IVWM analysis, a linear regression analysis using polygenic risk scores (PRS) as predictors was performed to test for associations with measured WBC counts in community-based (ARIC, n=8926) and medical-center derived cohorts (BioVU, n=40,461). The PRS analyses were also stratified by sex, as the incidence of AI diseases often varies by sex. **Results:** In the IVWM analysis, systemic lupus erythematosus (SLE) (Beta= -0.05 [95% CI, -0.06, -0.03]), multiple sclerosis (MS) (Beta= -0.05 [-0.07, -0.02]), and rheumatoid arthritis (RA) (Beta= 0.02 [0.01, 0.03]) were significantly associated with WBC counts. PRS for these diseases showed associations with measured WBC counts in ARIC and BioVU, following the same direction of effect as the IVWM analysis. In ARIC, a higher PRS was associated with lower log-transformed WBC count for SLE (beta -0.01, 95% CI: -0.017, -0.005 per SD increase in the PRS) and MS (-0.01, 95% CI: -0.017, -0.005). In BioVU, the association estimates for each AI PRS were very similar to those from the ARIC cohort, except that the RA association was significant. The WBC changes were larger among females, compared to males, consistent with known differences in sex-specific incidence rates. **Conclusions:** A genetic predisposition to SLE, RA, and MS was significantly associated with WBC counts. The gender differences in effect size suggest a non-constitutive association between genetic predisposition and WBC counts. This, combined with the fact that associations were observed in populations where prevalence of these AI diseases is expected to be low, suggests that subclinical AI disease may be more common than would be expected in the general population.

Complex Traits Posters - Thursday

PB1409. Genetic Variability of *CYP2B6* Polymorphisms (785A>G, 64C>T and 516G>T) in Southwest Nigerian Population: Implications for Malaria Treatment

Authors:

S. Ademola, O. Bamikole, A. Dibia, Y. S. Imo, J. Etim, A-L. Ajige, O. K. Amodu; Univ. of Ibadan, Ibadan, Nigeria

Abstract Body:

Malaria remains a disease of public health importance in spite of the efforts made over the years to reduce its burden globally. It is responsible for most deaths among under 5 children in sub-Saharan Africa. Although there was recorded gains in malaria reduction, the COVID-19 pandemic stalled the progress due to disruption in malaria management and treatment provision. The artemisinin combination therapy (ACT) is the recommended first line treatment for *Plasmodium falciparum* malaria in Nigeria. However, single nucleotide polymorphisms of the cytochrome P450 2B6 gene (known for the primary induction of artemisinin) can affect the enzyme function as it either enhances, decreases, or abolishes its enzymatic activities depending on the drug substrates. This study investigated the frequency of commonly observed *CYP2B6* variants (516G>T {rs3745274}, 785A>G {rs2279343}, 64C>T {rs2279343}) and the association with malaria in a Yoruba population of Nigeria. Genotyping of *CYP2B6* genetic polymorphisms was carried out among 200 children using PCR-RFLP protocol. Statistical package for social sciences (SPSS) was used in analyzing demographic and clinical data. Genotypic & allelic frequencies were calculated, conformity with Hardy Weinberg equilibrium and associations were tested for under three different genetic models using Plink v1.90 and haploview 4.2. We observed a minor allelic frequency (MAF) of 0.12, 0.375, 0.423 for 785A>G, 64C>T & 516G>T allelic variant respectively. The Three SNPs were not in linkage disequilibrium, SNP variants 785A>G and 516G>T were in conformity with Hardy Weinberg equilibrium. The genotype distributions of 785A>G and 64C>T differed significantly across all malaria groups. Further analysis showed a statistically significant association ($P < 0.01$; $OR < 1$) of 785 A>G with malaria under the additive and dominant models. This finding indicates that some cytochrome p450 gene variants may confer protection against malaria development in addition to its drug metabolism functions. Further large-scale studies on *CYP2B6* variants are recommended to explore its impact in malaria development or susceptibility.

Complex Traits Posters - Wednesday

PB1410. Genetic variants in Olfactory Response Genes associated with Osteoarthritis progression to Hip or Knee Replacement among African and Hispanic American Veterans.

Authors:

A. Nair^{1,2}, V. Srinivasasainagendra^{2,3}, H. K. Tiwari³, M. M. Bamman^{2,4,5}, J. Richman^{2,6}, J. Singh^{2,7,8}, M-L. McDonald^{1,2,7,9}; ¹Div. of Pulmonary, Allergy and Critical Care Med., Dept. of Med., Sch. of Med., Univ. of Alabama at Birmingham (UAB), Birmingham, AL, ²Birmingham VA Hlth.care System (BVAHS), Birmingham, AL, ³Dept. of Biostatistics, Sch. of Publ. Hlth., UAB, Birmingham, AL, ⁴Dept. of Cell, Dev.al, and Integrative Biology, Sch. of Med. (UAB), Birmingham, AL, ⁵Florida Inst. for Human & Machine Cognition, Pensacola, FL, ⁶Dept. of Surgery, Sch. of Med., UAB, Birmingham, AL, ⁷Dept. of Epidemiology, Sch. of Publ. Hlth., UAB, Birmingham, AL, ⁸Div. of Rheumatology and Clinical Immunology, Dept. of Med. at the Sch. of Med., UAB, Birmingham, AL, ⁹Dept. of Genetics, Sch. of Med., UAB, Birmingham, AL

Abstract Body:

Background/Aims: Osteoarthritis (OA) is a progressive disease marked with increasing pain and mobility limitations necessitating total hip or knee replacement (THR/TKR). Veterans who undergo THR/TKR experience higher comorbidity, poor functional status and higher health care utilization. We aimed to identify genetic variants associated with THR/TKR.

Methods: Ancestry stratified genome wide association scans (GWAS) were performed among European White (EW: N=56,466), African American (AA: N=16,876) and Hispanic (HIS: N=5,579) Veterans from the Million Veteran Program (MVP) adjusting for age, sex, BMI and population structure. ICD9/ICD10 codes were used to identify THR/TKR cases and two sets of age matched controls: 1) Healthy set: no OA history and 2) Possible progressive set: have OA but not THR/TKR.

Results: Using the healthy set for controls, no variant was significantly associated with THR/TKR at a level reaching genome-wide significance (GWS) among EUR or HIS Veterans. Two variants were significantly associated with THR/TKR among AA Veterans: 1) rs181133140 (OR=5.5, P=1.7x10⁻⁸, MAF=0.0006) intergenic to *OR52H1* and *OR52B6* genes which encode olfactory receptors 52H1 and 52B6; and 2) rs145338039 (OR=2.1, P=2.2x10⁻⁸, MAF=0.015) intronic to *FAM184A* gene which has been previously implicated in medulloblastoma. Using the possible progressive set as controls, several variants were significantly associated with THR/TKR among AA and HIS but not EUR Veterans. Among AA Veterans one variant in the UTR3 of *WFDC13* gene in the telomeric region of chromosome 20 and an intronic *HAOI* (Hydroxyacid Oxidase 1) variant were significantly associated. Among HIS Veterans a non-synonymous variant within *OR10G9* (olfactory receptor *10G9*) was associated with THR/TKR.

Conclusions: Variants in or near olfactory receptor genes were associated with THR/TKR in AFR and HIS Veterans. Olfactory loss is prevalent in aging populations often attributed to increased chronic diseases and medications. A direct link to OA or THR/TKR has not previously been reported. However, olfactory stimulation has been shown to mediate pain responses and progression to THR/TKR is largely motivated by increasing pain. Thus, it is possible that genetic variation in olfactory receptors influence pain perception in Veterans with OA increasing the necessity for THR/TKR.

Complex Traits Posters - Thursday

PB1411. Genetic variation in fatty acid amide hydrolase (*FAAH*) influences early drinking and smoking behaviors.

Authors:

A. Alsaafin¹, M. Chenoweth¹, M-P. Sylvestre², J. O'Loughlin², R. Tyndale³; ¹Univ. of Toronto, Toronto, ON, Canada, ²Université de Montréal, Montreal, QC, Canada, ³CAMH and Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Background: The endocannabinoid system plays a role in psychiatric disorders and drug dependence. Within the endocannabinoid system, genetically variable fatty acid amide hydrolase (FAAH) metabolizes endocannabinoids. Individuals with the C/A or A/A genotypes (A-group) of a common missense variant (rs324420; C>A; Pro129Thr) in *FAAH* have reduced enzyme activity compared to those with the C/C genotype (C-group). In animal studies, low FAAH activity is associated with alcohol and nicotine use, but in contrasting directions. Among human adults, those in the A- (vs. C-) group have an increased risk of harmful alcohol drinking (e.g., binge drinking days), but do not differ in risk for regular smoking or tobacco dependence. The associations between low FAAH activity (A-group) and early drinking and smoking behaviors during adolescence have not yet been investigated.

Methods: We studied European-ancestry participants from the Nicotine Dependence in Teens cohort. Adolescents were surveyed in a total of 20 cycles (from age ~12 to 17), then again during young adulthood (from age ~20 to 30). Hazard ratios (HR) were estimated to assess the association between genotype (C vs. A- group) and drinking initiation among baseline never-drinkers. Among those who initiated drinking, HR were estimated for attaining monthly, weekly, and daily drinking. The association between genotype and past-year binge drinking when participants were young adults was estimated using logistic regression. We estimated HR for smoking initiation according to genotype among baseline never-smokers, and from first inhalation to ICD-10 dependence among those who initiated smoking.

Results: Adolescents in the A- (vs. C-) group had a higher risk of drinking initiation (HR=1.39, 95% CI 1.09-1.77) and daily drinking (HR=2.24, 95% CI 1.05-4.76). As young adults, those in the A-group had higher odds of past-year binge drinking at ages 20 (odds ratio (OR)=2.18; 95% CI 1.38-3.45) and 30 (OR=1.66, 95% CI 1.13-2.45). An association between *FAAH* variation and smoking initiation was observed (HR=1.20, 95% CI 1.04-1.39). In contrast to early alcohol behaviors, those in the A- (vs. C-) group had generally slower or similar rates to early smoking outcomes (HR's range from 0.43 to 1.15). Sensitivity analyses showed robust findings when examining the three genotype groups separately (i.e., C/C vs. C/A vs. A/A).

Conclusions: Low FAAH activity (i.e., A-group) is a risk factor for drinking and smoking initiation and in the development of early harmful drinking. In contrast, low FAAH activity may protect against early smoking outcomes. *FAAH* variation appears to alter susceptibility to drug use.

Complex Traits Posters - Wednesday

PB1412. Genetically Informed Prediction of Short Term Parkinson's Disease Progression

Authors:

H. Javedani Sadaei¹, A. Cordova-Palomera², L. Jonghun², J. Padmanabhan², C. Shang-Fu¹, N. Wineinger¹, R. Dias¹, D. Prilutsky², S. Szalma², A. Torkamani¹; ¹Scripps Res. Translational Inst., La Jolla, CA, ²Takeda Dev. Ctr. Americas, San Diego, CA

Abstract Body:

Background: Parkinson's disease (PD) treatments modify disease symptoms but do not halt progression, characterized by slow and varied motor and non-motor changes overtime. Variation in PD progression hampers clinical research, resulting in long and expensive clinical trials prone to failure. Objective: Development of models for short-term PD progression prediction, which could be useful for informing individual-level disease management, and shortening the time required to detect disease-modifying drug effects in clinical studies. Methods: PD progressors were defined by an increase in MDS-UPDRS scores at 12-, 24-, and 36- months post-baseline. Using only baseline features, PD progression was separately predicted across all timepoints and MDS-UPDRS subparts in an optimized XGBoost framework. These predictions were combined into a meta-predictor for 12-month MDS UPDRS Total progression. Data from the Parkinson's Progression Markers Initiative (PPMI) were used for training with independent testing on the Parkinson's Disease Biomarkers Program (PDPB) cohort. Results: 12-month PD total progression was predicted with an F-measure 0.77, ROC AUC of 0.77, and PR AUC of 0.76 when tested on a hold-out PPMI set. When tested on PDPB we achieve a F-measure 0.75, ROC AUC of 0.74, and PR AUC of 0.73. Exclusion of genetic predictors led to the greatest loss in predictive accuracy; ROC AUC of 0.66, PR AUC of 0.66-0.68 for both PPMI and PDPB testing. Conclusions: Short-term PD progression can be predicted with a combination of survey-based, neuroimaging, physician examination, and genetic predictors. Dissection of the interplay between genetic risk, motor symptoms, non-motor symptoms, and longer-term expected rates of progression enable generalizable predictions.

Complex Traits Posters - Thursday

PB1413. Genetically-associated body-mass index, vitamin D levels, and age at puberty are not associated with Parkinson's Disease related phenotypes.

Authors:

E. Misicka, F. Briggs; Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Abstract Body:

Background: Parkinson Disease (PD) is a neurodegenerative disorder affecting 1 million persons in the United States. There is an urgent need to characterize the biological mechanisms underlying PD onset and presentation as prevalence is expected to increase by 20% by 2030. Robustly investigating the effects of potential risk factors can be difficult due to the limited availability of epidemiologic data (i.e., there are conflicting findings for body mass index [BMI] and vitamin D [vitD] on PD risk). Fortunately, genetic instrumental variable (GIV) analyses can examine genetically driven causal relationships that may be confounded in epidemiologic studies. **Objective:** To investigate the relationships for factors associated with risk and presentation of other common neurodegenerative diseases and multiple PD phenotypes using Mendelian randomization (MR). **Methods:** MR is a form of GIV analysis where genome-wide association (GWA) summary statistics are utilized to investigate the relationship between exposure and outcome. Genetic variants associated with the exposure serve as GIVs. GIVs were selected from GWA studies for BMI (n=681,275), serum vitD levels (n=79,366), and age at puberty (AAP; n=329,245). GWA studies were also selected for PD phenotypes as outcomes, including risk (n=1.34×10⁶), age at onset (AAO; n=28,568), and binary and ratio measures of PD motor phenotype (MP, MPR; n=3,212). For each exposure, GIVs were clumped at a linkage disequilibrium threshold of r²<0.05 using the 1000 Genomes EUR Reference Panel. Inverse variance weighted (IVW) meta-analysis was performed for each exposure on each outcome, alongside other MR approaches. Significant associations were investigated for horizontal pleiotropy using the MR-Egger intercept test. **Results:** Neither VitD (GIVs=6) nor AAP (GIVs=377) were significantly associated with any of the PD phenotypes (p>0.05). However, BMI (GIVs=656) was significantly associated with PD AAO under IVW meta-analysis (β=-0.87, p=0.018). A significance test of the MR-Egger intercept for the BMI-PD AAO association revealed that there were horizontally pleiotropic effects present (p=0.044). **Conclusions:** MR analyses do not implicate genetically-driven BMI or vitD as risk factors for PD, which adds resolution to conflicting epidemiological studies. AAP was also not associated with PD risk, which is in contrast to the relationship between AAP and multiple sclerosis risk. BMI, vitD and AAP were also not associated with PD presentation (MP, MPR, or AAO); however, an association between BMI and PD AAO appeared to be influenced by horizontal pleiotropy, and merits further investigation.

Complex Traits Posters - Wednesday

PB1414. Genetics determinants of Renin-Angiotensin-Aldosterone System (RAAS) in the Cooperative Health Research In South Tyrol study

Authors:

L. Foco¹, M. W. Arisido¹, E. Koenig¹, M-C. Zennaro², M. Poglitsch³, P. P. Pramstaller¹, C. Fuchsberger¹, C. Pattaro¹; ¹Inst. for Biomedicine, Eurac Res., Bolzano, Italy, ²Université Paris-Cité, PARCC, Inserm, Paris, France, ³Attoquant Diagnostics GmbH, Vienna, Austria

Abstract Body:

Background The Renin-Angiotensin Aldosterone system (RAAS) is the major player in the long-term blood pressure (BP) regulation. Following cleavage of angiotensinogen by renin, the resulting angiotensin I peptide is further converted to angiotensin II by the Angiotensin Converting Enzyme (ACE). Angiotensin II can then bind to its type 1 receptor, stimulating the production of aldosterone, a steroid hormone inducing sodium and water retention, thus raising BP. The RAAS is targeted by commonly used anti-hypertensive drugs, but despite its importance, little is still known about its genetics, partially due to the technical difficulties in the quantification of the involved biomarkers. **Methods** Using mass-spectrometry, we simultaneously quantified in serum equilibrium angiotensin I (eqAngI), angiotensin II (eqAngII) and aldosterone (Aldo) on 2105 participants to the Cooperative Health Research In South Tyrol study (CHRIS), who were either normotensive or under documented anti-hypertensive drug (AHD) treatment. We derived proxies of renin activity ($PRA-S=eqAngI+eqAngII$), angiotensin-converting enzyme activity ($ACE-S=eqAngII/eqAngI$) and adrenal gland function ($AA2-ratio=Aldo/eqAngII$). We tested association of these biomarkers with 1,085,897 SNPs imputed against a population specific whole-exome-sequencing-based panel, adjusting for age, sex, AHD, and assuming an additive genetic model, using Regenie v 3.1.1. **Results** ACE activity (ACE-S) was associated with variants spanning the ACE gene. The top association was seen at rs4363 ($p=2.8E-18$), already associated with serum ACE level, and we replicated the association with rs4343 ($p=6.7E-18$) and ACE activity described in an Asian population. We found strong evidence of association of eqAngII and adrenal function (AA2-ratio) with the missense/stop gained rs27044 in the ubiquitous endoplasmic reticulum aminopeptidase 1 (ERAP1, $p=8.2E-10$), involved in angiotensin II degradation. We observed a borderline association between aldosterone and the synonymous/non-coding transcript exon rs7275 at SMARCA4/LDLR ($p=5.9E-8$), previously related to lipid metabolism and coronary artery disease. **Conclusions** While some results await replication in an independent cohort, findings are consistent with known physiology, and shed light on undescribed mechanisms possibly involved in RAAS biology.

Complex Traits Posters - Wednesday

PB1415. Genetics of Gulf War illness, a genome wide association study

Authors:

J. Vahey^{1,2}, X. Qin², A. Stone³, W. C. Carter³, L. M. Griffin³, S. Pyarajan⁴, G. Turner⁴, C. Stafford², E. J. Gifford^{2,5}, K. J. Sims², C. Williams², E. Hauser^{6,2}; ¹Duke Univ., Durham, NC, ²Durham VA Med. Ctr., Durham, NC, ³Central Arkansas Veterans Hlth.care System, Little Rock, AR, ⁴VA Boston Hlth.care System, Boston, MA, ⁵Duke Margolis Ctr. for Hlth.Policy, Durham, NC, ⁶Duke Univ Med Ctr, Durham, NC

Abstract Body:

GWII is a set of debilitating symptoms afflicting 25-35% of US Veterans deployed in the 1990-91 Gulf War. The cause is unknown but is believed to be a complex mixture of military exposures and individual susceptibility. In the 30+ years since the war, a diagnostic definition, biomarkers and treatments remain elusive. Veterans continue to suffer major decrements to quality of life. The GWIECB was developed to perform genome-wide genetic analysis to explore underlying genetic susceptibility to GWII with the hope that enriched genes and associated genotypes may eventually help identify new biomarkers or treatments. Gifford et al. (Life Sciences 2021) applied the recommended GWII research definitions to the GWIECB. Based on the consistent association with deployment, the CDC Severe GWII case definition was chosen as the primary case definition for the analysis. Genotyping was done using Illumina Omni2.5-8v1.4 using standard protocols. Plink 1.9 was used for genetic data cleaning and analysis of 1,777,976 SNPs. GWAS was completed on 1,061 Veterans, using a logistic regression model, adjusting for age, sex, and ten genetic principal components (PCs) with 247 CDC Severe GWII cases (23.6%). Pathway analysis was performed using gene and gene set association, as implemented in MAGMA 1.6b, using the individual level data. Genome-wide significance was not reached by any gene in the association analysis. The top 10 genes in the gene-based association were SLC22A4, LANCL2, FAM172A, ZNF17, DPT, EPB41L1, ZNF516, ZNF804B, SCFD1, and ZNF394. SLC22A4 is an integral membrane transport protein that transports small organic cations. The top ten ranked gene sets reproduce prior work and are consistent with findings related to exposures thought to be implicated in GWII. The top result, response to cadmium ion (beta=0.43, p=2.36×10⁻⁴) suggests genetic susceptibility to heavy metals as a contributing factor. The 2nd and 3rd pathways, regulation of response to interferon gamma (beta=0.62, p=3.08×10⁻⁴) and regulation of autophagosome maturation (beta=0.81, p=3.35×10⁻⁴), reinforce neuroinflammation and neurodegeneration pathways in GWII. This unbiased genetic analysis of CDC Severe GWII suggests the need to incorporate environmental exposures in genetic models of GWII.

Complex Traits Posters - Wednesday

PB1416. *Genetics of ventricular septal defects: A novel genetic interaction between Sox7 and Wnt4 is associated with abnormal endocardial cushion morphogenesis*

Authors:

A. Hernandez-Garcia¹, K. Pendleton¹, S. Kim¹, Y. Li¹, K. Bum Jun¹, H. Zaveri¹, V. Jordan², C. Ljungberg^{3,4}, R. Chen¹, R. Lanz¹, D. Scott^{5,6}; ¹Baylor Coll. of Med., Houston, TX, ²Dept. of Molecular Physiology and Biophysics, Baylor Coll. of Med., Houston, TX, ³Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, ⁴Jan and Dan Duncan Neurological Res. Ctr. at Texas Children's Hosp., Houston, TX, ⁵Baylor Coll. Med, Houston, TX, ⁶Dept. of Molecular Physiology and Biophysics, Baylor Coll. of Med., Houston, TX

Abstract Body:

SOX7 is located in a region on chromosome 8p23.1 that is recurrently deleted in individuals with septal defects. *Sox7*^{-/-} embryos die of heart failure around E11.5 due to defects in vascular remodeling. These embryos have hypocellular endocardial cushions with severely reduced numbers of mesenchymal cells. We also observed a ventricular septal defect (VSD) in a rare *Sox7*^{lox/-};Tie2-Cre embryo that escaped early lethality. This led us to hypothesize that SOX7 plays a critical developmental role in the endocardium of the atrioventricular (AV) canal. We subsequently used AV explant studies to show that SOX7 deficiency leads to a severe reduction in endocardial-to-mesenchymal transition (EndMT). Since SOX7 is a transcription factor, we hypothesized that it functions in the endocardium by regulating the expression of EndMT-related genes. To identify these genes in an unbiased manner, we performed RNA-seq on pooled E9.5 hearts tubes harvested from *Sox7*^{-/-} embryos and their wild-type littermates. We found that *Wnt4* transcript levels were severely reduced, which we confirmed by RNA in situ hybridization. This was of particular interest since WNT4-deficiency causes VSDs in humans. We went on to perform *in silico* analyses that suggested that multiple SOX family binding sites exist in the predicted *Wnt4* promoter. We and others have proved that WNT4 is expressed in the endocardium and promotes EndMT by acting in a paracrine manner to increase the expression of BMP2 in the myocardium. Consistent with these findings, we found that *Bmp2* transcript levels were diminished in *Sox7*^{-/-} embryonic hearts. To verify if *Wnt4* interacts with *Sox7* to modulate EndMT process during the endocardial cushion development, we determined the density of mesenchymal cells in the endocardial cushions of E10.5 mouse embryos. Consistent with our hypothesis, *Wnt4*^{+/-}; *Sox7*^{+/-} double heterozygous had reduced numbers of mesenchymal cells in their endocardial cushions compared to controls, due to decreased levels of EndMT and mesenchymal cell proliferation. We conclude that SOX7 promotes EndMT in the developing AV canal by positively regulating *Wnt4* and *Bmp2* expression. Additionally, we have novel evidence that *Sox7* interacts genetically with *Wnt4* to modulate EndMT and mesenchymal cell proliferation in the endocardial cushions of the developing AV canal. These data also provide additional evidence that haploinsufficiency of *SOX7* contributes to the congenital heart defects seen in individuals with recurrent 8p23.1 microdeletions.

Complex Traits Posters - Thursday

PB1417. Genome scans of early childhood caries implicate bitter taste receptors

Authors:

E. Orlova¹, T. Dudding², J. M. Chernus³, R. N. Alotaibi⁴, S. Haworth², R. J. Crout⁵, M. Lee³, N. Mukhopadhyay⁶, E. Feingold¹, S. M. Levy⁷, D. W. McNeil⁵, B. Foxman⁸, R. J. Weyant³, N. J. Timpson⁹, M. L. Marazita¹⁰, J. R. Shaffer¹; ¹Univ of Pittsburgh, Pittsburgh, PA, ²Univ Bristol, Bristol, United Kingdom, ³Univ. of Pittsburgh, Pittsburgh, PA, ⁴King Saud Univ., Riyadh, Saudi Arabia, ⁵West Virginia Univ., Morgantown, WV, ⁶Univ Pittsburgh/Sch Dental Med, Pittsburgh, PA, ⁷Univ. of Iowa Coll. of Dentistry, Iowa City, IA, ⁸Univ. of Michigan, Ann Arbor, MI, ⁹Bristol Univ., Bristol, United Kingdom, ¹⁰Univ. of Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA

Abstract Body:

Although genetic risk factors are known to affect susceptibility to early childhood caries (ECC), few studies have focused on finding genetic determinants of ECC. Here we performed a meta-analysis of genome-wide association studies (GWAS) testing ~3.9 million genetic variants in five cohorts of children (aged up to 6 years, total N=2974, cohorts: COHRA1, COHRA2, Iowa Fluoride Study, Iowa Head Start, ALSPAC) using Stouffer's method, aiming to identify genes with potential roles in ECC biology. There was suggestive evidence for association (p -value $< 1 \times 10^{-5}$) at genetic regions and gene families previously associated with caries in primary and permanent dentition. There was also suggestive evidence for association with genetic regions related to tooth morphology, immune response to bacteria, nociception, periodontal disease, and others. We then integrated the meta-analysis results with gene expression data in a transcriptome-wide association study (TWAS). This approach identified 6 genes whose genetically predicted expression was associated with ECC (p -values $< 3.09 \times 10^{-6}$; *LINC02905*, *CDH17*, *TAS2R43*, *SMIM10L1*, *TAS2R14*, *NRAD1*). Some of the strongest associations were with genes encoding members of the bitter taste receptor family (TAS2R); other members of this family have previously been associated with caries in primary, mixed and permanent dentition. Of note, we identified the receptor encoded by *TAS2R14* (z -score 0.47; p -value = 1.06×10^{-6}), which stimulates rapid innate immunity and anti-microbial defense in response to molecules released by cariogenic bacteria *S. mutans* and *S. aureus*. The other gene transcripts identified by the TWAS did not have known biological relationships with caries. The transcript *CDH17* (z -score 1.74; p -value = 3.32×10^{-8}) stands out as one of the most significant and reliable associations. *CDH17* influences the permeability of the intestinal epithelium. These findings provide insight into ECC genetic architecture, underscore the importance of host-microbial interaction in caries risk, and identify novel risk genes for follow-up study.

Complex Traits Posters - Wednesday

PB1418. Genome sequencing in brain samples from the ROSMAP cohort to examine the relationship between telomere length, mitochondrial copy number, and β -amyloid

Authors:

M. Lynch¹, M. Taub², J. Farfel³, J. Yang³, D. Bennet³, R. Mathias⁴; ¹Johns Hopkins Univ., Baltimore, MD, ²Johns Hopkins, Baltimore, MD, ³Rush Med. Coll., Chicago, IL, ⁴Johns Hopkins Univ, Baltimore, MD

Abstract Body:

Telomere length (TL) changes regulation and mitochondrial DNA (mtDNA) alterations are well established hallmarks of aging and may have associations with cognitive phenotypes. Little is known about the relationship between TL and β -amyloid, especially in brain samples. Previously, the role of TL in Alzheimer's Disease (AD) has been studied in small sample size groups using leukocyte samples and few studies investigated AD related phenotypes including β -amyloid.

We examined N=256 DLPFC samples extracted from human aged (range: 71.3-108.2, mean: 88.6 years old) and AD brains from The Religious Orders Study (ROS), and the Rush Memory and Aging Project (MAP). In this sample, 43.4% were diagnosed with dementia and 28.5% with MCI. TL was estimated from whole genome sequencing data (WGS) by Telseq which has high correlation with Southern blot methods (Taub et al Cell Genomics 2022). Raw mtDNAcn estimates were calculated as $(covmt/covnuc) \times 2$ using mtDNA and nuclear DNA from the WGS data (Klein et al Molecular Neurodegeneration 2021).

We found TL and mtDNAcn measures to be positively correlated ($R=0.13$, $p=0.041$), even after adjustment for age and sex ($R=0.13$, $p=0.031$). While not significant, we found age of the samples to be negatively correlated with TL (~ 11 bp decrease/year, $p=0.29$) and males to have shorter TL than females (106bp shorter, $p=0.47$), consistent with prior knowledge. We did not find mtDNAcn to be significantly associated with age of sample (13.4 increase/year, $p=0.19$) or sex (92.2 higher in males, $p=0.53$). Previous studies have reported mixed results on the effect of aging and sex on mtDNAcn in brain samples.

In independent linear regression models, we identified a significant negative association ($\beta=-0.23$, $p=0.012$) between mtDNAcn and β -amyloid levels and between TL and β -amyloid levels ($\beta=-0.15$, $p=0.023$). Additionally, in a multivariate regression, we found both brain mtDNAcn and TL to be independently related to β -amyloid. Higher β -amyloid is associated with lower mtDNA copy number ($\beta=-0.21$, $p=0.024$), and lower telomere length ($\beta=-0.13$, $p=0.049$). We found these associations to be robust to cell composition; results did not change after adjusting for brain cell composition estimated from these same samples using bulk RNASequencing data.

These results suggest that higher β -amyloid levels are related to lower mtDNAcn and shorter TL in the DLPFC, and therefore that mtDNAcn and TL may be involved in AD pathogenesis. These findings suggest for the first time that the two genomic measures of aging in brain tissue (mtDNAcn and TL) may be independently related to AD pathogenesis.

Complex Traits Posters - Thursday

PB1419. Genome wide association study of clinically predicted suicide liability.

Authors:

A. Jespersen, O. Plana-Ripoll, J. Steinbach, C. Albiñana, J. McGrath, P. Mortensen, F. Privé, E. Agerbo, B. Vilhjálmsón; Aarhus Univ., Aarhus, Denmark

Abstract Body:

Suicide accounts for 1 in every 100 deaths worldwide, resulting in more than 700,000 deaths annually. Globally the number of suicides has decreased by 36% in the last 20 years whilst the Americas have seen a 17% increase in deaths by suicide. Although there is considerable evidence that attempted suicide and death by suicide have distinct aetiologies, they share risk factors, which include previous suicide attempt, depression, and alcohol use disorders.

Heritability estimates of suicidal behaviour range from 17% to 55% suggesting a significant genetic component to suicide risk. However, as suicides are fortunately relatively uncommon, genetic studies of suicidal behaviour suffer from insufficient number of genotyped cases and are generally underpowered. In this study, we aim to address this limitation by leveraging Danish electronic health care records to perform a GWAS of clinically predicted suicide liability. We achieve this by using health registers containing all hospital admissions, diagnoses, and prescriptions of the Danish population, to clinically predict death by suicide. In order to do this, we will first use a Danish population sample comprising 323,000 individuals, 869 of whom died by suicide, to train a penalised regression. The resulting weights will be used to clinically predict suicide liability in an independent sample of 134,000 genotyped individuals. We then perform a GWAS of the clinically predicted suicide liability and compare the genetic overlap with the most recent case-control GWAS of death by suicide (Docherty et al. 2020), as well as genetic correlations between clinically predicted suicide liability and major psychiatric disorders. We expect results from this study to shed light on the contribution of common genetic variants to the clinically predicted risk of death by suicide. The study will demonstrate the level of genetic similarity between clinically predicted death by suicide and observed death by suicide. Finally, this method of leveraging electronic health records for clinical prediction to increase power in genetic studies could pave the way for genetic studies of phenotypes otherwise suffering from low numbers of genotyped cases.

Complex Traits Posters - Wednesday

PB1420. Genome-wide analysis in over 1 million individuals reveals over 2,000 independent genetic signals for blood pressure

Authors:

J. Keaton¹, Z. Kamali², A. Vaez³, T. Xie⁴, A. Ani⁴, E. Evangelou⁵, J. Hellwege⁶, J. Denny⁷, D. Levy⁸, T. Edwards⁶, P. Munroe⁹, H. Snieder⁴, H. Warren¹⁰, Million Veterans Program, Lifelines Cohort Study, ICBP Consortium; ¹NIH/NHGRI, Bethesda, MD, ²Isfahan Univ. of Med. Sci. (IUMS), Isfahan, Iran, Islamic Republic of, ³Univ. of Groningen, Groningen, Groningen, Netherlands, ⁴Univ. of Groningen, Groningen, Netherlands, ⁵Imperial Coll. London, London, United Kingdom, ⁶Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁷NIH, Kensington, MD, ⁸NHLBI/NIH, Framingham, MA, ⁹Queen Mary Univ. of London, London, United Kingdom, ¹⁰William Harvey Res. Inst., Barts and The London Sch. of Med., QMUL, London, United Kingdom

Abstract Body:

Hypertension is a leading cause of premature death affecting more than a billion individuals worldwide. Here we report on the genetic determinants of blood pressure (BP) traits (systolic, diastolic, and pulse pressure) in the largest single-stage genome-wide analysis to date (N=1,028,980 European-descent individuals). We identified 2,103 independent genetic signals ($P < 5 \times 10^{-8}$) for BP traits, including 113 novel loci. These associations explain ~40% of common SNP heritability of systolic and diastolic BP. Comparison of top versus bottom deciles of polygenic risk scores (PRS) based on these results reveal clinically meaningful differences in BP (12.9 mm Hg for systolic BP, 95% CI 11.5-14.2 mm Hg, $p = 9.08 \times 10^{-73}$) and hypertension risk (OR 5.41; 95% CI 4.12 to 7.10; $P = 9.71 \times 10^{-33}$) in an independent dataset. Compared with the area under the curve (AUC) for hypertension discrimination for a model with sex, age, BMI, and genetic ancestry, adding systolic and diastolic BP PRS increased discrimination from 0.791 (95% CI = 0.781-0.801) to 0.814 (95% CI = 0.805-0.824, delta AUC=0.023, $P = 2.27 \times 10^{-22}$). Our transcriptome-wide association study detected 2,793 BP colocalized associations with genetically-predicted expression of 1,070 genes in five cardiovascular tissues, of which 500 are previously unreported for BP traits. These findings represent an advance in our understanding of hypertension and highlight the role of increasingly large genomic studies for development of more accurate PRS, which may inform precision health research.

Complex Traits Posters - Thursday

PB1421*. Genome-wide association analyses of sepsis and septic shock in a large practice-based biobank

Authors:

L. Jiang¹, A. Oeser¹, A. Dickson², L. Daniel², N. Cox³, R. Tao³, C. Chung², W-Q. Wei⁴, M. Stein², Q. FENG⁵; ¹Vanderbilt Med. Ctr., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Vanderbilt Univ Med Ctr., Nashville, TN, ⁴Vanderbilt Univ, Nashville, TN, ⁵Vanderbilt Univ. Med. Ctr.I, Nashville, TN

Abstract Body:

Introduction: Sepsis accounts for one in every three in-hospital deaths and for 13% of hospital costs in the U.S. Classic twin studies of infections and recent genetic studies on COVID19 suggest that host genetics contribute to sepsis. However, we know little about genetic predictors of sepsis, especially in patients with infections. In the current study, we leveraged the de-identified electronic health records (EHRs) linked to DNA samples in the biobank at Vanderbilt (BioVU). We conducted GWAS on the risk of (1) sepsis and (2) septic shock in patients hospitalized with infections.

Methods: We identified 11,989 White, and 2,096 Black adults who (1) were admitted to Vanderbilt University Medical Center with an infection requiring antibiotics and (2) had genome-wide genotyping available. Sepsis was defined by the Sepsis-3 criteria of concurrent infection and organ dysfunction using diagnosis codes (ICD), procedure codes, lab tests, and keywords from the chart. Septic shock was defined by the ICD codes for septic shock/severe sepsis (995.92, 785.52, R65.20, and R65.21). We tested associations between genotypes and risk of sepsis or septic shock using logistic regression and adjustment for sex, age, age², and 10 PCs of ancestry to account for residual population structure.

Results: During the hospital stay, 3,546 White and 521 Black patients developed sepsis, including 1386 and 152 with septic shock, respectively. Among patients who developed sepsis, there were fewer women (46.5%) than men, and Black patients were younger (51.8±16.9 years) than White patients (58.2±15.9 years). In GWAS analyses, rs77960372-A was associated with an increased risk of septic shock in White patients (upstream of CD166; odds ratio [OR], 2.98; p=4.6×10⁻⁸); CD166 is a member of a subfamily of immunoglobulin receptors and it may be involved in inflammation through its regulation of members of the NF-κB family. There was a suggestive association with an increased risk of septic shock in Black patients for rs111587601-A (close to DYRK2; OR, 3.7; p=6.9×10⁻⁸); DYRK2 was previously reported to be part of an expression signature of 7 genes associated with outcomes in patients with sepsis.

Conclusion: In a GWAS on the risk of sepsis and septic shock in patients hospitalized with infections we identified two potential genetic signals close to CD166 and DYRK2 genes.

Complex Traits Posters - Wednesday

PB1422. Genome-wide association analysis of composite sleep scores in 413,904 individuals

Authors:

H. Wang¹, M. O. Goodman¹, J. M. Lane¹, H. S. Dashti², J. Chung¹, T. Sofer¹, S. M. Purcell¹, X. Zhu³, M. K. Rutter⁴, S. Redline¹, R. Saxena²; ¹Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, ²Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, ³Case Western Reserve Univ., Cleveland, OH, ⁴Manchester Univ., Manchester, United Kingdom

Abstract Body:

Recent genome-wide association studies (GWAS) of self-reported sleep traits have identified hundreds of genetic loci, often in genes and biological pathways that are shared across sleep phenotypes. Modeling the multiple dimensions of sleep may allow identification of common underlying genetic mechanisms and complement analyses of individual traits. Moreover, accounting for phenotypic and genetic correlations between traits has the potential to enhance signal in noisy measurements. To help identify the biological bases common to multiple sleep characteristics, we performed GWAS of two composite sleep scores derived from sleep questionnaires characterizing 413,904 individuals of European ancestry from the UK Biobank. Questionnaire items included information on self-reported sleep duration, chronotype, insomnia, snoring, and daytime sleepiness. We constructed an additive sleep score as a sum of up to five favorable sleep behaviors (having sleep duration of 7-8 hours, early chronotype, few insomnia symptoms, no snoring, and no excessive daytime sleepiness). We also performed principal component (PC) analysis to extract a composite sleep score (PC1) as the weighted sum explaining the highest proportion of the combined phenotypic variance. Higher PC1 value is interpretable as a sleep deficiency phenotype (shorter sleep duration, frequent insomnia symptoms, and excessive daytime sleepiness). Separate GWAS of additive and PC1 sleep scores identified, respectively, 7 and 15 novel loci, containing genes that function in nervous system development (*CNTN3*), mTORC1 signaling (*FNIP2*), and immune response pathways (*IGSF21* and *KLRG1*). Both composite sleep approaches revealed novel genetic mechanism of related sleep behaviors, while retaining clinical and biological interpretability. In the next step, we will construct an analogous heritability-maximizing composite sleep score, and study the genetic overlap between composite sleep scores with cardiometabolic, inflammatory, and psychiatric traits using genetic correlation and Mendelian randomization analyses.

Complex Traits Posters - Thursday

PB1423. Genome-wide association meta-analysis and burden of rare coding variants in African American patients with Acne

Authors:

A. Saeidian¹, H-Q. Qu², M. March², Y. Liu³, J. Glessner², H. Hakonarson⁴; ¹Ctr. for Applied Genomics, Children's Hosp. of Philadelphia, Philadelphia, PA, ²Children s Hosp. of Philadelphia, Philadelphia, PA, ³The children's Hosp. of philadelphia CHOP, Philadelphia, PA, ⁴Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

While Acne is mostly a disorder of adolescence, it persists into middle age in a significant number of individuals. A broad range of evidence primarily from twin studies suggests that acne is a disorder of genetic susceptibility. Acne consists of a broad spectrum of diseases varying in the age of onset and resolution and in severity, type, and distribution of lesions that primarily impact facial skin. Here, we perform a GWAS meta-analysis and evaluated the aggregate contribution of predicted deleterious low-frequency and rare variants in the minority African American (AA) population. We performed a GWAS meta-analysis on blood-derived DNA of 4853 AA individuals, including 2811 males, 2011 females, and 22578 AA controls without acne. The cases and the controls are matched in age of diagnosis and sex. The samples are extensively stratified into sub-cohorts by coexisting obesity, polycystic ovary disease, a steroid medication, and specific combinations of >1 comorbid conditions. For rare variants, SNP-set Kernel Association Test (SKAT) was done in stratified sub-cohorts of a total of 5000 AA cases and controls sequenced by whole genome sequencing (WGS). We identified several candidate genes with the strongest enrichment ($P < 5E-08$ in GWAS and $P < 1E-06$ in WGS, respectively), results that are currently being replicated in independent samples of European populations. Further experimental study is being designed to clarify the functional mechanisms of the genes in acne, pending for validation results in European.

Complex Traits Posters - Wednesday

PB1424. Genome-wide association meta-analysis reveals 99 risk loci for pain susceptibility and pleiotropic relationships with psychiatric, metabolic, and immunological traits.

Authors:

E. Mocchi^{1,2,3}, K. Ward¹, S. Dorsey^{1,3}, S. Ament^{2,4}; ¹Dept. of Pain & Translational Symptom Sci., Univ. of Baltimore, Sch. of Nursing, Baltimore, MD, ²Inst. for Genome Sci., Sch. of Med., Univ. of Maryland, Baltimore, MD, ³Ctr. to Advance Chronic Pain Res. (CACPR), Univ. of Maryland Baltimore, Baltimore MD, Baltimore, MD, ⁴Dept. of Psychiatry, Univ. of Maryland Sch. of Med., Baltimore, MD

Abstract Body:

Chronic pain is at epidemic proportions in the United States, represents a significant burden on our public health system and is coincident with a growing opioid crisis. There are substantial individual differences in pain susceptibility, with patients who have seemingly identical injuries often reporting different pain intensities, likely explained by genetic factors.

We conducted a meta-analysis of genome-wide association study (GWAS) summary statistics from seventeen pain susceptibility traits in the UK Biobank. The selected traits included eight phenotypes describing types of pain experienced in the last month (headache, neck/shoulder, back, stomach/abdominal, hip, knee, pain all over the body and pain any), two phenotypes related to chronic pain lasting three months or longer (headache and back), five phenotypes related to the use of pain relief medications, and two phenotypes related to cardiovascular pain (chest pain and leg pain on walking). Combined analysis of the 17 selected pain related traits was carried out using the software PLEIO, which corrects for environmental correlations due to sample overlap. This analysis revealed 99 genome-wide significant risk loci, 65 of our 99 risk loci overlap loci from previous studies, while the remaining 34 loci are novel.

We used a leave-one-trait-out meta-analysis approach to evaluate the robustness of our results and the influence of each individual trait on the joint analysis. Thirty-seven of the 99 loci reached genome-wide significance in all leave-one-trait-out analyses, including many of those with the strongest joint p-values, e.g., 12q13.3 (P-value 6.8e-72; *LRP1*, *STAT6*), 1q22 (P-value 6.2e-29; *MEF2D*), and 17q21.31 (P-value 1.96e-28; *CRHRI*), while 22 regions were mostly driven by a single trait, among them 12p12.1 (P-value 3.56e-11; *SOX5*) for back pain, 20q11.22 (P-value 2.8e-17; *GDF5*) for knee pain and 7p14.1 (P-value 4.7e-17; *SUGCT*) for headache.

Risk loci were enriched for genes involved in neurological and inflammatory pathways. Furthermore, genetic correlations and two-sample Mendelian randomization indicated that depression, neuroticism, metabolic, cardiovascular and immunological traits mediate the risk of pain in these loci, with the strongest effects from neuropsychiatric (odds ratio=1.48; 95%CI:1.23-1.77; P-value=2.95E-05) and immunological traits (odds ratio=2.05; 95%CI:1.40-3.00; P-value=2.09E-04).

Our analyses suggest a complex genetic architecture linking genetically correlated pain phenotypes and related traits.

Complex Traits Posters - Thursday

PB1425. Genome-wide association of Copy Number Variation for Cleft Lip with or without Cleft Palate in families from a multi-ethnic study sample.

Authors:

N. Mukhopadhyay¹, E. Leslie², E. Feingold³, M. L. Marazita⁴; ¹Univ Pittsburgh/Sch Dental Med, Pittsburgh, PA, ²Emory Univ., Atlanta, GA, ³Univ of Pittsburgh, Pittsburgh, PA, ⁴Univ Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA

Abstract Body:

Orofacial clefts (OFCs) comprise one of the most common forms of birth defects worldwide, posing a large public health burden, due to the social, emotional and financial consequences for individuals affected with OFCs, their families, and society. Thus, identifying the etiologic factors for OFCs is vital for assessing risk, developing prevention methods, and determination of the extent of therapeutic and social support needed by affected individuals and families. OFCs vary by manifestation and severity, and are grouped broadly into cleft lip with or without cleft palate (CL/P) and cleft palate alone (CP). The majority of OFCs - 70% of CL/P and 50% of CP - occur without any other detectable cognitive or structural abnormality. Identifying the genetic causes of non-syndromic OFCs has been largely limited to the study of polymorphic markers such as SNPs; very few have explored the association between copy number variation (CNV) and OFC risk. Further, prior CNV studies were limited by the sample size and/or target regions analyzed. Our current study investigates whether the presence/absence of an abnormal copy number (specifically a deletion or duplication) is associated with non-syndromic CL/P. In our study, *de novo* and inherited CNVs are treated and analyzed similarly, under the assumption that both may be similarly disruptive to gene functioning. Subjects drawn from a multiethnic GWAS sample from the Pittsburgh Orofacial Cleft Study (POFC) are included in this study. CNVs were identified based on genotype intensities from a GWAS panel of SNPs, and the copy numbers at predetermined genomic coordinates located across the genome - normal vs. abnormal - analyzed within a variance-component based case-control GWAS framework. GWASs were run on the entire POFC sample, as well as subsets defined based on their ancestry. A strong association signal (most significant p-value ~ 10E-32) was observed within a region spanning a 50 KB region in 7p14.1 in the total POFC sample. The most strongly associated ancestry group was the Central and South American subset (p-value ~ 10E-26), with the European subset showing a modest association as well (p-value < 10E-05). In an earlier study of only European trios, *de novo* CNVs in this region were reported as being associated with CL/P risk. Thus our study confirms the earlier finding in Europeans, and extends the finding to trios from Central and South America. Fine-mapping and functional investigation of these and other potential associations are currently underway.

Complex Traits Posters - Wednesday

PB1426*. Genome-wide association of NASH and Mendelian randomization with plasma protein levels identifies putative protein changes resulting from disease.

Authors:

E. Smith¹, L. Sundararajan², S. Held², E. Sundaram², E. Uberbacher², M. Leuze², S. Badola³, R. Mikkilineni³, J. Schneider², D. Diogo³, L. Ekstrom², J. Stover², S. Szalma¹; ¹Takeda Dev. Ctr. Americas, Inc., San Diego, CA, ²Nashville BioSci.s, Nashville, TN, ³Takeda Dev. Ctr. Americas, Inc., Cambridge, MA

Abstract Body:

Nonalcoholic steatohepatitis (NASH) is estimated to affect 5% of adults in the United States, however only a fraction of those affected are diagnosed due to risks associated with liver biopsy. Currently, noninvasive blood biomarkers have been identified that can help identify patients that are likely have NASH, but have limitations in predicting stage or disease progression. Identifying proteins with levels that change as a result of disease could point to novel biomarkers or mechanisms in disease progression that could help diagnose and monitor patients with NASH. Recent advances in high throughput plasma proteomics now allows for profiling of 1000's of proteins in the plasma proteome. Large-scale studies, such as the UK Biobank Pharma Proteomics Project (UKB-PPP), are enabling well-powered Mendelian randomization studies that could identify causal links between disease and protein levels. To improve the genetic understanding of NASH and allow for Mendelian randomization analysis, we performed a genome-wide association study (GWAS) of non-alcoholic fatty liver disease (NAFLD) patients that we further classified as likely-NASH or likely non-alcoholic fatty liver (NAFL). NAFLD patients were identified from electronic health record data in participants of European ancestry in the BioVU database. Patients were classified as likely-NASH or likely-NAFL using a combination of diagnosis codes, biopsy reports, imaging reports, ALT measures, and keywords. Controls had no reported liver disease and were matched to cases by age and sex. Genome-wide association of 1,542 likely-NASH patients vs 6,538 control participants identified two previously reported loci: PNPLA3 and HSD17B13. We performed meta-analysis with a recently published NAFLD GWAS including NASH patients (Anstee et al.) and identified 4 additional loci that have been previously implicated in NASH: MARC1, GCKR, TRIB1, and TM6SF2. To identify proteins that might be affected as a result of NASH, we tested for association between the 6 lead variants and 1469 protein levels in 36,3741 unrelated participants of White British ancestry from the UKB-PPP. We performed two-sample Mendelian randomization and identified 108 proteins significantly associated using the weighted median test (FDR < 0.05). Significantly associated proteins were enriched for being metabolic enzymes and were highly expressed in liver hepatocytes. These results highlight the utility of large-scale proteomic studies for identifying putative biomarkers of disease and implicate a large number of proteins that may have utility in identifying patients with NASH or in monitoring disease course.

Complex Traits Posters - Thursday

PB1427. Genome-wide association studies identify genetic variants associated with comorbidities in atopic dermatitis

Authors:

A. Hartley¹, J. Saklatvala², R. Ramessur³, A. Budu-Aggrey¹, C. Smith³, S. Brown⁴, J. Min¹, N. Dand², L. Paternoster¹; ¹MRC-Integrative Epidemiology Unit, Population Hlth.Sci., Bristol Med. Sch., Bristol, United Kingdom, ²Dept. of Med. and Molecular Genetics, King's Coll. London, London, United Kingdom, ³St John's Inst. of Dermatology, King's Coll. London, London, United Kingdom, ⁴Ctr. for Genomic and Experimental Med., Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract Body:

Background: Atopic dermatitis (AD), also referred to as eczema, is a common skin condition with a substantial comorbid burden, including other atopic conditions (e.g. asthma and food allergy). To further elucidate the genetic aetiology of comorbidities in AD, we performed GWAS of relevant comorbidities stratified by AD status in UK Biobank.

Methods: Using the white European subset of UK Biobank, we analysed 36 comorbidities with a prevalence $\geq 2\%$ in UK Biobank, selected by expert review and epidemiological evidence. AD and comorbidities were identified based on self-report, hospital episode statistics, cancer and death registry data. GWAS were performed for each comorbidity, separately for 13,965 AD cases and 349,805 controls and we then performed an interaction Z-test to test for differences in SNP-comorbidity associations between cases and controls. Bolt linear mixed models were fitted on SNPs with a minor allele frequency ≥ 0.01 and an imputation INFO score ≥ 0.7 . Analyses were adjusted for age, sex and ten principal components.

Results: Sixty-nine independent genome-wide significant ($p \leq 5 \times 10^{-8}$) SNPs were identified across 20 comorbidities in the AD case population. Most of these SNPs displayed stronger associations with comorbidity in AD cases compared to controls, as evidenced by genome-wide significant p-values in the interaction test for 43 of these SNPs. Additionally, six SNPs with stronger effects in cases compared to controls, but with p-values $> 5 \times 10^{-8}$ in the case-only GWAS, were identified by the interaction test. Of the 49 SNPs associated with a condition more strongly amongst AD cases than controls, 22 SNPs were intronic, including *IGF1R* for congestive heart failure, *ST6GALNAC3* for type 2 diabetes and *ACE* for myocardial infarction.

In stratified analyses such as these, index-event bias may influence results. To assess the potential influence of index-event bias, we examined the SNP associations with AD in the latest 'Early Genetics and Lifecourse Epidemiology' consortium AD GWAS. Twenty-seven of the 49 SNPs showed no association with AD ($p > 0.05$), excluding the possibility of index-event bias. However, 18 SNPs were associated with AD ($p \leq 5 \times 10^{-8}$) and therefore interactions may be explained by this bias. These SNPs were associated with asthma and were mainly clustered around the strongly-replicated *FLG* locus on chromosome 1q21.3.

Conclusions: Our work has identified SNPs with potential roles in comorbidity associated with AD. A better understanding of the genetic mechanisms contributing to the burden of comorbidity in AD may in the future aid attempts to target these patient subgroups for more effective preventative or therapeutic strategies.

Complex Traits Posters - Wednesday

PB1428. Genome-wide association studies of human and rat body mass converge on a conserved molecular network.

Authors:

B. Leger¹, S. Wright¹, B. Rosenthal¹, J. Kreisberg¹, S. Liu¹, T. Jia¹, A. Chitre¹, O. Poleskaya¹, K. Holl², J. Gao¹, R. Cheng¹, A. Martinez³, G. Tony⁴, A. Gileta⁵, H. Chen⁶, S. Fligel⁷, P. Meyer⁸, T. Robinson⁷, L. Solberg-Woods⁹, T. Ideker¹, A. Palmer¹; ¹Univ. of California San Diego, La Jolla, CA, ²Med. Coll. of Wisconsin, Milwaukee, WI, ³The Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, ⁴Univ. of Toronto, Toronto, ON, Canada, ⁵Univ. of Chicago, Chicago, IL, ⁶Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, ⁷Univ. of Michigan, Ann Arbor, MI, ⁸Univ. of Buffalo, Buffalo, NY, ⁹Wake Forest Univ., Winston-Salem, NC

Abstract Body:

Translation of polygenic findings across species is critical for understanding the mechanisms driving normal and disease phenotypes, and for gaining clearer understanding of how well model organism paradigms recapitulate human biology. Here, we demonstrate a network genetic approach to identify the conserved and divergent mechanisms identified in GWASs for body mass index (BMI) in humans and rats. While the specific genes associated with BMI from GWAS show non-significant cross-species agreement, we show highly-significant alignment when analyzing the genes in the context of a network (PCNet, a network of 2.7 million physical and functional gene associations). In particular, we find that the conserved network highlights brain-expressed genes and mechanisms, including hormone regulation, epigenetic regulation, and cellular signaling. To clarify the unshared biology between the two species, we also identify divergent mechanisms for the two species using a similar approach. These disparate processes include water transport in rats, and additional epigenetic processes in humans. We validate these networks using functional studies in mice that are summarized by the Mouse Genome Database. This validation demonstrates conserved function across three distinct mammals evolutionarily separated by ~100M years. This study provides a roadmap for factoring any genotype-phenotype association carried out in multiple species into discrete sets of conserved and divergent molecular pathways and processes. Further, it is a novel method for clarifying whether studies in model organisms accurately model the polygenic signals identified by genetic studies in humans.

Complex Traits Posters - Thursday

PB1429. Genome-wide association study analysis of glucose response in oral glucose tolerance test (OGTT).

Authors:

V. Thaker¹, W. Gu², S. Cao², R. Salem²; ¹Columbia Univ Med Ctr, New York, NY, ²Univ. of California, San Diego, La Jolla, CA

Abstract Body:

The shape of the glucose response curve during an oral glucose tolerance test (OGTT), monophasic versus biphasic, identifies physiologically distinct group of individuals with differences in insulin secretion and β -cell sensitivity. The subjects with biphasic glucose curve, defined as a rebound of glucose following a nadir ≥ 0.025 mmol/L, have higher insulin sensitivity and β -cell function. The genetic variants linked with such glucose response have not been defined. **Method:** We performed a genome wide association study (GWAS) to identify genetic variants associated with biphasic profile using data from the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study retrieved from dbGaP. Eligible pregnant women from diverse ethnicities underwent a standard 75-g OGTT between 24-32 weeks gestation: Afro-Caribbean- (AC, n = 1345), European- (EA, n = 1507), Hispanic- (HL, n = 889), and Thai-Asian (TA, n= 1253) ancestry. Two glucose response outcomes were constructed: i) Quantitative: glucose-difference (Gluc-diff) between glucose levels at hours 2 and 1 of OGTT and ii) Binary: biphasic (Gluc-diff ≥ 0) vs. monophasic (Gluc-diff < 0). GWAS analyses were performed for each ancestry group using SAIGE, controlling for age, BMI, type 2 diabetes (T2D), and first three principal components. Ancestry specific results were combined using METAL and loci annotated using BioThings. **Results:** We identified 814 subjects with biphasic OGTT response (AC=299, EA=244, HL=131 and TA=140). The ancestry specific analyses identified an intronic variant rs10830963 in the melatonin receptor 1B (*MTNR1B*, $\beta=-6.67$, $p=4.51E-09$) with Gluc-diff in EA, that remained significant in transancestry meta-analysis ($p=2.37E-09$). *MTNR1B* is established gene for T2D and glycemia related traits. Known prioritized genes for suggestive threshold loci ($10E-6$) in transethnic metanalysis are involved in insulin biology/metabolic pathways: *ATXN1* (adipocyte differentiation, waist-hip ratio), *CNDP2* (diabetic nephropathy), *SGCZ* (T2D), *TCF12* (T2D), *NKAIN1* (BMI, BMI-adjusted hip ratio), *KIF17* (ciliary transcription factor) for binary outcome and *SGMS1* (lipid metabolism) for quantitative outcome. Other intergenic/intronic loci of unknown significance were also prioritized. **Conclusion:** Use of a novel phenotype for glucose regulation may identify additional loci and/or genes relevant to insulin biology/metabolism that may or may not be known with more traditional measures. Prior studies have validated the relevance of genes prioritized in gestational diabetes for general population; future replication studies are planned to establish their role in non-pregnant state.

Complex Traits Posters - Wednesday

PB1430*. Genome-wide association study for 233 circulating metabolic traits in 136,000 participants reveals extensive pleiotropy and novel associations.

Authors:

M. Karjalainen¹, S. Karthikeyan², E. Sliz¹, NMR Metabolomics and Genetics Consortium, M. Ala-Korpela¹, A. S. Butterworth², J. Kettunen¹; ¹Univ. of Oulu, Oulu, Finland, ²Univ. of Cambridge, Cambridge, United Kingdom

Abstract Body:

Introduction: Genome-wide association studies utilizing high-throughput metabolomics have identified multiple loci associated with circulating metabolic traits. Knowledge on the genetic determinants of systemic metabolism gives novel insights into the biological pathways affecting complex diseases. **Methods:** We performed a GWAS of 233 circulating metabolic traits quantified by nuclear magnetic resonance spectroscopy (NMR) in over 136,000 participants from 33 cohorts, thereby considerably expanding our previous GWAS which included 123 traits and 25,000 participants. GWAS was followed by multiple downstream analyses, including detailed characterization of metabolic profiles of lipid and lipoprotein-associated variants, characterization of disease associations of the metabolic trait-associated variants, and Mendelian randomization of twenty non-lipid traits. **Results:** Significant associations were identified for all 233 metabolic traits. Associations were detected at 276 genomic regions that corresponded to over 400 independent loci, which is a substantial improvement compared to our previous study which identified 62 loci. Novel loci were identified for multiple metabolic traits. As an example, associations at 26 loci were detected for glutamine, several of which were novel. Many loci showed extensive pleiotropy. Detailed characterization of metabolic profiles of 84 novel lipid- and lipoprotein-associated loci revealed that the metabolic profile of *TRIM5*, a novel coronary artery disease-associated locus, is similar to that of *HMGCR* whose effects are concordant to statin therapy. Many metabolic trait-associated variants associated with diseases. One such disease was intrahepatic cholestasis of pregnancy, whose genetic determinants and metabolic underpinnings were characterized here for the first time. Mendelian randomization analyses demonstrated causal relationships between non-lipid traits and diseases, implying, e.g., a causal association of acetoacetate (a ketone body) with hypertension. **Conclusions:** Our GWAS of 233 circulating metabolic traits revealed genetic associations in over 400 independent loci. The five-fold increase in sample size and doubling the number of metabolic traits compared to our previous NMR metabolomics GWAS resulted in a large increase in the number of significant associations, leading to a substantial improvement in understanding of genetic regulation of systemic metabolism. Extensive pleiotropy, novel associations and causal relationships were detected.

Complex Traits Posters - Thursday

PB1431. Genome-Wide Association Study in a Rat Model of Temperament Identifies Multiple Loci for Exploratory Locomotion and Anxiety-Like Traits.

Authors:

A. Chitre¹, E. Hebda-Bauer², P. Blandino², K-M. Nguyen¹, P. Maras², F. Li², S. Flagel², A. Ozel², O. Polesskaya¹, R. Cheng¹, S. Watson, Jr.², J. Li³, A. A. Palmer⁴, H. Akil²; ¹Univ. of California San Diego, La Jolla, CA, ²Univ. of Michigan, Ann Arbor, MI, ³Univ Michigan, Ann Arbor, MI, ⁴Univ. of California San Diego, LA JOLLA, CA

Abstract Body:

Common genetic factors likely contribute to multiple psychiatric diseases including mood and substance use disorders. Certain stable, heritable traits reflecting temperament, termed externalizing or internalizing, play a large role in modulating vulnerability to these disorders. To model these heritable tendencies, we selectively bred rats for high and low exploration in a novel environment (bred High Responders (bHR) vs. Low Responders (bLR)). Although selective breeding has been carried on for tens of generations, the bHR and bLR retain significant genetic diversity, although the sites under the strongest selection are likely to be divergently fixed or nearly fixed when contrasting the bHR and bLR. To identify genes underlying the response to selection, we phenotyped and genotyped 538 rats from an F2 cross between bHR and bLR. The SNP heritability estimates for these behavioral traits ranged from 0.06 to 0.79. There were significant phenotypic and genetic correlations between tests that capture facets of exploratory locomotion and anxiety. Ten significant and conditionally independent loci for six behavioral traits were identified. Five of the six traits reflect different facets of exploratory locomotion that were captured by three behavioral tests. Distance traveled measures from the open field and the elevated plus maze map onto different loci and may, thus, represent different aspects of novelty-induced locomotor activity. The sixth behavior trait is the only anxiety-related trait mapping to a significant locus on chromosome 18 within which the *Pik3c3* gene is located. We also identified *Fancf* and *Gas2* as potential candidate genes that may drive the chromosome 1:107 Mb QTL for exploratory locomotion traits. The identification of a locomotor activity-related QTL on chromosome 7 encompassing the *Pkhd111* and *Trhr* genes is consistent with our finding of these genes being differentially expressed in the hippocampus of bHR vs. bLR rats. The strong heritability coupled with identification of several loci associated with exploratory locomotion and emotionality provide compelling support for this selectively bred rat model in discovering relatively large effect causal variants tied to elements of internalizing and externalizing behaviors relevant to psychiatric and substance use disorders.

Complex Traits Posters - Wednesday

PB1432. Genome-wide association study in mesial temporal lobe epilepsy with hippocampal sclerosis

Authors:

E. Bruxel^{1,2}, P. H. M. Magalhães^{1,2}, T. K. de Araujo^{1,2}, T. C. de Oliveira^{1,2}, M. K. M. Alvim^{1,2}, M. E. Morita-Sherman^{1,2}, R. Secolin^{1,2}, C. L. Yasuda¹, B. Leal³, J. Chaves³, A. M. Ferreira³, A. M. da Silva³, L. E. Betting⁴, K. Lin⁵, R. Walz⁵, M. R. Rodrigues⁶, T. Hünemeier⁶, L. V. Pereira⁷, A. C. Pereira⁸, F. Cendes^{1,2}, I. Lopes-Cendes^{1,2}; ¹Univ. of Campinas, Campinas, Brazil, ²Brazilian Inst. of NeuroSci. and Neurotechnology (BRAINN), Campinas, Brazil, ³Univ. of Porto, Porto, Portugal, ⁴São Paulo State Univ., Botucatu, Brazil, ⁵Federal Univ. of Santa Catarina, Florianópolis, Brazil, ⁶Univ. of São Paulo, São Paulo, Brazil, ⁷Univ. de São Paulo, São Paulo, Brazil, ⁸Harvard Med. Sch., Boston, MA

Abstract Body:

Introduction: Mesial temporal lobe epilepsy (MTLE) is a chronic neurological disorder characterized by the occurrence of spontaneous seizures initiated in the mesial temporal lobe structures. MTLE is a type of focal epilepsy frequently associated with a characteristic histopathological lesion denominated hippocampal sclerosis (HS). Many patients with MTLE+HS have severe epilepsy, which is often resistant to treatment with antiseizure medication. Recent genome-wide association metanalysis found suggestive evidence of association to two loci on chromosomes (ch)s 3q25.41 and 6q22.31 in patients with focal epilepsy and HS, and a locus on ch 2q24.3 for all focal epilepsies combined [1]; however, to our knowledge, these results have not been replicated. **Materials and Methods:** We studied 501 patients and 1.146 controls. Patients with MTLE+HS were ascertained in several epilepsy centers in Brazil and Portugal. All patients fulfilled the diagnostic criteria for MTLE+HS proposed by the International League Against Epilepsy. In addition, controls were ethnically matched to patients. We used the Affymetrix 6.0 array for SNP genotyping. **Results:** Principal component analyses showed no significant differences in the genomic composition of patients and controls and were used as covariates in the association study. After quality control, we used an additive model for the association to perform a logistic regression. We found four novel association signals at chs 7 ($p = 5.689e^{-10}$), 9 ($p = 2.68e^{-10}$), 12 ($p = 2.989e^{-10}$) and ch 14 ($p = 3.645e^{-8}$). **Discussion/Conclusion:** The SNPs associated with the phenotype are in introns and intergenic regions, suggesting that regulatory elements are likely to be involved in the genetic predisposition to MTLE+HS. Future steps in this project aim to perform local ancestry inference and integrate transcriptomic and epigenomic data to identify putative causal variants. **Funding:** Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), SP, Brazil. **Reference:** [1] International League Against Epilepsy Consortium on Complex Epilepsies. Nat Commun 9(1): 5269, 2018.

Complex Traits Posters - Thursday

PB1433. Genome-wide association study of diabetic retinopathy in the Million Veteran Program identifies four significant loci

Authors:

J. Breeyer^{1,2}, A. Giri^{3,2}, S. L. Mitchell^{3,2}, J. N. Hellwege^{3,2}, B. Charest⁴, A. Khakharia^{5,4}, C. W. Halladay⁶, P. D. Reaven^{7,8}, J. B. Meigs^{9,10}, M. K. Rhee^{4,5}, S. Yang¹¹, M. G. Lynch⁴, C. Nealon¹², S. K. Iyengar^{13,14}, M. A. Brantley^{1,3}, N. S. Peachey^{15,14}, Y. V. Sun^{5,4}, P. W. Wilson^{4,5}, L. S. Phillips^{5,4}, T. L. Edwards^{3,2}, VA Million Veteran Program; ¹Vanderbilt Univ., Nashville, TN, ²VA Tennessee Valley Hlth.care System, Nashville, TN, ³Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁴Atlanta VA Med. Ctr., Decatur, GA, ⁵Emory Univ., Atlanta, GA, ⁶Providence VA Med. Ctr., Providence, RI, ⁷Phoenix VA Hlth.Care System, Phoenix, AZ, ⁸Arizona Univ., Phoenix, AZ, ⁹Massachusetts Gen. Hosp., Boston, MA, ¹⁰Broad Inst., Cambridge, MA, ¹¹Stanford Univ., Palo Alto, CA, ¹²VA Northeast Ohio Hlth.care System, Cleveland, OH, ¹³Case Western Reserve Univ., Cleveland, OH, ¹⁴Louis Stokes Cleveland VA Med. Ctr., Cleveland, OH, ¹⁵Cole Eye Inst., Cleveland Clinic, Cleveland, OH

Abstract Body:

Diabetic retinopathy (DR) is the leading cause of vision loss and preventable blindness in adults, affecting an estimated 93 million individuals worldwide. The course of diabetes and severity of complications vary substantially among individuals, and this variability is not fully explained by risk factors such as glycemic control. Several genome-wide association studies (GWAS) of DR have been conducted, and while significantly associated loci have been reported, few have replicated. We used a validated electronic health record algorithm to identify DR cases (proliferative, severe non-proliferative, moderate non-proliferative, and unspecified DR; total N = 33,627) and controls with diabetes but without retinopathy (N = 52,388) in the Million Veteran Program. GWAS of TOPMed imputed SNPs were conducted using logistic regression adjusting for age, sex, body mass index, hypertension status, duration of diabetes, and the top ten principal components. Analyses were stratified using the harmonized ancestry and race/ethnicity (HARE) approach and three groups were included in the present study: European (21,486 cases/35,986 controls), African American (8,559 cases/12,079 controls), and Hispanic (3,582 cases/4,323 controls). We identified one, two, and one genome-wide significant loci in the European, African American, and Hispanic populations, respectively. All genome-wide significant SNPs identified in the analysis of European ancestry individuals were located in the intronic regions of TCF7L2, a gene previously associated with type 2 diabetes. The African American analysis identified loci in the intronic regions of CNGB3 (rs7824288 (Odds ratio [OR] = 0.83 (95% confidence interval [CI] 0.78 - 0.88), p = 3.44x10⁻⁹) and ARHGAP29-AS1 (rs55937938 (OR = 1.29 (95% CI 1.18 - 1.41), p = 1.75x10⁻⁸)). The Hispanic analysis identified rs61833024, intergenic on chromosome one, (OR = 0.35 (95% CI 0.24 - 0.51), p = 2.71x10⁻⁸). While the significant locus identified in the European population may be a proxy for glycemic control, the loci identified in the Hispanic and African American analyses have not been previously associated with DR and their role is not yet defined. With the resources of the Million Veteran Program, we conducted the largest multi-ancestry GWAS of Hispanic, African American, and European populations to date, identifying three previously unreported loci associated with DR in populations underrepresented in genomic research.

Complex Traits Posters - Wednesday

PB1434. Genome-wide association study of longitudinal changes in motor symptomology in Parkinson's Disease

Authors:

W. Kim¹, F. Tao¹, S. Hall^{2,3}, O. Hansson^{2,3}, M. Nagle¹; ¹Genetics Guided Dementia Discovery (G2D2) Eisai Inc., Cambridge, MA, ²Dept. of Clinical Sci., Lund Univ., Malmo, Sweden, ³Memory Clinic, Skane Univ. Hosp., Malmo, Sweden

Abstract Body:

Introduction: To date, there are no therapeutics to slow the progression of Parkinson's Disease (PD), which would have tremendous impact for patients. While genome-wide association studies (GWAS) have identified genetic factors associated with risk of PD, genetic determinants of motor progression of PD remain largely unknown. At Eisai's G2D2 research institute, we apply human genetics to discover novel drug targets for neurodegenerative disease. Given our ongoing work, this study aims to identify the genetic influences of motor progression of PD in two longitudinal cohorts, Accelerating Medicines Partnership Parkinson's disease (AMP-PD) and Biomarkers For Identifying Neurodegenerative Disorders Early and Reliably (BioFINDER), and the published GWAS of progression of PD clinical endpoints in 12 longitudinal patient cohorts (Iwaki et al., 2019).

Methods: We examined seven PD clinical measures, including the Unified Parkinson's Disease Rating Scale (UPDRS) or Movement Disorder Society revised UPDRS Part I to IV and total score, modified Schwab & England activities of daily living scale, and Hoehn and Yahr (HY) staging scale. In AMP-PD and BioFINDER cohorts, genetic and longitudinal phenotype data were available on a total of 3,164 subjects with 19,540 observations for a median of 1.5 years. We derived a slope of longitudinal change using the linear mixed model with each subject's deviation as the random effect. We then regressed this slope estimate on an imputed genome-wide marker under an additive genetic model for each cohort. Age at baseline, sex, clinical site if needed, and principal components were included as covariates. We meta-analyzed 2 GWAS results from AMP-PD and BioFINDER and one published GWAS result from a total of 4,093 patients with 22,307 observations for a median of 3.81 years under the inverse-variance fixed model. We then performed candidate gene analysis focusing on 78 reported PD risk loci to determine whether genome-wide significant PD susceptibility loci are associated with the progression of PD clinical measures.

Results: We did not identify genetic loci reaching genome-wide significance in our meta-analysis. However, in the candidate gene analysis, rs12528068-T near *RIMS1* (Regulating Synaptic Membrane Exocytosis 1) was associated with the annual increase in HY score, passing the Bonferroni corrected p-value (Beta=0.27, P=6.26E-04, Minor allele frequency=0.29). This genetic locus was reported to be associated with an increased risk of PD (Beta=0.07, P=1.63E-10).

Conclusions: Our study identified a genetic locus associated with both susceptibility and progression of PD. Further study is required to replicate our findings.

Complex Traits Posters - Thursday

PB1435. Genome-wide association study of nephrotic syndrome replicates *APOLI* and MHC loci, while predicted gene expression analysis identifies *C4A*.

Authors:

J. Hellwege¹, J. H. Breeyear², E. S. Torstenson³, G. P. Jarvik⁴, O. Dikilitas⁵, I. J. Kullo⁵, R. Rowley⁶, K. Kiryluk⁷, K. Susztak⁸, T. L. Edwards¹; ¹Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Vanderbilt Univ., Nashville, TN, ³Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁴Univ Washington Med Ctr., Seattle, WA, ⁵Mayo Clinic, Rochester, MN, ⁶Natl. Human Genome Res. Inst., Bethesda, MD, ⁷Columbia Univ., New York, NY, ⁸Univ Pennsylvania, Philadelphia, PA

Abstract Body:

Nephrotic syndrome is characterized by heavy proteinuria (>3g/24 h), hypoalbuminemia, hyperlipidemia, and edema. The etiology of nephrotic syndrome is heterogeneous, with diverse renal pathophysiology leading to similar clinical manifestations, and prevalence varying by both age and race/ethnicity. Although some cases may be monogenic in origin, genome-wide association studies have identified common variation contributing to risk of nephrotic syndrome. We integrated imputed genotype data from Vanderbilt University's BioVU and the Electronic Medical Records and Genetics (eMERGE) network with summary statistics from Biobank Japan, UK Biobank and FinnGen, for a cross-ancestry total of 2,683 cases and 731,916 controls. Cases were identified by diagnostic codes for nephrotic syndrome, while controls were without renal disease diagnosis codes. Genetic associations with nephrotic syndrome status were modeled as a function of additive genotype, sex, and the top 10 principal components of ancestry, followed by inverse-variance weighted fixed-effects meta-analysis. Across SNPs passing quality filters and with minor allele frequency >0.01, 21 independent SNPs at 17 unique loci were significant at $p < 5 \times 10^{-8}$. The smallest p-value was found for rs2763980 ($p = 3.3 \times 10^{-16}$, odds ratio [OR] = 1.8) near *LSM2* in the class III MHC region on chromosome 6. The G1 variants of *APOLI* were also strongly associated ($p = 5.9 \times 10^{-16}$ and 9.2×10^{-16} , OR = 2.6), driven by African ancestry individuals in BioVU and eMERGE. *APOLI* is a well-known risk factor for various forms of kidney disease in African ancestry individuals, including nephrotic syndrome. Genetic correlations were estimated with LD score regression between nephrotic syndrome and estimated glomerular filtration rate (-0.3 [standard error (SE) = 0.2]), blood urea nitrogen (0.1 [SE = 0.1]), systolic blood pressure (0.2 [SE = 0.2]), and type 2 diabetes (0.7 [SE = 0.5]), but none were significant ($p > 0.1$). We also evaluated associations between nephrotic syndrome and genetically predicted gene expression using GTEx kidney cortex tissue models. One gene, *C4A*, was significant in this analysis ($p = 9 \times 10^{-11}$), with increased expression associated with decreased risk of nephrotic syndrome. Our results identified numerous loci associated with nephrotic syndrome, and suggest more shared genetic architecture with type 2 diabetes than with kidney function or blood pressure, which is consistent with diabetic kidney disease leading to nephrotic syndrome. This study provides insight into nephrotic syndrome through identification of genetic loci, shared genetic architecture, and translation to predicted renal gene expression.

Complex Traits Posters - Thursday

PB1437. Genome-wide association study of plasma amyloid β levels in older adults

Authors:

M. Aslam¹, K. Fan¹, B. Snitz¹, S. DeKosky², O. Lopez³, E. Feingold⁴, M. Kamboh⁴; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²Univ. of Florida, Gainesville, FL, ³Univ. of Pittsburgh, Department of Neurology, PA, ⁴Univ of Pittsburgh, Pittsburgh, PA

Abstract Body:

In addition to the deposition of amyloid plaques in the brain, plasma levels of amyloid-beta ($A\beta$) are becoming of great interest as an important endophenotype for Alzheimer's disease (AD). Identification of genetic variants modulating plasma $A\beta$ levels could help to better understand the underlying pathways regulating $A\beta$ levels in plasma. Baseline plasma $A\beta_{40}$ and $A\beta_{42}$ levels were measured in approximately 3,000 older subjects (>75 years; 94% self-identified white), derived from the Ginkgo Evaluation of Memory (GEM) study. Plasma $A\beta_{42}/40$ ratio was calculated on subjects with both measurements. Genome-wide genotyping was performed on the available ~2,700 DNA samples using Illumina Infinium Multi-Ethnic Global-8 Kit followed by imputation through the Michigan imputation server. Genome-wide association analyses on $A\beta_{40}$, $A\beta_{42}$, and $A\beta_{42}/40$ ratio were performed on white subjects having both genotype and phenotype data. Analyses were performed using a linear regression framework implemented in PLINK, including age, sex, and the first four principal components of ancestry as covariates. FUMA GWAS (<https://fuma.ctglab.nl/>) was used for the pathway and gene-based analyses. A genome-wide significant signal was observed for plasma $A\beta_{42}$ levels on chromosome 19 in *KLHL26* (*rs11665980*, $\beta = 0.06304$, $p = 4.76E-08$). The same variant (*KLHL26/rs11665980*) was also associated with increased plasma $A\beta_{42}/40$ ratio ($\beta = 0.06578$, $p = 5.66E-08$). Interestingly, in the gene-based analysis *KLHL26* was the top gene for plasma $A\beta_{42}$ ($p = 8.0914E-06$) and $A\beta_{42}/40$ ratio ($p = 5.1024E-06$) and also showed association with $A\beta_{40}$ ($p = 0.00071$). *KLHL26* belongs to *KLHL* family, which is involved in ubiquitin-mediated protein degradation, cytoskeletal rearrangement, and actin binding. In conclusion, our study has identified a novel *KLHL26* gene modulating plasma amyloid levels, which needs to be confirmed in independent and larger samples.

Complex Traits Posters - Wednesday

PB1438. Genome-wide association study of venous thromboembolism in women using combined oral contraceptives

Authors:

D-A. Trégouët¹, M. GERMAIN¹, P. Suchon², G. Munsch¹, N. Saut², D. Smadja³, J. Emmerich⁴, D. Bacq-Daian⁵, A. Boland⁵, J-F. Deleuze⁵, P-E. Morange²; ¹Bordeaux Population Hlth.Res. Ctr., INSERM U1219, Bordeaux, France, ²Aix Marseille Univ., La Timone Hosp., Marseille, France, ³Hematology Dept. and Biosurgical Res. Lab (Carpentier Fndn.), European Georges Pompidou Hosp., Assistance Publique Hôpitaux de Paris, Paris, France, ⁴Dept. of vascular medicine, Paris Saint-Joseph Hosp. Group, INSERM 1153, Paris, France, ⁵CNRGH, Evry, France

Abstract Body:

Millions of women worldwide use combined oral contraceptives (COC). In some age groups, more than 50% of women are using COC. COC use induces a prothrombotic state and associates with a ~five-fold increased risk of venous thromboembolism (VTE). In 2013, the pill scare has underlined the need for identifying at risk women before the first prescription of COC. However, no efficient and powerful tool is currently available to achieve this goal, in particularly because the exact underlying pathophysiological mechanisms are still not well characterized. To fill this gap and identify novel molecular players involved in COC induced VTE, we here performed the first genome-wide association study (GWAS) for VTE in COC users. Capitalizing on a sample of 964 women who developed VTE on COC and 662 COC users with no personal history of VTE, we tested the association with VTE of 10,442,711 single nucleotide polymorphisms (SNPs) with imputation criteria > 0.3 and minor allele frequency > 0.005 . One genome-wide association signal was observed at the known *ABO* locus (lead SNP rs505922, p-value = $5.28 \cdot 10^{-13}$). Haplotype analysis revealed that ABO blood A1 and B groups were associated with increased risk of VTE (OR = 1.80 [1.50 - 2.17], p = $9.4 \cdot 10^{-10}$; OR = 1.57 [1.19 - 2.08], p = $1.5 \cdot 10^{-3}$, respectively). Intriguingly, A2 blood group was found protective in this population (OR= 0.57 [0.35 - 0.92], p = 0.022). Of note, among the 4 additional loci (*DTHD1*, *IRF2*, *JADE2*, *SLC39A11*) that exhibited suggestive statistical association at $p < 10^{-6}$, two (*IRF2* and *JADE2*) have previously been found associated with plasma levels of Sex Hormone Binding Globulin (SHBG). Increased SHBG levels have already been observed in VTE patients under COC compared to healthy COC users, suggesting that genetically determined SHBG levels could serve as a biomarker for VTE in COC users. Polygenic risk score analyses are ongoing together with the replication of the *IRF2* and *JADE2* association in an independent sample of 735 VTE on COC and 735 women without VTE on COC.

Complex Traits Posters - Thursday

PB1439. Genome-wide association study on skin tags

Authors:

M. Liu¹, J. Lee¹, B. Riley-Gillis¹, N. Smaoui¹, J. F. Waring¹, M. Reppell¹, X. Zheng¹, S. Manson Brown², T. Cheng², R. Mehta², S. Peterson², A. Abbasi¹; ¹AbbVie Inc., North Chicago, IL, ²Allergan Aesthetics, Irvine, CA

Abstract Body:

Skin tags, also known as acrochordons, are common benign lesions that are more likely to develop as people age. Skin tags are mainly found where the skin folds, such as the armpits and the neck, which may be due to constant friction of skin. Although usually not painful, skin tags can impact self-image and self-esteem; therefore, understanding their etiology with a goal of avoidance or treatment merits effort. Aging, obesity and insulin resistance have been associated with skin tags, however, the biological mechanisms underlying this condition remains unknown. Moreover, as they typically manifest later in life, there could be a natural part of the aging process that is yet to be fully understood. To understand the genetic basis of skin tags a genome-wide association study (GWAS) for skin tags was run using a large-scale population cohort, UKBiobank. We created a case-control cohort using a combination of electronic health records (EHR) with 9,324 cases and 158,113 controls. The analysis was limited to participants of European ancestry and GWAS was performed using SAIGE mixed model and adjusted for age, sex, principal components, and relatedness in the mixed models. In total, we identified 3 loci for skin tags around the genes *HMCNI*, *TERT* and *ABCA1*. We found significant eQTL in skin-related sun-exposed and not-sun-exposed tissue for genes *HMCNI* and *TERT*. Interestingly, *HMCNI* has been shown to play a part in basement membrane organization of the skin. We will present genetic correlation and mendelian randomization with other obesity and skin-related traits to better understand the shared underlying biology. To our knowledge, this is the largest GWAS to date on the skin tag phenotype providing insights into the genetic contribution and biological processes involved.

Complex Traits Posters - Wednesday

PB1440. Genome-wide association study to identify genetic variants for obesity in Korean population

Authors:

S. Yang, Y. Kwak, M. Yuk, J. Lee, J. Youn, N. Song; Coll. of Pharmacy, Chungbuk Natl. Univ., Cheongju, Korea, Republic of

Abstract Body:

Background: Obesity is affected by approximately 40% of genetic factors. The previous genome-wide association studies (GWAS) for obesity have reported more than 15 genes and 100 genetic loci. But most of GWAS were performed in European and only about 20% of studies were researched in Asian. Asian have a difference in obesity criteria and might have different genetic variants from Europeans. We conducted a GWAS on obesity in Korean population to replicate previously reported genetic loci and identify novel genetic loci in East Asian.

Methods: From the Korean Genome and Epidemiology Study (KoGES) including the Health EXAminee (HEXA) study (n=58,690), the CARdioVascular disease Association Study (CAVAS, n=8,105), and the Ansan and Ansung study (n=5,493), a total of 72,288 study participants with epidemiological and genome-wide single nucleotide polymorphism (SNP) data were selected. Obesity was defined as a body mass index $\geq 25\text{kg/m}^2$ according to World Health Organization obesity classification for Asian. The previously identified 60 obesity susceptibility SNPs from the GWAS ($P < 5.0 \times 10^{-8}$) were selected for replication. In each cohort, we performed GWAS for obesity using multivariable logistic regression model with adjustment for age, sex, history of chronic diseases (i.e., diabetes, hypertension, and hyperlipidemia), and smoking status. Furthermore, the GWAS results of 3 cohorts were combined by a meta-analysis.

Results: Among the 60 SNPs previously identified by GWAS on obesity, 28 SNPs were statistically significantly replicated ($P < 0.05$), where 11 SNPs (rs633715, rs6752378, rs2030323, rs1421085, rs1558902, rs17817449, rs7185735, rs8043757, rs10871777, rs11152213, and rs17782313) had a nominal genome-wide significance ($P < 5.0 \times 10^{-8}$) in Korean population. In the meta-analysis of GWAS results for 3 cohort, we identified 69 SNPs (novel 65 SNPs and known 4 SNPs) associated with obesity ($P < 5.0 \times 10^{-8}$) showing extensive overlapping of genetic susceptibility regions (*FTO*, *ADCY3*, *BDNF*, *BDNF-AS*, *LINC00678*, *DNAJC27*, and *CRYZL2P-SEC16B*) previously identified in European. The most significant SNP for obesity was rs8050136 (16q12.2; $P = 2.2 \times 10^{-18}$) at well-known obesity gene *FTO*.

Conclusion: We identified 69 obesity-related SNPs in Korean population showing a similar aspect of obesity-associated genetic loci to European suggesting the worldwide obesity-risk genes of *FTO*, *ADCY3*, *BDNF*, *BDNF-AS*, *LINC00678*, *DNAJC27*, and *CRYZL2P-SEC16B*.

Complex Traits Posters - Thursday

PB1441. Genome-wide Interaction Study with Smoking Identifies *FHIT* and *SLC22A23* Associated with Alzheimer's Disease.

Authors:

X. Han¹, M. R. Moore², J. Chung¹, S. Durape², M. Rosenthaler², M. Uretsky², B. Abdolmohammadi², J. Farrell¹, X. Zhang¹, G. Jun^{1,2}, W. Qiu², M. Logue³, T. Stein², D. A. Bennett⁴, P. Crane⁵, M. Kamboh⁶, K. A. Walter⁵, E. B. Larson⁵, R. Au², J. L. Haines⁷, M. A. Pericak-Vance⁸, G. D. Schellenberg⁹, R. Mayeux¹⁰, K. L. Lunetta¹¹, L. A. Farrer^{1,2,11}, J. Mez²; ¹Dept. of Med., Boston Univ. Sch. of Med., Boston, MA, ²Boston Univ. Alzheimer's Disease Res. Ctr., Boston Univ. Sch. of Med., Boston, MA, ³VA Boston Hlth.care System, Boston, MA, ⁴Dept. of Neurological Sci., Rush Med. Coll., Chicago, IL, ⁵Dept. of Med., Univ. of Washington, Seattle, WA, ⁶Dept. of Human Genetics, Univ. of Pittsburgh, Pittsburgh, PA, ⁷Dept. of Epidemiology and Biostatistics, Case Western Reserve Univ., Cleveland, OH, ⁸The John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL, ⁹Dept. of Pathology and Lab. Med., Univ. of Pennsylvania, Philadelphia, PA, ¹⁰Dept. of Neurology, Columbia Univ., New York, NY, ¹¹Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA

Abstract Body:

Introduction: Alzheimer's disease (AD) is a progressive neurologic disorder with memory loss and dementia. Gene-environment interactions may explain missing heritability for AD. Cigarette smoking has been clearly implicated as a risk factor for AD in meta-analysis. We conducted a genome-wide (GW) study assessing the joint main and interaction effect with smoking to further investigate AD genetic architecture. **Methods:** A total of 20,687 non-Hispanic White (NHW; 7,821 cases and 12,866 controls) and 2,093 African American (AA; 761 cases and 1,332 controls) participants from the AD Genetic Consortium and the Framingham Heart Study who had lifetime smoking information available were analyzed. "Ever smoking" (i.e., past or current smoking) was considered to be a positive exposure. Across 33 datasets, we conducted a GW joint analysis (2 degrees of freedom [2DF] test) of the combined SNP main and SNP-by-smoking interaction effects. The joint effect was estimated using logistic regression and the F-Test. Age, sex, and principal components for population structure were included as covariates. Meta-analysis with METAL within ethnic background was conducted by combining p-values from the 2DF tests weighted based on sample size. **Results:** In addition to the known *APOE* locus, we identified two novel GW significant associations within *FHIT* (top SNP: rs2367086 minor allele frequency [MAF]=0.37, P=4.07×10⁻⁸, OR=1.05 in smokers, OR=1.01 in non-smokers, intronic, chromosome 3) in the NHW sample and within *SLC22A23* (top SNP: rs9392488 MAF=0.31, P=2.11×10⁻⁹, OR=1.24 in smokers, OR=0.58 in non-smokers, intronic, chromosome 6) in the AA sample. The minor alleles of rs2367086(T) and rs9392488(C) were significantly associated with decreased expression levels of *FHIT* (anterior cingulate cortex [P=3.40×10⁻⁴]) and *SLC22A23* (frontal cortex [P=0.036]) respectively in the GTEx database and also associated with increased methylation levels in blood at CpGs near the candidate genes (cg13679804 for rs2367086 [P=1.32×10⁻¹⁵⁵]; cg18842474 for rs9392488 [P=3.97×10⁻¹¹⁷]) in the GoDMC database. *FHIT*, which has also been implicated in a GWAS of blood total tau levels, encompasses the common fragile site FRA3B susceptible to carcinogen-induced damage leading to translocations and aberrant transcripts. *SLC22A23* codes for a transmembrane protein implicated in smoking cessation and resilience to AD. **Conclusion:** In this GW interaction study with smoking of AD, we identified two novel and promising ancestry-specific loci. Our findings highlight the value of considering environmental, genetic and ancestry-specific factors together towards a personalized medicine approach to AD.

Complex Traits Posters - Wednesday

PB1442. Genome-wide meta-analysis identifies a potential therapeutic targets for new loci and functional pathways influencing Alzheimer's disease risk

Authors:

J. Kim¹, S. Park¹, J. Cha¹, D-H. Park¹, H. Lim¹, J. Kim¹, S. Lee¹, H. Kim¹, W. Chung²; ¹Basgenbio Co., Ltd., Seoul, Korea, Republic of, ²Harvard Sch. of Publ. Hlth., Boston, MA

Abstract Body:

Alzheimer's disease (AD) has a complex causal pathway in which environmental characteristics and individual characteristics, such as aging and genetics, interact. In particular, the genetic factors involved in its pathogenesis are complex. Previous studies suggested that the pathogenesis of AD is influenced by multiple genetic components, rather than a single genetic factor. It may bring the benefits of GWAS-based research to identify biomarkers and potential therapeutic targets for AD. While previous studies on potential biomarkers and therapeutic targets have suggested genes (variants and locus), results may depend on the characteristics of the cohort used in each study, such as race, sample size, and the genomic data used for the analysis. In this study, we performed a GWAS analysis on AD using samples excluding non-white in the UK Biobank data. First, we selected a proxy AD group for people who do not have AD in consideration of the heritability of AD, such as whether the parents have AD, the timing of the onset of the parents, and the period of survival. It was confirmed that the GWAS results of the proxy AD group replicated the GWAS results of the clinical AD patient group in the previous studies. Next, a meta-GWAS was performed based on the recently reported AD research data. A total of 20 significant loci were identified, excluding the APOE gene region, and 4 significant novel loci of new AD were discovered. Four loci are coding regions for HS3ST1, ANK3, MADD, and CYYR1 genes, which have functional relationships directly or indirectly with AD. For functional annotation of new loci, the eQTL of the brain region for the corresponding SNP was checked in the GTEx portal, and functional evaluation was performed through colocalization. When single-tissue eQTLs for the corresponding SNPs were checked on the GTEx portal for functional annotation of new loci, three genes (HS3ST1, ANK3, and MADD) showed strong expression patterns in brain tissue and differences in expression according to genotypes. In the CYYR1, it was presented that the different expression levels in adipose and function evaluation were performed through colocalization. The results of this study are expected to serve as a basis for functional follow-up studies and may be utilized in pre-screening of potential risk groups for AD risk. In addition, it can provide research data for screening novel biomarkers of AD and targets for drug development. We anticipate that functional interpretation strategies and follow-up experiments will result in a comprehensive understanding of late-onset AD aetiology, which will serve as a solid foundation for the improvement of AD therapy.

Complex Traits Posters - Thursday

PB1443. Genome-wide meta-analysis identifies novel susceptibility loci for acne vulgaris

Authors:

M. Kals^{1,2}, M. Teder-Laving¹, A. Reigo¹, R. Ehin^{1,3,4}, T. Objärtel¹, M. Vaht¹, T. Nikopensius¹, A. Metspalu^{1,5}, K. Kingo^{6,7}, ¹Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, ²Inst. for Molecular Med. Finland (FIMM), HiLIFE, Univ. of Helsinki, Helsinki, Finland, ³Inst. of Hlth. Technologies, Tallinn Technical Univ., Tallinn, Estonia, ⁴BioCC Ltd., Tallinn, Estonia, ⁵Inst. of Molecular and Cell Biology, Univ. of Tartu, Tartu, Estonia, ⁶Inst. of Clinical Med., Faculty of Med., Univ. of Tartu, Tartu, Estonia, ⁷Tartu Univ. Hosp., Tartu, Estonia

Abstract Body:

Acne vulgaris is a common inflammatory disease that primarily impacts facial skin in adolescence and in young adulthood, causing long-term psychosocial consequences. It is characterized by chronic inflammation of the pilosebaceous unit resulting from bacterial colonization of hair follicles by *Propionibacterium acnes*, androgen-induced increased sebum production, and altered keratinization. To undertake the analysis of genetic effects on outcome in acne, we performed a genome-wide meta-analysis comprising 34,422 individuals with acne and 364,991 controls from three independent cohorts with European ancestry (Estonian Biobank, FinnGen, and LifeLines). We identified four novel genome-wide significant loci: 11q12.2 (*FADS2*); 12q21.1 (*LGR5*); 17q25.3 (*FASN*); and 22q12.1 (*ZNRF3*) and replicated 19 previously implicated risk loci for European populations, bringing the total number of reported acne susceptibility loci to 50. We estimated that 9.4% of the phenotypic variance (SNP-based heritability) of acne was explained by common variation across the genome. Evaluation of polygenic risk scores (PRS) indicated a strong association with acne pathogenesis - individuals reporting acne had significantly higher mean acne PRSs than those who reported no acne. Furthermore, difference in PRS distributions was more significant in individuals with severe form of the disease, acne conglobata, compared to normal controls. Our findings highlight Wnt, p38 MAPK, and TGF β -mediated signaling pathways as key factors in genetic predisposition to acne vulgaris together with variation in fatty acid metabolism-related network of genes affecting the structure and maintenance of the hair follicle, and particularly the pilosebaceous unit. Understanding the genetic susceptibility factors and underlying biological pathways involved in the pathogenesis of acne will promote gaining insights for development of more effective acne treatment schemes.

Complex Traits Posters - Wednesday

PB1444. Genome-wide metabolite quantitative trait loci analysis in red blood cells from routine blood donors

Authors:

A. Moore¹, M. P. Busch², K. Dziewulska³, E. A. Hod⁴, J. C. Zimring³, A. D'Alessandro⁵, G. P. Page⁶, for the National Heart, Lung, and Blood Institute (NHLBI) Recipient Epidemiology Donor Evaluation Study; ¹RTI Intl., Durham, NC, ²Vitalant Res. Inst., San Francisco, CA, ³Univ. of Virginia, Charlottesville, VA, ⁴Columbia Univ. Med. Ctr., New York City, NY, ⁵Univ. of Colorado Denver-Anschutz Med. Campus, Aurora, CO, ⁶RTI Intl., Atlanta, GA

Abstract Body:

The Red Blood Cell (RBC)-Omics study, part of the larger NHLBI-funded Recipient Epidemiology and Donor Evaluation Study (REDS-III), aims to understand the genetic contribution to blood donor and RBC characteristics, and to outcomes in recipients of RBC from these donors. Previous work identified genetic and metabolic underpinnings to blood donation, storage, and transfusion outcomes, but none have yet linked the two bodies of work. We performed a pilot Genome-Wide Association (GWA) analysis using RBC-Omics study participants with generated untargeted metabolomics data to identify metabolite QTL (mQTL) in RBCs. RBC-Omics participants were recruited from four blood donation centers across the United States from among volunteer donors and oversampled for racial/ethnic diversity and high-frequency donors. Metabolomics data was generated from stored, frozen, leukoreduced RBC (LR-RBC) aliquots using ultra-high-pressure liquid chromatograph-mass spectrometry. 512 metabolites were assigned molecule names. As part of metabolite quality control and processing, we excluded metabolites after manual review (n=5), pharmaceuticals with implausible distributions (n=22), and high missingness (n=103). We imputed missing values in the remaining metabolites using quantile regression imputation of left-censored data, natural log-transformed the relatively quantified metabolites, and normalized values of each metabolite. We performed GWA analyses of 382 metabolites in 243 individuals imputed using the 1000 Genomes Project phase 3 all-ancestry reference panel. Analyses were conducted using ProbABEL and adjusted for sex, age, donation center, number of whole blood donations in the past two years, and first ten principal components of ancestry. Our preliminary results identified 423 independent loci associated with 132 metabolites ($p < 5 \times 10^{-8}$). Potentially novel locus-metabolite associations were identified for *FLVCR1* and choline; and for *LPCAT3* and the lysophosphatidylserines 16.0, 18.0, 18.1, and 18.2. These associations are supported by published rare disease and mouse studies. We confirm previous metabolite GWA results for associations including N(6)-Methyl-L-lysine and *PYROXD2*, and various carnitines and *SLC22A16*. We demonstrate that it is possible to perform metabolomics-scale GWA analyses with a modest, trans-ancestry sample size. Given the many published GWA reports linking top SNPs identified in the present analysis to various blood cell traits, careful follow-up is needed to characterize these associations found in a population enriched for repeat blood donors. These data will also provide insights into the system biology of RBCs.

Complex Traits Posters - Thursday

PB1445. Genome-wide Polygenic Risk Scores in Diabetes Risk Prediction: Results from the Asian Indian Diabetic Heart Study/Sikh Diabetes Study

Authors:

D. Sanghera; Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK

Abstract Body:

Background and Aims: Genome-wide polygenic risk scores (PRS) have shown high specificity and sensitivity in predicting type 2 diabetes (T2D) risk in Europeans. However, PRS-driven information on other ethnic groups is sparse, including Asian Indians (AI), who have a heightened risk for T2D. We examined the predictive efficacy of ancestry-derived (PRSAI) and European-derived (PRSEU) with a clinical risk score (CRS) of T2D.

Methods and Results: Weighted PRSs were computed using 63 variants from (Punjabi/Sikh) genome-wide association studies and 59 variants from European meta-analyses studies in 4,602 individuals (2,574 cases and 2,028 controls) of the Asian Indian Diabetic Heart Study/Sikh Diabetes Study. Ancestry-specific PRSAI showed 23.7% improved efficacy for T2D risk prediction over PRSEU. In sensitivity analysis, the area under the curve (AUC) for PRSAI and PRSEU was 0.79 and 0.72, respectively, after adjusting for CRS. Interestingly, the AUC remained unchanged on combining the PRS (AI+EU), i.e., 0.79, suggesting a superior performance of the ancestry-specific model. Additionally, PRSAI was also associated with the increased risk for glycosylated hemoglobin (HbA1c) (Beta=0.15, $p=2.16 \times 10^{-16}$) and decreased beta-cell function (HOMA-B) (Beta=-0.32, $p=1.12 \times 10^{-32}$).

Conclusions: The ancestry-specific PRSAI demonstrated greater efficacy and a 9.7% enhanced AUC over PRSEU. More genetic evaluations on underserved non-European populations are required to increase the transferability of PRS across racial and ethnic groups and facilitate its smoother integration into clinical practice.

Complex Traits Posters - Wednesday

PB1446. Genome-Wide SNP Interaction Tests with Polygenic Risk Score for Ischemic Stroke Identifies Associations with Alzheimer Disease at *VCPKMT* and *KNDC1* by Age-Stratified Analysis

Authors:

J. Chung¹, X. Sun¹, X. Han¹, J. Mez¹, R. Mayeux², J. Haines³, M. Pericak-Vance⁴, G. Schellenberg⁵, K. Lunetta⁶, L. Farrer⁷; ¹Boston Univ., Boston, MA, ²Columbia Univ, New York, NY, ³Case Western Reserve Univ, Cleveland, OH, ⁴Univ. of Miami Miller Sch. of Med., Miami, FL, ⁵Univ Pennsylvania Sch Med, Philadelphia, PA, ⁶Boston Univ SPH Crosstown Ctr, Rm 313, Boston, MA, ⁷Boston Univ Sch Med, Boston, MA

Abstract Body:

Background: Amelioration of hypertension and other vascular risk factors has been associated with lower Alzheimer's disease (AD) risk. We aimed to identify genetic factors for AD influenced by cerebrovascular genetics using AD genome-wide association study (GWAS) with polygenic risk score (PRS) for ischemic stroke (IS). **Method:** We computed the IS PRS for European-ancestry individuals in GWAS datasets from Alzheimer's Disease Genetic Consortium and Framingham Heart Study (AD cases: 16,343, AD controls: 21,091) using the SNP-level effect estimates for IS from a MEGASTROKE GWAS. The PRS consisted of 1,438 SNPs pruned for linkage equilibrium ($r^2 < 0.4$) with minor allele frequency (MAF) $> 1\%$ and was normalized in each study. We evaluated the relationship between age (at onset for AD cases and at the last exam for controls) and the IS PRS. Next, we conducted genome-wide (GW) SNP-by-PRS interaction tests for AD in different age groups using logistic regression models adjusting for age, sex, and principal components of ancestry. Top findings were further evaluated using a transcriptome dataset in Religious Orders Study and Memory and Aging Project (ROSMAP). **Result:** We found a significant interaction effect between the IS PRS and age in all sample (OR=1.01, P=0.04) for AD. In particular, we observed significant associations between the IS PRS and AD in younger age groups defined using age cutoffs < 75 (OR=1.07, P=5.9x10⁻⁵) or < 80 (OR=1.04, P=0.01), but no association in older age groups > 75 (OR=0.98, P=0.16) or > 80 (OR=0.99, P=0.74). Following these observations, we tested GW SNP-by-PRS interactions in the age groups > 80 and < 75 and the entire sample. No genome-wide significant (GWS) interaction was detected in the total sample, but we identified GWS interactions near *VCPKMT* in the < 80 group (best SNP: rs4900982, MAF=0.44, interaction [INT] effect=-0.12, INT P=7.0x10⁻⁸) and within *KNDC1* in the > 75 group (best SNP: rs9419027, MAF=0.42, INT effect=0.13, INT P=3.3x10⁻⁸). In GTEx, rs4900982 and rs9419027 are significant expression SNPs for *VCPKMT* in the frontal cortex (P=5.0x10⁻³) and *KNDC1* in the caudate nucleus (P=5.0x10⁻⁵). In ROSMAP, we found a nominally significant association of AD risk with the interaction between *VCPKMT* expression and IS PRS among subjects < 80 (n=66, P=0.056), but not > 80 (n=572, P=0.75) or total sample (n=638, P=0.43). There were no significant interactions of the IS PRS with *KNDC1* in portion of the sample. **Conclusion:** We identified association of AD risk with interactions between two novel loci (*VCPKMT* and *KNDC1*) and IS PRS. We are evaluating the associations of these interactions with cognitive test scores and AD-related neuropathological traits.

Complex Traits Posters - Thursday

PB1447*. Genomic architecture of Autism Spectrum Disorder from comprehensive whole-genome sequence annotation.

Authors:

B. Trost¹, B. Thiruvahindrapuram¹, A. J. S. Chan², W. Engchuan¹, E. J. Higginbotham¹, J. L. Howe¹, L. O. Loureiro¹, M. S. Reuter¹, D. Roshandel¹, J. Whitney¹, M. Zarrei¹, M. Bookman³, C. Somerville¹, R. Shaath¹, M. Abdi⁴, E. Aliyev⁵, J. Sebat⁶, T. W. Frazier⁷, J. A. S. Vorstman¹, K. Fakhro⁵, B. A. Fernandez⁸, S. Lewis⁹, R. Weksberg¹, M. Fiume¹⁰, R. K. C. Yuen¹, E. Anagnostou¹¹, N. Sondheimer¹, D. Glazer¹², D. M. Hartley⁷, S. W. Scherer¹; ¹The Hosp. for Sick Children, Toronto, ON, Canada, ²Children's Hosp. of Philadelphia, Philadelphia, PA, ³Verily Life Sci., San Francisco, CA, ⁴Hamad Bin Khalifa Univ., Doha, Qatar, ⁵Sidra Med., Doha, Qatar, ⁶UC San Diego, La Jolla, CA, ⁷Autism Speaks, Princeton, NJ, ⁸Children's Hosp. Los Angeles, Los Angeles, CA, ⁹BC Children's & Women's Hlth.Ctr, The Univ. of British Columbia, Vancouver, BC, Canada, ¹⁰DNASTack, Toronto, ON, Canada, ¹¹Holland Bloorview Kids Rehabilitation Hosp., Toronto, ON, Canada, ¹²Google, Mountain View, CA

Abstract Body:

Fully understanding the genetic factors involved in Autism Spectrum Disorder (ASD) requires whole-genome sequencing (WGS), which theoretically allows the detection of all types of genetic variants. With the aim of generating an unprecedented resource for resolving the genomic architecture underlying ASD, we analyzed genome sequences and phenotypic data from 5,100 individuals with ASD and 6,212 additional parents and siblings (total n=11,312) in the Autism Speaks MSSNG Project, as well as additional individuals from other WGS cohorts. WGS data and autism phenotyping were based on high-quality short-read sequencing (>30x coverage) and clinically accepted diagnostic measures for ASD, respectively. For initial discovery of ASD-associated genes, we used exonic sequence-level variants from MSSNG as well as whole-exome sequencing-based ASD data from SPARK and the Autism Sequencing Consortium (>18,000 trios plus additional cases and controls), identifying 135 ASD-associated protein-coding genes with false discovery rate <10%. Combined with ASD-associated genes curated from the literature, this list was used to guide the interpretation of all other variant types in WGS data from MSSNG and the Simons Simplex Collection (SSC; n=9,205). We identified ASD-associated rare variants in 789/5,100 individuals with ASD from MSSNG (15%) and 421/2,419 from SSC (17%). Considering the genomic architecture, 57% of ASD-associated rare variants were nuclear sequence-level variants, 41% were nuclear structural variants (SVs) (mainly copy number variants, but also including inversions, large insertions, uniparental isodisomies, and tandem repeat expansions), and 2% were mitochondrial variants. Several of the ASD-associated SVs would have been difficult to detect without WGS, including an inversion disrupting *SCN2A* and a nuclear mitochondrial insertion impacting *SYNGAP1*. Polygenic risk scores did not differ between children with ASD in multiplex families versus simplex, and rare, damaging recessive events were significantly depleted in multiplex families, collectively suggesting that rare, dominant variation plays a predominant role in multiplex ASD. Our study provides a guidebook for exploring genotype-phenotype correlations in the 15-20% of ASD families who carry ASD-associated rare variants, as well as an entry point to the larger and more diverse studies that will be required to dissect the etiology in the >80% of the ASD population that remains idiopathic. All data resulting from this study are available to the medical genomics research community in an open but protected manner.

Complex Traits Posters - Wednesday

PB1448. Genomic discovery with functional annotation of *CNTNAP5* in the phenotypic extremes of primary angle closure glaucoma

Authors:

S. Chakraborty, S. Saha Roy, J. Sarma, T. Ahmed, S. Bhattacharjee, M. Acharya; Natl. Inst. of BioMed. Genomics, Kalyani, India

Abstract Body:

Primary angle closure glaucoma (PACG) is one of the major leading causes of blindness. Several risk factors include narrow iridocorneal angle, and increased lens thickness, leading to shallow anterior chambers in anatomically predisposed eyes. Earlier, the two comprehensive GWAS on PACG case-control identified eight novel loci where control populations were included without considering any anatomically predisposed parameters. Indeed, in India, ~30% of people show a narrow iridocorneal angle, but from here only 0.5-1% of people develop PACG. To exclude heterogeneity, we conducted an age-dependent model of progressive angle closure GWAS where relatively early-onset PACG patients (PACG: age ≤ 50 years) compared with anatomically predisposed (narrow iridocorneal angle $< 15^\circ$) non-glaucomatous older individuals (PACS: age ≥ 60 years) to improve power to detect genetic factors for glaucomatous neurodegeneration. In our discovery phase, we identified genomic alterations using a genome-wide association study in 148 PACG and 92 PACS individuals. A total of 11 genome wide suggestive significant ($P < 1e^{-05}$) SNPs were found. Of them, 5 loci were mapped at *TLL2*, *CNTNAP5*, *TNF*, *PAPLN* and *MCCDI*. Genotype-phenotype correlation prioritized rs780010 of *CNTNAP5*, which was significantly associated with the cup to disc ratio ($P = 0.0024$), which is a clinical parameter directly correlated with glaucomatous neurodegeneration. Conditional analyses identify 13 loci of *CNTNAP5* associated with PACG. We further validated the sentinel SNP rs780010, with $P = 2.131e^{-06}$, in a separate replication cohort (PACG = 50; PACS = 39) and observed a significant association (odds ratio (OR) = 2.307, $P = 0.012$). Integrative bioinformatics analyses suggested higher retinal neuronal expression of *CNTNAP5* with active enhancer marks, which were subsequently validated through a luciferase assay. Further, translation blocker morpholinos against *cntnap5a* and *cntnap5b* were injected into zebrafish embryos at a single cell stage (2.5ng each). There were significant eye size reductions and retinal nerve thinning was observed in zebrafish morphants in comparison to 5 base mismatch-pair control morphants. In conclusion, genomic discovery results not only indicate a genomic association of the *CNTNAP5* with PACG but also imply that it may play a role in glaucomatous neurodegeneration. Further, functional annotation results lead us to hypothesize that *CNTNAP5* is the candidate gene to perturb the development of the neural retina, leading to reduced eye size and neurodegeneration of the retinal ganglion cells, thereby increasing the risk of PACG-associated vision loss.

Complex Traits Posters - Wednesday

PB1449. Genomic structural equation modeling to unravel kidney-specific variants from multiple biomarkers

Authors:

R. Fujii^{1,2,3}, C. Pattaro¹; ¹Eurac Res., Bolzano, Italy, ²Fujita Hlth.Univ., Toyoake, Japan, ³Nagoya Univ., Nagoya, Japan

Abstract Body:

Genome-wide association studies (GWAS) of kidney function have focused almost exclusively on the estimated glomerular filtration rate derived from serum creatinine (eGFR_{crea}), which may also reflect muscle mass metabolism. To identify kidney-specific loci, researchers typically compare the GWAS results with those from alternative kidney function markers such as cystatin C-based eGFR (eGFR_{cys}), uric acid (UA), and blood urea nitrogen (BUN). Assuming the existence of an underlying kidney function trait affecting the levels of multiple markers, we applied genomic structural equation modeling (GenomicSEM) combining publicly available univariate GWAS results of eGFR_{crea} (n=1,201,909), eGFR_{cys} (n=460,826), UA (n=457,690), and BUN (n=852,678) from the CKDGen Consortium. Z-scores of the results were compared against 424 lead variants of each locus in the previous GWAS for eGFR_{crea}. The merged summary statistics across four previous GWAS included >6,000,000 SNPs. The highest genetic correlation was observed between eGFR_{crea} and eGFR_{cys} ($r_g=.593$, $se=.041$), followed by eGFR_{cys}-UA ($r_g=-.393$, $se=.037$), eGFR_{cys}-BUN ($r_g=-.387$, $se=.042$), eGFR_{crea}-BUN ($r_g=-.364$, $se=.044$), eGFR_{crea}-UA ($r_g=-.217$, $se=.035$), and UA-BUN ($r_g=.211$, $se=.039$). Confirmatory factor analysis (CFA) showed good fitting (comparative fit index=.936, root mean square residual=.028) and identified factor loadings (λ) for each biomarker: $\lambda_{eGFR_{crea}}=.654$; $\lambda_{eGFR_{cys}}=.896$; $\lambda_{BUN}=-.405$; and $\lambda_{UA}=-.479$. Of the 373 available SNPs which were significant in the previous eGFR_{crea} GWAS, the absolute Z-score of the common latent kidney trait were bigger for 119 variants (31.9%). Of these, 112 variants (94.1%) showed significant effects on eGFR_{crea}, eGFR_{cys} and BUN, underlying their kidney function relevance. Among 254 variants with smaller absolute Z-score in the latent kidney trait analysis, only 86 (33.9%) were significantly associated with eGFR_{crea}, eGFR_{cys} and BUN. These findings highlight the strong enrichment of loci directly relevant to the kidney obtained with GenomicSEM. In the absence of a kidney function-specific marker, GenomicSEM analysis shows the potential to prioritize kidney function-relevant loci by combining multiple kidney-related biomarkers.

Complex Traits Posters - Thursday

PB1450. GenoPipeR: An Automated Pipeline for Predictive Modelling of Complex Phenotypes using Genome-wide Data

Authors:

V. Marshe¹, S. Elsheikh², E. Tio², M. Maciukiewicz³, A-C. Hauschild⁴, X. Men², D. Felsky², D. J. Mueller²; ¹Columbia Univ. Irving Med. Ctr., New York, NY, ²Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, ³Univ. Hosp. Zurich, Zurich, Switzerland, ⁴Dept. of Med. Informatics, Univ. Med. Ctr. Göttingen, Göttingen, Germany

Abstract Body:

Machine Learning (ML) offers novel opportunities for modelling complex traits using high-dimensional SNP data. However, rigorous ML analyses are challenging and require the integration of ML best practices with modality-specific expert knowledge. We present the *genoPipeR* R package: an automated, user-friendly ML pipeline for complex trait modelling using SNP and covariate data. In contrast to SNP-modelling pipelines with custom architectures, *genoPipeR* is a standardized, supporting *PLINK*-formatted files and allows the user to select among state-of-the-art ML models (R caret) and types of cross-validation (CV) for model selection and hyperparameter tuning. The pipeline automatically performs feature selection and modelling with consideration for genetic architecture. To evaluate *genoPipeR*, we built models for beta-amyloid levels and Alzheimer's disease consensus diagnosis, adjusted for demographic and batch covariates in the ROS/MAP cohort (N = 2,067). For genetic features, we prioritized across SNPs using PC-corrected genome-wide GLMs. A final set of the top 25 clinical and genetic features were selected based on non-zero elastic net coefficients and passed to model selection and hyperparameter tuning. We assessed seven models (GLM, CART, RF, GBM, linear/radial SVM, kernel PLS) and two CV types (5-fold vs. nested 5-fold). Final models were tested in a 30% holdout. For genome-wide models (9,329,439 SNPs before prioritization), the computational performance was reasonable (7-37 mins). For beta-amyloid levels, our best model was a non-nested GLM model ($RMSE = 0.99$, $MAE = 0.82$, $adj. R^2 = 0.17$; actual vs. predicted, Spearman $\rho = 0.42$, $p = 8.52 \times 10^{-18}$) including top features rs11728484, rs2215637, age at death, *APOE* rs1065853, and *ZDHHC21* rs113159914. Other model-selected SNPs were in *KCNT2*, *APOC1* and *APOE* $\epsilon 4$ (rs429358). Compared to a non-ML linear regression model including sex, age at death, batch effects and *APOE* $\epsilon 4$ status ($R^2 = 0.12$), our model achieved a 5% higher explained variance. A non-nested kernel PLS model showed the best performance for Alzheimer's disease status ($AUC = 0.68$; $Acc = 0.62$, $p = 4.01 \times 10^{-4}$; $Sens = 0.68$, $Spec = 0.65$). The top predictors were age at death, *APOE* rs429358, *NECTIN2* rs12972156, rs6956154, and *BCL2L14* rs1641717. Our analysis shows that *genoPipeR* can identify features at the genome-wide level that have previously been associated with a complex outcome, as well as potentially interesting features for further investigation. The pipeline is suitable for local and distributed computing settings, depending on the sample size, and integrates extensive post-processing, including publication-ready plots.

Complex Traits Posters - Wednesday

PB1451. Genotypic and Phenotypic Heterogeneity in Prader-Willi Syndrome: A Bangladeshi cohort study

Authors:

R. Mim¹, H. Akter^{1,2}, M. Rahman³, S. Sarker^{1,4}, M. Basiruzzaman^{1,5}, M. Chowdhury^{1,6}, M. Rahaman¹, T. Eshaque¹, M. Morshed¹, A. Islam^{1,7}, G. Kundu⁸, S. Kanta⁴, R. Biswas⁹, K. Uddin^{1,10}, M. Uddin¹¹; ¹Genetics and Genomic Med. Ctr., NeuroGen Hlth.care, Dhaka, Bangladesh, Dhaka, Bangladesh, ²Dept. of Biochemistry and Molecular Biology, Univ. of Dhaka, Dhaka, Bangladesh, ³Dept. of Paediatric Neurology, Bangabandhu Sheikh Mujib Med. Univ., Dhaka, Bangladesh, Dhaka, Bangladesh, ⁴Dept. of Paediatric NeuroSci., Dhaka Shishu Hosp., Dhaka, Bangladesh, Dhaka, Bangladesh, ⁵Dept. of Neurology, Natl. Inst. of NeuroSci.s and Hosp., Dhaka, Bangladesh, Dhaka, Bangladesh, ⁶Natl. Inst. of NeuroSci.s, Dhaka, Bangladesh, Dhaka, Bangladesh, ⁷Cellular Intelligence Lab, GenomeArc Inc., Toronto, ON, Canada, ⁸Dept. of Child Neurology, Bangabandhu Sheikh Mujib Med. Univ., Dhaka, Bangladesh, Dhaka, Bangladesh, ⁹Dept. of Paediatric Endocrinology Dhaka Shishu (Children) Hosp. Shere Bangla Nagar, Dhaka, Bangladesh, Dhaka, Bangladesh, ¹⁰Dept. of Biochemistry, Holy Family Red Crescent Med. Coll., Dhaka, Bangladesh, Dhaka, Bangladesh, ¹¹Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates

Abstract Body:

Background: Prader-Willi syndrome (PWS) is a neurodevelopmental disease characterized by aberrant imprinted gene expression in the 15q11-13 region. Dysregulation of the genes located in this region shows variable clinical phenotypes such as Prader-Willi syndrome (PWS, #176270), Angelman syndrome (AS, #105830) and 15q11-q13 duplication syndrome (Dup15q syndrome, #608636). In this study, we applied chromosomal microarray analysis (CMA) to detect chromosomal abnormalities throughout the genome of the clinically suspected Prader-Willi syndrome patients.

Methods: We have conducted genome-wide CMA (642,824 probes spanning the genome) analysis for 6 clinically suspected Prader-Willi patients to identify chromosomal abnormalities (deletion/duplication/translocation and rearrangements). For the negative patient, we used Whole Exome Sequencing (WES) to uncover single nucleotide variants (SNVs), small insertion and deletion. We have used human genome build GRCH38/UCSC hg38 as reference. Variant classification analysis was conducted based on the American College of Medical Genetics (ACMG) guidelines.

Results: Out of the 6 patients, 5 patients were carrying different lengths of pathogenic deletion and duplication CNVs in the 15q11.2q13.1 region of chromosome 15. The overlapping region is 4.92Mb (Chr15:23370969-28289373). The diagnostic yield of CMA in clinically suspected Prader-Willi syndrome patients is 83.33%. No clinically relevant variants were found in one patient. This patient was further included in WES. From WES, we have found a variant of uncertain significance (VOUS) c.1642T>A (p.Tyr548Asn) in the *PPP2R5D* gene which is associated with Intellectual developmental disorder 35 (MRD35, #616355).

Conclusion: In Bangladesh, CMA and WES testing are not yet implemented in clinical practice as a first-tier diagnostic test for Prader-Willi syndrome. Our results from this study show the utility of CMA and WES for the precise genetic diagnosis and its integration in clinical practice.

Key words: Prader-Willi syndrome (PWS); Chromosomal microarray analysis (CMA); Whole Exome Sequencing (WES); Single nucleotide variants (SNVs); American College of Medical Genetics (ACMG).

Complex Traits Posters - Thursday

PB1452. Global analysis of RNA editing in Alzheimer's disease across multiple brain regions.

Authors:

E. Huang, M. Choudhury, X. Xiao; UCLA, Los Angeles, CA

Abstract Body:

RNA editing is a biological process that refers to the alteration of RNA transcripts through insertion, deletion, or substitution of nucleotides. The nucleotide changes induced by RNA editing events have been shown to regulate gene expression through a diverse range of mechanisms, such as altering protein-coding sequences, splice sites, and sequences related to microRNA gene-targeting. It is increasingly recognized that RNA editing is globally dysregulated in many neurological disorders. Here, we aim to gain a comprehensive understanding of the relationship between RNA editing and Alzheimer's Disease (AD) pathology. We utilized bulk RNA-seq data from two large-scale patient cohorts: the Mount Sinai Brain Bank (MSBB) study and the ROS/MAP project. Each cohort encompass samples from hundreds of individuals across multiple brain regions, including the prefrontal cortex, superior temporal gyrus, parahippocampal gyrus, and inferior frontal gyrus. Using a computational pipeline designed for de novo identification of editing sites and quantification of their corresponding editing levels, we identified over hundreds of thousands of editing sites in each brain region. Thousands of sites were found to have significantly different editing levels between AD and control samples, among which are protein-recoding sites in various genes that have been previously implicated in Alzheimer's Disease pathology. To better understand regulators of editing levels, we identified RNA binding proteins (RBPs) that significantly correlate with the editing levels across the samples in each brain region. Finally, we performed ordinal regression analyses to identify sites with editing levels that are strongly related to the severity of three measures of Alzheimer's Disease pathology: cognitive status, neurofibrillary tangle levels, and plaque density. Our work presents a systematic view of how RNA editing contributes to multiple facets of AD progression and emphasize the importance of its consideration in future studies that aim to better understand the disease's biological underpinnings.

Complex Traits Posters - Wednesday

PB1453. Global Long COVID Host Genetics Initiative identifies *FOXP4* locus as the first genetic risk factor associated to Long COVID

Authors:

V. Lammi¹, S. E. Jones¹, T. Nakanishi^{2,3}, Long COVID Host Genetics Initiative, H. M. Ollila^{1,4,5,6}; ¹Inst. for Molecular Med. Finland, University of Helsinki, Finland, ²McGill Univ., Montreal, QC, Canada, ³Kyoto-McGill Intl. Collaborative Sch. in Genomic Med., Graduate Sch. of Med., Kyoto Univ., Kyoto, Japan, ⁴Broad Inst. of MIT and Harvard, Cambridge, MA, ⁵Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁶Anesthesia, Critical Care, and Pain Med., Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA

Abstract Body:

Introduction: While most people recover from COVID-19 within 3 months, a substantial proportion suffer from long-term post COVID-19 conditions or so-called “Long COVID”. The most common symptoms include fatigue, cognitive dysfunction and breathlessness, though symptoms can vary widely by proband, and severity can range from mild to debilitating. To understand the disease etiology and genetic liability behind Long COVID, we built a global working group under the COVID-19 Host Genetics Initiative.

Materials and methods: At present (data freeze 2), 20 studies from 14 countries representing six genetic ancestry groups contributed data to our meta-analyses. Long COVID was primarily captured using questionnaires, with Long COVID being defined as ‘any symptoms that cannot be explained by alternative diagnoses, or impact on everyday functioning, 3 months after the onset of COVID’. In studies with electronic health record (EHR) data, specific diagnosis codes for Post COVID-19 conditions (ICD-10 code U09(.9), or corresponding SNOMED codes) were used.

Results: We performed a GWAS meta-analysis comparing Long COVID cases (N=2,592) to population controls (N=942,312) and identified a genetic risk locus on chromosome 6 upstream of the *FOXP4* gene (rs9367106, $P = 7.3e-9$, OR = 1.6[1.4-1.9], minor allele frequency = 0.04). The lead variant is in linkage disequilibrium with another variant previously associated with COVID-19 severity (rs1886814, $r^2=0.62$, $D'=0.93$), showing the same direction of effect for the risk haplotype. At present, none of the other reported lead variants for COVID-19 severity showed associations with Long COVID. In addition, comparing Long COVID cases (N=3,877) to COVID+ controls without Long COVID (N=28,446), showed a nominal association (rs9367106, $P = 0.046$, OR = 1.2[1.0-1.4]).

Conclusions: Our early findings indicate that *FOXP4* may affect susceptibility to Long COVID. The transcription factor (Forkhead box protein P4) encoded by *FOXP4* is involved, inter alia, in antigen-specific recall responses in T cells, and lung secretory epithelium regeneration. Thus, this association suggests a biological link for infection-related immunological memory and lung mucus functions with the development of Long COVID. Furthermore, our data support epidemiological findings that connect Long COVID with initial COVID-19 disease severity. The Long COVID Host Genetics Initiative is continuing to recruit additional studies from diverse ancestries and incorporate information from ongoing data collection efforts within existing contributors to better understand the genetic factors predisposing to Long COVID and to explore its effect on long-term health

Complex Traits Posters - Thursday

PB1454. GWAS and Polygenic prediction of anthropometric traits in sub-Saharan African populations

Authors:

S. Adebamowo¹, A. Adebayo², P. Achara³, B. Alabi⁴, R. Bakare⁵, G. Chen⁶, A. Doumatey⁷, A. Famooto⁸, K. Obende⁹, R. Offiong¹⁰, O. Olaniyan¹¹, S. Ologun¹², J. Zhou⁷, A. Adeyemo¹³, C. Rotimi¹³, C. Adebamowo¹, ACCME Research Group as part of the H3Africa Consortium; ¹Univ. of Maryland Sch. of Med., Baltimore, MD, ²Asokoro District Hosp., Abuja, Nigeria, ³Federal Med. Ctr. Keffi, Keffi, Nigeria, ⁴Wuse Gen. Hosp., Abuja, Nigeria, ⁵Univ. of Ibadan, Ibadan, Nigeria, ⁶NHGRI/CRGGH, Bethesda, MD, ⁷NIH/NHGRI, Bethesda, MD, ⁸Inst. of Human Virology Nigeria, Abuja, Nigeria, ⁹Garki Hosp. Abuja, Abuja, Nigeria, ¹⁰Univ. of Abuja Teaching Hosp., Abuja, Nigeria, ¹¹Natl. Hosp. Abuja, Abuja, Nigeria, ¹²Kubwa Gen. Hosp. Abuja, Abuja, Nigeria, ¹³NIH, Bethesda, MD

Abstract Body:

Background: Evidence suggests that polygenic risk scores (PRS) for complex traits which are currently mostly derived from European-ancestry (EA) populations perform poorly in other populations because of underrepresentation in the underlying genome-wide association studies (GWAS) data, differences in allele frequencies, and effect sizes (among other factors). **Methods:** We conducted GWAS, meta-analyses and gene-set analyses of height, body-mass index (BMI), and waist-hip ratio (WHR), and generated and compared PRS from EA (n~694,000), Multiethnic ancestry (MEA, n~49,300), African American and Afro-Caribbean ancestry (AA/AC, n~41,300) and Sub-Saharan African (AFR, n~14,000) populations. We used the African Collaborative Center for Microbiome and Genomics Research data (ACCME, n~15,000) for GWAS discovery and PRS target, and the Africa America Diabetes Mellitus study data (AADM, n~5,200), for GWAS replication and PRS validation. **Results:** The top replicated variants associated with each trait were in *HMGAI* (rs6937622, p=3.70 x 10⁻¹², for height), *MDGAI* (rs188795185, p=1.18 x 10⁻⁷, for BMI) and *HAAO* (rs140002152, p=1.95 x 10⁻⁹, for WHR) gene regions. The EA-derived PRS had the highest proportion of variance explained (R^2) for the selected anthropometric traits in SSA populations, compared to the MEA, AA/AC and AFR derived PRS. The ancestry-specific parameters for the best predictive PRS were EA ($R^2=2.39\%$, p=4.24x10⁻⁵⁶), AA ($R^2=2.08\%$, p=5.12x10⁻⁴⁹) and AFR ($R^2=0.54\%$, p=1.89x10⁻¹³) for height; EA ($R^2=1.21\%$, p=2.19x10⁻²⁸), AFR ($R^2=0.18\%$, p=2.49x10⁻⁰⁵) and MEA ($R^2=0.79\%$, p=4.30x10⁻¹⁹) for BMI; and EA ($R^2=0.59\%$, p=1.19x10⁻¹⁴), AFR ($R^2=0.06\%$, p=0.018) and MEA ($R^2=0.06\%$, p=0.013) for waist-hip ratio. The results of comparing participants in the highest PRS quantile versus those in the lowest quantile were consistent with these findings. **Conclusions:** PRS derived from large-scale EA populations had high predictability for anthropometric traits in SSA populations. For the selected anthropometric traits, PRS prediction R^2 improved with increasing sample size, indicating the need for more genomic data from diverse populations to serve as a basis for PRS.

Complex Traits Posters - Wednesday

PB1455. GWAS of the age-of-onset of type 1 diabetes reveals HTATIP2 as a novel T cell regulator.

Authors:

C. Cardinale¹, X. Chang², H. Qu², H. Hakonarson¹; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

The age at which type 1 diabetes first presents is a parameter of clinical interest. We conducted a quantitative trait genome-wide association study of type 1 diabetes age-of-onset in a discovery cohort of 4,014 cases and a replication cohort of 493 cases. Two loci were genome-wide significant: the previously-identified signal in the human leukocyte antigen locus on chromosome 6, and a novel signal in the *NELL1* gene on chromosome 11. Bayesian fine-mapping nominates rs10833518 as the causal SNP, and homozygosity for the risk allele is associated with an average one-year earlier onset. The risk allele of this SNP is bound by an allele-specific protein complex in gel shift. Four genes in this locus are expressed in immune cells, and we used siRNA knock-down of these candidates with RNA-seq in primary human CD4⁺ T cells to determine that inhibition of HTATIP2 reduces expression of kinases, receptors, and transcription factors involved in signal transduction. Functionally, knock-down of HTATIP2 in these cells increases viability and reduces expression of activation markers. Taken together, we nominate HTATIP2 as a new type 1 diabetes risk gene acting via T cell regulation.

Complex Traits Posters - Thursday

PB1456. Heterozygous loss-of-function mutation in *SORL1* causes neuronal dysregulation of endosomal trafficking and processing of APP.

Authors:

B. DeRosa¹, S. Simon¹, B. Kunkle¹, R. Carney^{1,2}, M. Cuccaro^{1,3}, H. Cukier^{1,4}, J. Vance^{1,3,4}, M. Pericak-Vance^{1,3,4}, D. Dykxhoorn^{1,3}; ¹John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, ²Miami VA Med. Ctr., Miami, FL, ³Dr. John T. Macdonald Fndn. Dept. of Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, ⁴Dept. of Neurology, Univ. of Miami Miller Sch. of Med., Miami, FL

Abstract Body:

Dysregulation in endolysosomal trafficking has emerged as a key pathogenic process in both sporadic and inherited forms of Alzheimer's disease (AD). *SORL1*, is a top AD risk gene that encodes a sorting receptor involved in amyloid precursor protein (APP) endosomal trafficking and the secretion of amyloid- β (A β). *SORL1* rare protein-truncating variants have been found to be causal in late onset AD and are enriched in cases with early-onset AD (EOAD), with some families exhibiting autosomal-dominant patterns of inheritance. We previously identified siblings affected with EOAD that had a rare heterozygous frameshifting deletion in *SORL1* (c.4293delC; rs1343336951) that results in a truncated protein lacking ~30% of the C-terminus (p.C1431WfsX2). To further understand the role of *SORL1* in AD pathogenesis, we first derived induced pluripotent stem cell (iPSC) lines from the two EOAD-affected siblings bearing the c.4293delC mutation. Next, we created isogenic control iPSC lines via CRISPR/Cas9-mediated homology-directed repair knock-in of the wildtype allele. These iPSC lines were evaluated for pluripotency, genomic stability, and the lack of off-target genome editing resulting from the CRISPR-based gene correction. Once validated, the isogenic pairs of *SORL1*^{+/-} and deletion-corrected control iPSC lines were differentiated into cortical neurons. Our data show that heterozygous loss of *SORL1* function leads to enlarged endosomes and the accumulation of APP in early endosomes. Additionally, we observe *SORL1*^{+/-} neurons have increased levels of secreted A β 42, a pathogenic isoform of A β that is highly fibrillogenic and enriched in senile plaques. In conclusion, our results indicate that loss of function in a single copy of *SORL1* is sufficient to induce neuronal defects in APP trafficking and processing and dysregulation of A β production. This study broadens our understanding of the contribution of *SORL1* mutations to AD and provides insight into possible early neuropathogenic effects of EOAD-associated mutations. These data, together with similar findings in neurons with other EOAD gene loss-of-function mutations, support an important mechanistic role for disturbances in endosomal APP trafficking and A β production in AD pathophysiology.

Complex Traits Posters - Wednesday

PB1457. Heterozygous *RYR2*, *ABCG8*, *PIK3C2G*, and *RASSF9* variants as the cause of premature coronary artery disease in five Iranian families.

Authors:

S. Mehvari¹, M. Asadnezhad¹, N. Karimian Fathi¹, F. Ghodrattpour¹, S. Saki², K. Hosseini³, S. Sadeghian³, M. Akbari⁴, R. Malekzadeh², R. Najafipour¹, **H. Najmabadi**¹, K. Kahrizi¹; ¹Genetics Res. Ctr., Univ. of Social Welfare and Rehabilitation Sci., Tehran, Iran, Islamic Republic of, ²Digestive Disease Res. Inst., Tehran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, ³Dept. of Cardiology, Tehran Heart Ctr., Tehran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, ⁴Women's Coll. Res. Inst., Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Background/Objective: Coronary artery disease (CAD), the major cause of death globally, is a complex multifactorial disorder influenced by both common and rare genetic variants. Several risk factors including family history has long been recognized to increase the risk of CAD development. In this study, we aimed to identify the potential rare disease-causing variants in an Iranian cohort of familial premature CAD (PCAD). **Methods:** Whole exome sequencing (WES) was performed on proband of each family followed by Sanger sequencing for confirmation of the results and co-segregation analysis. **Results:** We have hitherto identified heterozygous probably disease-causing variants in genes *RYR2*, *ABCG8*, *PIK3C2G*, and *RASSF9* in five unrelated families. Mutations in *RYR2* are known to cause cardiac arrhythmia and sudden cardiac death (SCD). We identified different heterozygous missense variants in *RYR2* gene in two families, i.e. c.5189C>T (p.Thr1730Met) and c.9778C>T (p.Arg3260Thr). The former was detected in a 58-year-old female with a history of chest pain and the onset of established CAD at 47 years of age. The second variant was identified in a female with a history of cardiac attacks twice, at 23 and 49 years old, with a positive history of SCD in her family. In the third family, we identified a heterozygous likely pathogenic (LP) variant, c.562G>C (p.Val188Lys), in *ABCG8* gene which is involved in sterol metabolism in a 38-year-old female with a history of chest pain, hyperlipidemia and the onset of established CAD at 33 years of age. In the other family in which the male proband presented with the onset of established CAD at the age of 48, we identified a novel heterozygous LP splice variant, c.3357+1G>T, in *PIK3C2G* gene which is involved in promoting AKT signaling, hence insulin signaling pathway, glucose homeostasis and glycogen synthesis. In the last family, we identified a frameshift variant, c.893_894del (p.Ile298LysfsTer2), in *RASSF9* gene in a 44-year-old male with a family history of PCAD. Finally, for apparently healthy heterozygous members of these families, coronary CT angiography confirmed the genotype and showed mild to/or moderate coronary stenosis. **Conclusion:** Rare variants could be responsible for the pathogenesis of PCAD in affected individuals with a positive family history of such disease. Identifying these variants can help to prevent CAD morbidity and mortality in asymptomatic patients. The findings of this study may contribute to a better understanding of the genetic components of CAD and reflect the need to take rare variants into consideration. However, further studies are needed to support the role of *PIK3C2G* and *RASSF9* genes in CAD development.

Complex Traits Posters - Thursday

PB1458. HLA Haplotype Analysis and Genome Wide Association Study of Nephrotic Syndrome in the Million Veteran Program

Authors:

A. Hung^{1,2}, G. Wang^{1,2}, H-C. Chen^{1,2}, E. Phillips¹, C. Chung^{1,2}, J. Hellwege^{1,2}, O. Wilson^{1,2}, B. R. Gorman³, S. Pyarajan^{4,5}, J. E. Huffman⁶, M. Salani¹, E. Fagan¹, C. Fang¹, J. Triozzi¹, E. Siew^{1,2}, Million Veteran Program, A. Bick^{7,1}, C. Robinson-Cohen¹, R. Tao^{1,2}; ¹Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Nashville VA, Tennessee Valley Hlth.care System, Nashville, TN, ³VA Boston Hlth.care System, Boston, MA, ⁴VA Cooperative Studies Program, VA Boston Hlth.care System, Boston, MA, ⁵Dept. of Med., Brigham and Women's Hosp. and Harvard Sch. of Med., Boston, MA, ⁶Massachusetts Veterans Epidemiology Res. and Information Ctr., VA Boston Hlth.care System, Boston, MA, ⁷Vanderbilt, Nashville, TN

Abstract Body:

Nephrotic syndrome (NS), is characterized by massive urinary excretion of protein, is caused by damage to the glomerular filtration barrier, and leads to heightened morbidity and mortality and faster progression to ESRD. Discovery of new therapies is urgently needed.

We identified veterans with NS, 1640 European ancestry (EA), and 924 African ancestry (AA). Controls included individuals with GFR > 60 ml/min and a negative UA, 218,300 EA, and 61,921 (AA). We performed a GWAS using logistic regression, an additive model (MAF > 1%), imputed to the 1000 Genomes reference panel, adjusted for age, sex, and the 10 PCs, stratified by race. A second model also adjusted for diabetes. HLA haplotypes were built by the MVP Genomic core using HIBAG with estimated accuracy >90%. For HLA haplotypes additive and dominant models were performed we here report the results for the additive model. We used Bonferroni adjustment for p-values.

Mean age was 59 (SD14), 89% were males. 31% of the cases were AA compared to 20% of the controls. 67% of the cases had diabetes while only 25% of the controls. 178 patients had FSGS (6%), 77 (3%) membranous (29 had PLA2R +), 14 (0.5%) crescentic GN, 11 (0.4%) MPGN, 6 (2%) diffuse or mesangio-capillary, and 4 (1%) dense deposits. 86% were NS of unspecified morphology. Our GWAS results for the EA patients showed an extensive GWAS association with different components of the HLA (979 SNPs). In the haplotype analysis, there were two risk HLA haplotypes in EA participants without diabetes in HLA class II; DQA1*05:01 (p=1.19E-06)- DRB1*03:01 (p=0.0001) and there were four risk HLA haplotypes in EA participants with diabetes in both HLA class I & II; DQA1*05:01 (p=3.36E-06)- DRB1*03:01 (p=1.56E-05)- DPB1*01:01(6.25E-05) C*07:01(p=0.0001). No risk HLA haplotypes were identified in AA participants. Other independent loci associations in EA were:

known: *TSBP1* (p=9.01x10-09), *APOM* (p=1.41x10-08), *PRRC2A* (p=1.52x 10-08).

Novel: *TNF* (p=3.08x10-08), *TNXB* (p=1.47x10-08), *VWA7* (4.10x10-08), *NOTCH4* (1.65E-8). For AA patients, association were *APOLI*; (rs73885319 p= 1.59x10-09, rs60910145; p= 1.95x10-09), and *RNF2* (p= 4.82x10-08).

Our GWAS and HLA haplotype analyses confirmed the strong association known between HLA-DQA1 and NS in EA individuals, particularly for membranous. It also confirmed the association of other HLA haplotypes and NS supporting the role of the immune response in this clinical condition. Our study identified novel associations with Von Willebrand factor, Tenascin, and NS. In AA, we confirmed the association between *APOLI* and proteinuric diseases like FSGS.

Complex Traits Posters - Wednesday

PB1459. HLA in autoimmune encephalitis.

Authors:

T. K. de Araujo^{1,2}, D. C. Rosa¹, F. R. Torres¹, W. Watanabe³, C. L. Yasuda^{3,2}, A. C. Coan³, F. Cendes^{3,2}, I. Lopes-Cendes^{1,2}; ¹Dept. of Translational Med., Sch. of Med. Sci., Univ. of Campinas (UNICAMP), Campinas, SP, Brazil., Campinas, Brazil, ²Brazilian Inst. of NeuroSci. and Neurotechnology (BRAINN), Campinas, SP Brazil., Campinas, Brazil, Campinas, Brazil, ³Dept. of Neurology, Sch. of Med. Sci., Univ. of Campinas (UNICAMP), Campinas, SP, Brazil., Campinas, Brazil

Abstract Body:

Autoantibody-mediated forms of encephalitis (AE) include neurological disorders characterized by subacute memory loss, movement disorders, and focal seizures. The Human Leucocyte Antigen (HLA) genes are involved in susceptibility to more than 100 inflammatory, infectious, and autoimmune diseases. HLA is among the most polymorphic regions of the human genome and presents considerable diversity among populations. Given that there are still controversies regarding the role of specific HLA alleles in AE we aim to search for HLA associations in a cohort of well-characterized patients with AE. To date, we have evaluated 36 patients: 11 with anti-NMDA-R antibodies, 6 with anti-GAD65 antibodies, 6 with paraneoplastic encephalitis, 1 with LGI-1 antibodies, 2 with anti-TPO antibodies, 6 without antibodies detection, and 5 patients with Rasmussen encephalitis. We also included data from 297 ethnically matched controls. The DNA from these individuals was extracted from peripheral blood or saliva samples and was genotyped for HLA using next-generation sequencing technology. The DNA libraries were loaded onto a MiSeq Sequencer (Illumina), and the data were analyzed with the NGSengine v. 2.25 software (GenDx). HLA genotyping was reported at a six-digit level whenever technically possible for HLA -A, -B, -C, -DRB1, -DRB3, -DRB4, -DRB5, -DPA1, -DPB1, -DQA1, -DQB1, and -G, but for comparison with previous published work, we set the results to two digits. Statistical analysis of relative allele frequencies was performed using an Excel spreadsheet; p-value and odds ratio (OR) were calculated using the R studio. The most common HLA alleles in AE was *HLA-A*02:01P* (0.20), *HLA-B*07:02P* (0.14), *HLA-C*04:01P* (0.19), *HLA-DPA1*01:03P* (0.63), *HLA-DPB1*04:01P* (0.29), *HLA-DQA1*01:02P* (0.29), *HLA-DQB1*02:01P* (0.23), *HLA-DRB1*07:01P* (0.16), *HLA-G*01:01P* (0.66). Most interestingly, we found a significant difference in the frequency of *HLA-B*07:02P* allele when comparing patients to controls, p-value=0.01299, 95%CI:1.134062-6.477955, OR:2.839002. These results may indicate a potential role for HLA-B in the predisposition to AE in the population studied.

Complex Traits Posters - Thursday

PB1460. HLA types are associated with hundreds of complex traits and diseases in an ancestry-dependent manner

Authors:

M. D'Antonio¹, D. Perry¹, A. Massarat¹, J. Margoliash², M. Gymrek³, K. Frazer⁴; ¹Univ. of California, San Diego, LA JOLLA, CA, ²Univ. of California, San Diego, San Diego, CA, ³Univ California San Diego, La Jolla, CA, ⁴UC San Diego, SAN DIEGO, CA

Abstract Body:

The 4 Mb major histocompatibility (MHC) region on chromosome 6p21.3, which encodes the human leukocyte antigen (HLA) gene complex, is highly polymorphic, gene dense (383 genes), and exhibits strong linkage disequilibrium. Genome-wide association studies (GWAS) have identified many traits and diseases with significant associations in the MHC region; however, ~90% of GWAS have been performed on European individuals. Additionally, GWAS have primarily focused on common SNPs, and omitted analysis of whole HLA genes alleles (HLA types), which can be poorly tagged by SNPs. For these two reasons, we believe many genetic associations located in the MHC region have yet to be discovered. We utilized 479,806 individuals in the UK BioBank (UKBB), including 443,019 individuals of European, 7,746 South Asian, 7,745 African and 2,366 East Asian descent, as well as 18,931 admixed individuals. We examined associations between the genotypes of 346 imputed HLA types at 4-digit resolution (distinguishing nonsynonymous variants between alleles) for 11 HLA genes and >75,000 SNPs in the MHC region and 1,213 traits, including ICD9 and ICD10 codes and blood biochemistry markers, and observed genome-wide significant associations for 419 traits, including 60 associations that are European-specific, 44 that are African-specific and 70 that are East Asian-specific. This indicates that the MHC region has strong ancestry-specific components. To understand the differences between genetic associations for 4-digit HLA types and SNPs, we performed fine-mapping using SuSiE on each genome-wide significant trait and found 2,215 independent signals. For 37 signals, the variant with the strongest posterior inclusion probability was an HLA type, indicating that in certain cases HLA types explain more trait variation than each single SNP. These cases include the previously reported association between HLA-DRB1*01:03 and ulcerative colitis across all individuals, as well as multiple ancestry-specific associations, including: HLA-C*15:02 and kidney failure in Africans; HLA-DRB1*04:04 and aortic valve disorders in East Asians; HLA-B*35:01 and lymphocyte percentage in Europeans; and HLA-B*14:01 and C-reactive protein levels in South Asians. Our study shows that the MHC region is associated with hundreds of traits in an ancestry-specific fashion, and that many of these associations are likely driven by HLA types rather than single SNPs. Expanding HLA types from 4-digit to 8-digit resolution (which distinguish nonsynonymous, synonymous, and non-coding regulatory variants between alleles) will likely identify hundreds more associations with specific HLA gene alleles.

Complex Traits Posters - Wednesday

PB1461. HLA typing and COVID-19 disease outcome associations: GENCOV Study Canada.

Authors:

E. Frangione^{1,2}, D. Di Iorio^{1,2}, J. J. Lee^{1,2,3}, S. Casalino^{1,2}, S. Chowdhary^{1,2}, C. Mighton^{1,2,3,4}, H. Faghfoury⁵, Y. Bombard^{3,4}, L. Strug⁶, T. Pugh^{5,7}, J. Simpson⁷, L. Hao⁸, M. Lebo^{8,9}, W. Lane⁹, L. Briollais^{1,2}, J. Taher^{1,3}, J. Lerner-Ellis^{1,2,3}, S. Jones¹⁰, R. Abraham¹⁰, S. Arnoldo^{3,11}, N. Aujla^{1,2}, E. Bearss¹, A. Binnie^{3,11}, B. Borgundvaag^{1,3}, M. Dagher¹², L. Devine^{1,3}, S. M. Friedman^{1,5}, C. Fung^{1,3}, A-C. Gingras^{1,2,3}, L. Goneau¹³, E. Greenfeld^{1,3}, D. Kaushik¹¹, Z. Khan¹⁴, E. Lapadula^{1,2}, T. Lu¹, T. Mazzulli^{1,3}, A. McGeer^{1,2,3}, G. Morgan^{1,2,3}, A. Noor^{1,3}, S. McLeod^{1,3}, D. Richardson¹¹, S. Stern¹⁴, A. Taher^{3,5,14}, I. Wong¹⁴, N. Zarei¹⁴; ¹Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, ²Lunenfeld-Tanenbaum Res. Inst., Sinai Hlth., Toronto, ON, Canada, ³Univ. of Toronto, Toronto, ON, Canada, ⁴Unity Hlth. Toronto, Toronto, ON, Canada, ⁵Univ. Hlth. Network, Toronto, ON, Canada, ⁶The Hosp. for Sick Children, Toronto, ON, Canada, ⁷Ontario Inst. for Cancer Res., Toronto, ON, Canada, ⁸Lab. of Molecular Med., Partners Personalized Med., Cambridge, MA, ⁹Harvard Med. Sch. & Brigham and Women's Hosp., Cambridge, MA, ¹⁰BC Cancer Res. Inst., Vancouver, BC, Canada, ¹¹William Osler Hlth. System, Brampton, ON, Canada, ¹²Women's Coll. Hosp., Toronto, ON, Canada, ¹³Dynacare Med. Lab., Toronto, ON, Canada, ¹⁴Mackenzie Hlth., Toronto, ON, Canada

Abstract Body:

Background: The susceptibility and severity of COVID-19 caused by severe acute respiratory syndrome coronavirus (SARS-CoV-2) is tied closely to host genetic factors. Association studies have shown that genetic variants of the human leukocyte antigen (HLA) locus influence the progression of SARS-CoV-2 infections by affecting immune response pathways. This study intends to evaluate additional evidence for the role of Class I and II HLA allelic variation with COVID-19 outcomes including factors such as age and sex in a large prospective cohort. This study is part of the GENCOV initiative, which aims to examine the relationship between COVID-19 disease risk factors, host genetics, and antibody response. **Aim:** The aim of this study is to identify novel associations between HLA alleles and COVID-19 symptom severity and antibody response. **Methods:** The GENCOV cohort consists of 1438 individuals undergoing genome sequencing and serological analysis. Blood samples are collected during COVID-19 infection as well as at 1, 6, and 12 months after diagnosis for assessment of serial antibody titres (IgG, IgA, and IgM), isotypes, antigen targets, and viral neutralization. The bioinformatics tools Optitype and HLA-VBSeq were used to call HLA variants from genome data. **Results:** To date, serology results have been acquired from 1135 participants, with ~50% over age 50. Self-identified ancestry included White/European (70%), Middle Eastern (13%), Asian (11%), and Ashkenazi Jewish (9%). Among the cohort, 36% were hospitalized due to COVID-19 and 46% had at least one comorbidity, such as cardiac disease (7%), diabetes (5%), and cancer (4%). Optitype results for three HLA loci identified the highest frequency alleles among 778 participants as A*02:01 (42%), B*08:01 (13%), and C*04:01 (28%). Similarly, HLA-VBSeq analysis performed on 605 of the 778 participants identified the highest frequency alleles as A*02:01:01:01 (43%), B*35:01:01:05 (11%), and C*12:03:01:01 (71%). Previously, C*04:01 has been associated with severe COVID-19 outcomes resulting in intubation and hospitalizations. Regression tree analyses will be used to identify associations between HLA alleles and serological profiles (IgA, IgG, IgM), as well as characteristics such as sex, BMI, age, symptom severity, and disease comorbidities. Findings related to HLA associations will be presented at the conference. **Conclusion:** This study will evaluate the contribution of HLA genotype to COVID-19 symptom severity, antibody response and disease phenotypes to further elucidate potential host biological factors that influence variability in SARS-CoV-2 outcomes.

Complex Traits Posters - Thursday

PB1462. *HLA* associations with autoantibody-defined subgroups in idiopathic inflammatory myopathies

Authors:

L. Diaz-Gallo¹, V. Leclair^{1,2}, A. Galindo-Feria¹, S. Sarrafzadeh Zargaw¹, S. Rothwell³, O. Krystufkova⁴, H. Mann⁴, L. P. Diederichsen⁵, H. Andersson⁶, M. Klein⁴, S. Tansley⁷, N. McHugh⁷, J. A. Lamp³, J. Vencovsky⁴, H. Chinoy⁸, M. Holmqvist¹, M. Bianchi⁹, I. E. Lundberg¹, L. Padyukov¹; ¹Karolinska Inst.t, Stockholm, Sweden, ²Div. of Rheumatology, Jewish Gen. Hosp. Lady Davis Inst., Montreal, QC, Canada, ³Univ. of Manchester, Manchester, United Kingdom, ⁴Charles Univ., Prague, Czech Republic, ⁵Copenhagen Univ. Hosp., Copenhagen, Denmark, ⁶Oslo Univ. Hosp., Oslo, Norway, ⁷Univ. of Bath, Bath, United Kingdom, ⁸Manchester Univ. NHS, Manchester, United Kingdom, ⁹Uppsala Univ., Uppsala, Sweden

Abstract Body:

Objective: In idiopathic inflammatory myopathies (IIM), autoantibodies define specific phenotypes suggesting a role of the adaptive immunity in the disease mechanisms. Therefore, we explored the relationship between autoantibody-defined subgroups, including both myositis-specific and -associated autoantibodies (MSA/MAA) and *HLA* genetic variants.

Methods: We included 1348 patients with probable or possible IIM and determined the occurrence of: anti-Jo1, -PL7, -PL12, -EJ, -OJ, -SRP, -Mi2, -TIF1, -MDA5, -SAE1, -NXP2, -PM/Scl, -U1RNP and -Ro52 autoantibodies. We used cluster analysis to identify autoantibody-defined subgroups and estimated their associations with clinical manifestations, *HLA-DRB1*, *-DQA1*, *-DQB1*, alleles, and class II and I amino acid frequencies, using logistic regression.

Results: We identified eight subgroups with the following dominant autoantibodies (% of patients): 1) anti-Ro52 (10%), 2) -U1RNP (14%), 3) -PM/Scl (8%), 4) -Mi2 (5%), 5) -Jo1 (9%), 6) -Jo1/Ro52 (10%), 7) -TIF1 (6%) and 8) negative for tested autoantibodies (39%). *HLA-DRB1**11, *15, *-DQA1**03, and *-DQB1**03 alleles were more frequent in the anti-U1RNP dominated subgroup, while *HLA-DRB1**03, *-DQA1**05, and *-DQB1**02 alleles were overrepresented in the anti-PM/Scl dominated subgroup. For the anti-Mi2 dominated subgroup, *HLA-DRB1**07 had the strongest association signal. In the case of anti-TIF1 dominated subgroup, signals from class I (*HLA-C* and *-A*) were detected additionally to other signals from class II. Although *HLA-DRB1**03 was overrepresented in both the anti-Jo1 and anti-Jo1/Ro52 subgroups, *HLA-DQA1**05, *-DQB1**02 and aspartic acid in position 9 of HLA-B were increased only in the latter. Lastly, *HLA-DRB1**13, *-DQA1**01 and *-DQB1**06 alleles and tryptophan and arginine at position 152 of HLA-A were overrepresented in the subgroup negative for tested autoantibodies.

Conclusion: Distinct *HLA* class II and I associations were uncovered for each of the autoantibodies-defined IIM subgroups. Combining *HLA* genotype to autoantibody profiles in IIM can help to unravel some of the pathogenic mechanisms in different subgroups and suggest a role of the adaptive immune system in the production of autoantibodies in IIM.

Complex Traits Posters - Wednesday

PB1463. *HLA-DRB1*15:01* modifies sphingolipid levels in multiple sclerosis

Authors:

F. Briggs^{1,2}, E. Misicka¹, W. Tobin³, J. Oksenberg⁴, S. Gregory⁵; ¹Case Western Reserve Univ. Sch. of Med., Cleveland, OH, ²Cleveland Inst. for Computational Biology, Cleveland, OH, ³Mayo Clinic, Dept. of Neurology, Rochester, MN, ⁴Univ. California San Francisco, Weill Inst. for NeuroSci.s, Dept. of Neurology, San Francisco, CA, ⁵Duke Univ., Dept. of Neurology, Durham, NC

Abstract Body:

Background. Multiple sclerosis (MS) is an autoimmune disease and mediated, in part, by myelin-specific autoreactive T cells. The primary genetic risk locus for MS, *HLA-DRB1*15:01* (DR2b), encodes an MHC class II molecule involved in antigen presentation, with high affinity for myelin antigens. Recently, an *HLA-DQA1* variant (rs13211653A; significant linkage disequilibrium with DR2b: $D'=1$, $r^2=0.03$, $p<0.0001$) was associated with lower plasma levels of sphingomyelin (SM; $p<10^{-8}$). SM is a key component of myelin, and SM hydrolysis is the major pathway of ceramide (Cer) generation, and these sphingolipids (SLs) modulate many signaling pathways including neuronal cell death and T cell activity, and may play a prominent role in MS. **Objective.** To determine if DR2b differentially modifies SM and ceramide serum levels in MS. **Design/Methods.** Metabolomic profiling was conducted for 70 non-Hispanic white MS cases and 82 healthy controls (HCs) using untargeted HD4 global technology (Metabolon), and 29 SM and 9 Cer metabolites were discerned. MS cases were disease modifying therapy naïve (96%) or free (>90 days), ≤ 2 years of diagnosis, and <5 years of 1st symptom. HCs were age/sex frequency matched. DR2b carriage (dominant model) was determined by rs3135388A tagging SNP. Logistic regression models were conducted with main effects and an interaction term between DR2b carriage and each metabolite (normalized & standardized), adjusting for age, sex, smoking history, and BMI. The Simes method was used to control the false discovery rate. Stratified models were conducted for metabolites with significant interactions ($q<0.05$). **Results.** There were 54% and 46% of MS cases and 32% and 69% HCs were DR2b positive or negative, respectively; therefore, MS was associated with DR2b carriage in the study sample (OR=2.5; $p=0.007$). There were 18 (47%) significant DR2b x metabolites interactions for 15 SM and 3 Cer metabolites ($OR_{\text{Ixn}}<0.15$; $q<0.05$). In DR2b non-carriers, all metabolites were trending in association with MS (OR>2.3; $p<0.2$), and 9 metabolites (8 SM; 1 Cer) were significantly associated with MS status (OR: 2.7-11.2; $p<0.05$). The strongest associations with MS in DR2b non-carriers were for SM_{d18:1/22:1, d18:2/22:0, d16:1/24:1} (OR=11.2, $p=0.03$), SM_{d18:1/17:0, d17:1/18:0, d19:1/16:0} (OR=10.4, $p=0.0096$), and stearyl SM_{d18:1/18:0} (OR=8.1, $p=0.02$). Interestingly, no metabolite was associated with MS status in DR2b carriers. **Conclusions.** We observed significantly elevated serum levels of several SL metabolites in DR2b negative MS cases compared to HCs but not in DR2b positive individuals, and warrants further investigation. Replication analyses are underway.

Complex Traits Posters - Thursday

PB1464. Host Genetics in Resistance to COVID-19: Hints from recovered Brazilian superelderly

Authors:

M. de Castro¹, E. C. Castelli², M. S. Naslavsky¹, M. O. Scliar¹, N. S. B. Silva², R. N. Pereira², V. A. O. Ciriaco², C. F. B. de Castro², C. T. Mendes-Júnior³, D. Meyer¹, K. Nunes¹, L. R. B. Matos¹, M. V. R. Silva¹, J. Y. T. Wang¹, J. Esposito¹, V. R. Cória¹, R. H. Bortolin¹, M. H. Hirata¹, L. P. Dell Aquila⁴, A. Razuk-Filho⁴, P. B. Batista-Júnior⁴, A. N. Duarte-Neto¹, M. Dolhnikoff¹, P. H. N. Saldiva¹, M. Passos-Bueno¹, M. Zatz¹; ¹Univ. of Sao Paulo, Sao Paulo, Brazil, ²São Paulo State Univ. (UNESP), Botucatu, Brazil, ³Univ. of Sao Paulo, Ribeirão Preto, Brazil, ⁴Prevent Senior Inst., Sao Paulo, Brazil

Abstract Body:

Since the onset of COVID-19 pandemic, it has been suggested that the clinical variability caused by SARS-CoV-2 infection could be influenced by the host genotype. So far, most of the reported genetic variants responsible for the disease vulnerability are associated with immune response, involving type I IFN immunity and modulation; *HLA* cluster genes; inflammasome activation, genes of interleukins; and chemokines receptors. On the other hand, little is known about the resistance mechanisms against SARS-CoV-2 infection. Interestingly, although elderly people are at a significantly higher risk of having severe and lethal forms of COVID-19, some cases of still unvaccinated centenarians who have recovered from COVID-19 or remained asymptomatic despite SARS-Cov-2 exposition have been identified worldwide. Based on these observations and aiming to identify candidate genes associated with host resistance we investigated 100 Individuals older than 90 years recovered from COVID-19 (Group 90+ / n=100) with mild symptoms as compared to the other extreme group, 50 individuals younger than 60 years who died due to COVID-19 (Group 60- / n=50). SARS-CoV-2 infection was confirmed by RT-PCR test or histopathological analysis. Biological samples (peripheral blood from the elderly or skin fragments from the younger at autopsy) were collected to obtain DNA. Whole-exome sequencing (WES) followed by a state-of-the-art method was performed to call genotypes and haplotypes across the highly polymorphic major histocompatibility complex (MHC) region. We found that the super elderly group displayed a higher frequency of missense variants in the *MUC22* gene (which is a member of the mucins' family), as one of the strongest signals in the MHC region as compared to the severe COVID-19 group and the general elderly control population. Since the pro-inflammatory basal state in the elderly may enhance the susceptibility to severe COVID-19, we hypothesised that *MUC22* might play an important protective role against severe COVID-19, by reducing overactive immune responses in the senior population. This work was supported by the São Paulo Research Foundation (FAPESP/Brazil) [grant numbers 2013/08028-1, 2013/17084-0, 2014/50931-3, 2017/19223-0 and 2020/09702-1], the National Council for Scientific and Technological Development (CNPq) [grant number 465355/2014-5], JBS S.A. [grant number 69004] and the United States National Institutes of Health (NIH) (R01 GM075091).

Complex Traits Posters - Wednesday

PB1465. How are *APOE4*, changes in weight/BMI, and longevity related? Insights from a Causal Mediation Analysis.

Authors:

R. Holmes, H. Duan, K. Arbeev, D. Wu, A. Yashin, S. Ukraintseva, Biodemography of Aging Research Unit, Social Science Research Institute; Duke Univ., Durham, NC

Abstract Body:

Introduction: The $\epsilon 4$ allele of the *APOE* gene (*APOE4*) is known for its negative association with human longevity, though the mechanism is unclear. *APOE4* was also linked to changes in weight/BMI, and the latter itself was associated with survival in late life. This indicates a possibility that some of the *APOE4* effects on longevity may be attributed to aging changes in weight/BMI.

Method: To see if the negative association of *APOE4* with longevity is mediated by key characteristics of age-trajectories of weight/BMI, we performed a Causal Mediation Analysis in the Health and Retirement Study data. The $\epsilon 4$ allele carrying status (yes vs. no) was the *causal agent*. We used two binary *mediators*: (1) rate of changes in bodyweight (SlopeW) between ages 65 and 80 being below vs. above median value; and (2) age at reaching maximum BMI value (AgeMaxBMI) being younger vs. older than 75, and all individuals survived age 80. The *outcome* was a binary survival variable (survived age 85+ vs. died before 85). 3292 and 3020 individuals with AgeMaxBMI and SlopeW values, respectively, were included in the analysis. SAS procedure PROC CAUSALMED under counterfactual framework was used, controlling for smoking and education to address confounding. Log link was used to address the non-rare outcome situation. Mediation effects were evaluated by the Total Effect (TE) and Natural Indirect Effect (NIE).

Results: Younger AgeMaxBMI (below 75) was a significant mediator of negative *APOE4* effect on survival to ages 85+ in total sample, and in samples stratified by sex and race (p-values between 0.02-0.04). The $\epsilon 4$ carriers were 17%-21% more likely to die before age 85, compared to noncarriers, when considering the mediator effect (TE p-values 0.02-0.047, NIE p-values 0.006-0.008). The percentage of the total effect mediated by AgeMaxBMI was 21% for the overall sample (p=0.044), and varied between 21%-22% in strata (p-values 0.042-0.046). SlopeW was a significant mediator of the *APOE4* effect on survival 85+ in unstratified data, with chances of dying before age 85 being 22% higher for $\epsilon 4$ carriers vs. noncarriers (TE p-value = 0.02). Indirect effect was statistically significant (NIE p-value=0.015). Similar effects were observed for both sexes.

Conclusion: Results of our Causal Mediation Analysis suggest that carriers of the *APOE* $\epsilon 4$ allele may have lower chances of surviving to age 85 and beyond in part because they reach peak values of weight/BMI at younger ages, and decline faster afterwards, compared to non-carriers. These findings are in line with the idea that detrimental effects of *APOE4* on longevity may involve accelerated physical aging, and that *APOE4* can be a promising target for anti-aging interventions.

Complex Traits Posters - Thursday

PB1466*. Hypermethylation of *PM20D1* promoter is associated with atherosclerosis in Dominican Families

Authors:

L. Wang¹, N. Dueker², H. Zhao³, C. Dong⁴, D. Cabral⁵, R. Sacco⁵, S. Blanton⁴, C. Dong⁴, T. Rundek⁴; ¹Univ. of Miami, PALMETTO BAY, FL, ²Univ. of Miami Miller Sch. of Med., Miami, FL, ³Yale Univ. Sch. of Publ. Hlth., New Haven, CT, ⁴Univ Miami, Miami, FL, ⁵Univ. of Miami, Miami, FL

Abstract Body:

Introduction: Carotid intima-media thickness (cIMT) is a marker of atherosclerosis and a predictor of vascular disease. Traditional vascular risk factors, e.g., hypertension, smoking, cholesterol, obesity, diabetes, and genetic loci do not completely explain the variation in cIMT. We sought to identify epigenetic factors that may explain the remaining cIMT variability. **Methods:** Methylome-wide association analyses (MWAS) was performed in 61 Dominican families with 799 individuals. CpG methylation in blood-derived DNA was measured using the Human MethylationEPIC BeadChip. *ComBat* was used to remove batch effects. Linear mixed model analyses were performed regressing bifurcation cIMT on beta values for CpG sites, adjusting for age, sex, 1st ancestry principal component, and cell composition. Family was included as a random effect. Differentially methylated regions (DMRs) analysis was performed using Comb-p. Methylation-quantitative trait loci (m-QTL) analysis was done using linear mixed effects models implemented in the R package 'lmer4'. One-sample mendelian randomization (MR) analysis was performed using the R package 'systemfit.' **Results:** 29 DMRs were significantly associated with cIMT (Sidak $p < 0.05$), averaging 5 ± 2 CpGs per DMR. The strongest DMR (Sidak $p = 1.2 \times 10^{-13}$) overlapped with the promoter of *PM20D1*. All 11 CpGs within the *PM20D1* DMR were associated with cIMT ($\beta = 0.20 \sim 0.34$; $p = 0.0004 \sim 0.002$). *PM20D1* deficiency has been previously implicated in obesity and diabetes. In our cohort, the *PM20D1* methylation was not associated with obesity (BMI: $r = 0.02$, $p = 0.54$; waist-hip ratio: $r = 0.03$, $p = 0.42$), or diabetes ($p = 0.09$), suggesting that the association of *PM20D1* methylation with cIMT is not mediated by obesity or diabetes. Furthermore, all 11 CpGs are associated with cis variants nearby, e.g. cg14159672_rs1775147 ($\beta = 0.14$, $p = 5.53 \times 10^{-36}$), which is consistent with The Accessible Resource for Integrated Epigenomic Studies (ARIES, $\beta = 0.52$, $p = 2.21 \times 10^{-22}$). In addition, rs1775147 is an expression QTL for *PM20D1* in the peripheral blood ($p = 6.6 \times 10^{-11}$), suggesting hypermethylation of *PM20D1* DMR directs lower expression of the gene. MR analysis supported that the hypermethylation at the *PM20D1* promoter is causatively associated with cIMT ($p = 0.02$). **Conclusion:** *PM20D1* acts as endogenous uncouplers of mitochondrial respiration, which is a cytoprotective strategy against oxidative stressors. Reduced expression of *PM20D1* leads to reduced uncoupling and increased reactive oxygen species production. Taken together, our data suggest that hypermethylation of *PM20D1* is a novel mechanism of atherosclerosis probably via oxidative stress.

Complex Traits Posters - Wednesday

PB1467. Hypertension Prediction From Metabolomics For Two Cardiovascular Disease Cohorts and Transfer Learning of Resultant Models

Authors:

H. Gong¹, D. Nguyen², A. Gaye¹, G. Gibbons¹; ¹NIH, Bethesda, MD, ²NIH, Bethesda, MD

Abstract Body:

Background. We hypothesize that hypertension and other cardiovascular disease will have predictive signal in metabolomics data, since small molecules are known to mediate and be produced by those diseases. We investigated two cohorts, the MH-GRID (Minority Health Genomics and Translational Research Bio-Repository Database) cohort of blacks (n=165) and GENE-FORECAST (Genomics, Environmental Factors, and the Social Determinants of Cardiovascular Disease in Africans Americans) (n=534).

Data. We have GC-TOF metabolomics data from Plasma in both MH-GRID and GENE-FORECAST. The data is corrected for sample drift using SERRF (Systematical Error Removal using Random Forest), and batch corrected into the same multidimensional space using the ComBat empirical Bayes method.

Results. We generated a random forest model in MHGRID with an AUC of 0.76. That model applied to GENE-FORECAST data resulted in a 62.6% accuracy for hypertension prediction. Additionally, we found WGCNA metabolite modules that were significantly associated with triglycerides, glucose, HDL, and systolic blood pressure.

Complex Traits Posters - Thursday

PB1468. Identification of a locus associated with major depressive disorder in a Colombian population

Authors:

C. Lattig¹, M. M. Velásquez¹, L. C. Hernández¹, Á. Arenas¹, Y. Gómez¹, E. Ferro²; ¹Univ. de los Andes, Bogotá D.C, Colombia, ²Clínica Montserrat, Bogotá D.C, Colombia

Abstract Body:

Determining the genetic architecture of major depression is relevant for the understanding of this multifactorial and highly prevalent disorder as well as designing strategies of personalized medicine. Several genome-wide association studies (GWAS) have allowed the identification of variants that may confer risk of developing depressive phenotypes. However, most of them have focused on European populations, leaving underrepresented those of admixed ancestry, such as Latin Americans. The present study included a sample of 171 healthy controls and 169 cases with a primary diagnosis of major depression. After genotyping, which was performed through the Infinium Global Screening Array-24 v2.0 (Illumina), the data was filtered according to genotyping call rates, minor allele frequencies and Hardy-Weinberg equilibrium. Then a multivariate logistic regression analysis was performed (Plink 1.9) to identify common variants associated with the clinical phenotype. The results provide evidence of a locus potentially associated with the disorder and that is in linkage disequilibrium with variants located within the *CNTN5* gene. This gene codes for the human contactin-5 protein, which has a role in the formation and maintenance of neuronal connectivity during development and have been associated with depression and autism. Although with a limited sample size, this study included a strategy of deep phenotyping and provides insights into the genetic architecture of major depression in an admixed population, that could be useful for subsequent meta-analyses.

Complex Traits Posters - Wednesday

PB1469. Identification of genetic loci associated with peanut specific IgG4 in high risk children from the LEAP study

Authors:

K. Kanchan¹, J. Perry², G. Shankar¹, I. Ruczinski³, C. Baloh⁴, G. Du Toit⁵, G. Lack⁵, K. Cerosaletti⁶, G. Nepom⁶, R. Mathias¹; ¹Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ²Univ. of Maryland Sch. of Med., Baltimore, MD, ³Bloomberg Sch. of Publ. Hlth. Johns Hopkins Univ., Baltimore, MD, ⁴The Immune Tolerance Network, Seattle, WA, ⁵King's Coll. London, and Guy's and St Thomas' NHS Fndn. Trust, London, United Kingdom, ⁶Benaroya Res. Inst. at Virginia Mason, Seattle, WA

Abstract Body:

Learning Early About Peanut Allergy (LEAP) was a prospective, randomized, clinical trial in young children at high risk of peanut allergy (PA), establishing that early peanut consumption is effective in preventing PA. LEAP participants were randomly assigned to dietary peanut consumption or avoidance, and levels of peanut-specific IgG4 (psIgG4) were measured in serum at multiple time points throughout the trial. The consumption group of the trial had increased levels of psIgG4. We aim to identify genetic determinants of psIgG4 among those consuming peanuts, to elucidate potential genetic mechanisms for tolerance to peanuts. Whole genome sequencing was carried out on all participants. Genome-wide association tests adjusting for genetic ancestry, age at baseline, and sex was performed on 10,830,622 variants for psIgG4 among the 267 consumption group participants at 60 months of age, the primary endpoint in the trial. We identified 45 variants with p-values below the suggestive genome-wide significance threshold of 1×10^{-5} , mapping to 17 genetic loci. These loci mapped to genes known to be associated with barrier function (e.g., *SEPT2* and *PNPLA1*) and previously documented PA genes (e.g., *HLA-DQA1* and *HLA-DQB1*). We performed fine-mapping for 16 of these novel loci using the colocalization approach coloc (the HLA locus on chr6 has been previously described) and tested for colocalization against eQTLs for genes +/-200kb around each sentinel variant from 54 tissues from the GTEx catalog. A chromosome 2 locus for psIgG4 colocalizes to regulatory evidence for *SEPT2* and a lncRNA gene *AC005104.3*. We identified two credible causal variants, rs115271170 (MAF=33%, posterior probability of colocalization > 97%) in the small intestine and rs1429386442 (MAF=10%, posterior probability of colocalization 98%-100%) in numerous gut tissues such as esophagus gastroesophageal junction, esophagus mucosa, esophagus muscularis, stomach, colon sigmoid, colon transverse, lung, and whole blood. *SEPT2* encodes Septin 2, which is required for correct formation and polarization of the epithelium and mediates epithelial barrier function. This novel locus, a determinant of the psIgG4 response to peanut consumption, offers further validation of the importance of barrier dysfunction in food allergy, and the value of looking for genetic underpinnings of food allergy in the context of environmental exposure to an allergen.

Complex Traits Posters - Thursday

PB1470. Identification of genetic loci with divergent effects between Crohn's disease and ulcerative colitis.

Authors:

Y. Kim¹, S. Jung¹, H-S. Lee¹, D. Park¹, Y. Lee¹, S. Park¹, J. Baek¹, S. Hwang², S. Park², S-K. Yang², B. Ye², K. Song¹; ¹Dept. of Biochemistry and Molecular Biology, Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, ²Dept. of Gastroenterology, Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of

Abstract Body:

Crohn's disease (CD) and ulcerative colitis (UC), the two main types of inflammatory bowel disease, show substantial differences in clinical course and the treatment response. Understanding the genetic factors underlying the distinct characteristics of the two diseases will be crucial for improving diagnosis and treatment. We performed a genome-wide association study (GWAS) between CD (n=2359) and UC (n=2175) in a Korean population, followed by replication in an independent samples of 772 CD and 619 UC cases. We used publicly available summary statistics (Liu et al., 2015; de Lange et al. 2017) for comparison with and replication of our results. Genetic correlation between CD and UC in the Korean population ($r_{gKOR}=0.17$) was significantly lower than that of a European population based on the publicly available summary statistics ($r_{gEUR}=0.60$, $Z\text{-score}=3.60$, $P=1.60\times 10^{-4}$). Through a meta-analysis of the GWAS, we identified 2 novel loci with divergent effects between CD and UC: rs9842650 in *CD200* at 3q13 and rs885026 in *NCOR2* at 12q24. In addition, the 7 established susceptibility loci (MHC, *TNFSF15*, *OTUD3*, *USP12*, *IL23R*, *FCHSD2*, and *RIPK2*) reached genome-wide significance. Of the 9 loci, 6 (MHC, *TNFSF15*, *OTUD3*, *USP12*, *IL23R*, and *CD200*) were replicated in the case-case GWAS (CC-GWAS) of European and/or East Asian populations using the summary statistics. Pathway analysis using MAGMA showed that the interferon-gamma mediated signaling ($P=1.16\times 10^{-8}$) and cytokine-mediated signaling pathway ($P=2.71\times 10^{-6}$) were associated with genes in the loci with divergent effects. The proportion of variance explained in CD-UC status by polygenic risk score analysis was up to 22.6% (Nagelkerke's R^2). The area under the receiver-operating characteristic curve value was 0.74, suggesting acceptable discrimination between CD and UC. CD-UC GWAS provides new insights into genetic differences between the two diseases with similar symptoms, which might be useful in improving their diagnosis and treatment. Future studies of large-scale data are needed to verify our observations in diverse populations.

Complex Traits Posters - Wednesday

PB1471. Identification of novel genetic variants for association with rheumatoid arthritis susceptibility in Pakistani population

Authors:

S. Ali¹, R. Fatima¹, A. Maigoro^{1,2}, M. Bashir¹, A. Kanwal¹, H. Sadia¹, A. Suhail¹, J. M. Malik³, S. Khurshid⁴; ¹COMSATS Univ. Islamabad, Islamabad, Pakistan, Islamabad, Pakistan, ²Chungnam Natl. Univ., Daejeon, Korea, Republic of, ³Rehmat Noor Clinic, Rawalpindi, Pakistan, ⁴Pakistan Inst. of Med. Sci., Islamabad, Pakistan

Abstract Body:

Rheumatoid arthritis (RA) is an autoimmune disease, primarily affecting synovial joints causing inflammation and joint degradation. RA is responsible for a considerable degree of morbidity. It is a complex disease with both genetic and environmental risk factors. In the present study, 740 participants (370 cases and 370 controls) of Pakistani descent were genotyped for 14 common variants in nine candidate genes. Prior informed written consent was sought from all the cases and controls. Genotyping of the selected variants was done through PCR-based techniques. The association of variants with RA susceptibility was statistically investigated through multivariate logistic regression analysis. This case-control study indicated some significant associations in the overall and sub-group analyses. The rs2275913, rs3819024 and rs8193036 of *IL17A*; rs2397084 of *IL17F*; rs11549465 and rs11549467 of *HIF1A*; rs1799983 of *NOS3*; rs874881 of *PADI4* and rs4803455 of *TGFB1* showed significant association with the modified RA risk. In the smoking status-based sub-group analysis, the rs3819024 and rs8193036 of *IL17A*; rs763780 and rs2397084 of *IL17F*; rs11549465 of *HIF1A*; rs1799983 of *NOS3* and rs874881 of *PADI4* were found to be associated with differential risk among smokers and non-smokers. Further sub-group analysis, on the basis of self-identified gender, revealed the significant associations of rs2275913 of *IL17A*; rs763780 of *IL17F*; rs11549465 of *HIF1A* and rs1800795 of *IL6*, with susceptibility to RA. In brief, the present study identifies novel common variants and also confirms some previously known associations for rheumatoid arthritis susceptibility.

Complex Traits Posters - Thursday

PB1472. Identification of platelet aggregation genetic determinants: Results from a cross-ancestry GWAS meta-analysis

Authors:

F. Thibord, M. V. Chan, B. B. Nkambule, A. Lachapelle, J. Grech, R. E. Pashek, J. A. Cunha, M-H. Chen, A. D. Johnson; NIH/NHLBI, Framingham, MA

Abstract Body:

Platelets can be activated in response to several endogenous agonists, such as collagen exposed at a site of vessel injury, resulting in aggregation and the formation of a platelet plug to preserve vascular integrity. However, platelet hyperactivity has been associated with an increased risk of cardiovascular disease, a major cause of mortality worldwide. As platelet reactivity is heritable, our aim was to identify genetic associations in a mixed population utilizing newly collected data.

Platelet reactivity was investigated using two assays: light transmission aggregometry (LTA) and multiplate impedance aggregometry (MP). Agonists used to induce platelet activation included ADP, collagen, ristocetin and TRAP-6 amide for MP, and arachidonic acid and epinephrine were additionally used for LTA. Analyses were conducted in up to 2901 participants of European ancestry from the Framingham Heart Study (FHS) and 262 individuals of multi-ancestry from the OMNI cohort. Genetic analyses were adjusted for relatedness, age, sex, aspirin use, and 10 principal components using a linear mixed model. Variants with minor allele frequency ≥ 0.01 and imputation quality ≥ 0.8 were included in the analysis. Results from FHS and OMNI were then meta-analyzed using a fixed-effect inverse variance weighted model, and associations with a concordant effect direction in both cohorts and reaching $P < 5 \times 10^{-8}$ were considered significant.

Our analyses revealed five significant signals: rs6510306-C with stimulation by epinephrine (*CEP89* intron, $P = 1.17 \times 10^{-9}$), rs72867273-G with TRAP-6 (intergenic, $P = 1.96 \times 10^{-9}$) and rs670337-C with collagen (intergenic, $P = 4.96 \times 10^{-8}$) in LTA, as well as the previously reported *ITGA2* and *ADRA2A* loci, associated with MP collagen and LTA epinephrine responses, respectively. Then, we investigated suggestive associations ($P < 10^{-5}$) that were previously reported to also influence the platelet transcriptome, using eQTL results from the CEDAR and GENESTAR studies. This analysis identified nine associations that also influenced the expression of the following genes in platelets: *SAP30L*, *PNP*, *TPMI*, *PYGB* and *NCOA3*, as well as the known *PEAR1*, *ITGA2*, *ADRA2A* and *GP6*.

In conclusion, our significant results, as well as our suggestive results combined with platelet eQTL data, point to novel molecules that influence platelet aggregation, which could help to develop new therapeutic strategies to reduce platelet hyperactivity. Future directions include expanding these analyses using other platelet assays, as well as additional analyses stratified by sex and anti-platelet medication use.

Complex Traits Posters - Wednesday

PB1473. Identification of shared genes and pathways between psychiatric GWAS and monogenic neurodevelopmental disorders

Authors:

P. Jansen^{1,2}, E. van Walree^{1,2}, S. de Lange³, D. Posthuma²; ¹Dept. of Human Genetics, Amsterdam UMC, Amsterdam, Netherlands, ²Dept. of Complex Trait Genetics, VU Univ., Amsterdam, Netherlands, ³Dept. of Complex Trait Genetics, VU Univ., Amsterdam, Netherlands, Amsterdam, Netherlands

Abstract Body:

Introduction: Large genome-wide association studies (GWAS) rapidly identified novel risk genes for psychiatric disorders. Mendelian neurodevelopmental disorders (NDD) caused by single gene mutations show overlapping behavioral symptomatology, however the extent to which genes are shared between psychiatric GWAS and monogenic NDD needs to be fully explored.

Methods: We carry out genetic overlap analyses between five large-scale psychiatric GWAS (ADHD, autism, bipolar disorder, depression, schizophrenia, sample size range: N=46,077–807,553) and monogenic neurodevelopmental disorders (NDDs). Consistent gene-mapping was done in FUMA for each GWAS. Overlap analyses were carried out with monogenic disease genes in the Developmental Disease Gene - Phenotype (DDG2P) database. Using various bioinformatic resources, we analyzed gene properties and gene-expression of genes implicated in psychiatric disorders, NDDs or both.

Results: We found strong and statistically significant enrichment of psychiatric GWAS genes in all monogenic NDDs (enrichment=1.18, $P=1.70 \times 10^{-5}$), and NDDs genes related to abnormal brain development (enrichment=1.39, $P=1.10 \times 10^{-9}$). Of all psychiatric disorders, ADHD showed the strongest overlap with NDDs ($P=0.009$), and bipolar the lowest ($P=0.96$). We show that differences in gene function (synaptic and neurogenesis), properties (pLI score, missense sensitivity) and expression patterns predict involvement of genes in psychiatric GWAS, NDD or overlapping in both. We observed that psychiatric GWAS genes are significantly more common in known copy number variant (CNV) regions that are associated with a high risk of psychiatric symptoms (8% of psychiatric genes located CNVs).

Conclusions: Our findings provide novel insights into partly shared genetic etiologies between polygenic psychiatric disease and monogenic NDDs, and highlight mechanisms and gene properties that explain overlap between seemingly separate monogenic and polygenic disease. These results show that GWAS can inform the search of novel NDD genes, whereas monogenic NDDs can aid in the interpretation of GWAS loci.

Complex Traits Posters - Thursday

PB1474. Identifying compounds to treat opiate use disorder by leveraging multi-omic data integration and multiple drug repurposing databases.

Authors:

J. Stratford¹, M. Carnes¹, M. Schu¹, B. Quach¹, C. Willis¹, R. Mathur¹, E. Johnson¹, J. Carter¹, T. Nolen¹, N. Vandergrift¹, T. Kosten², B. Webb¹; ¹RTI Intl., Research Triangle Park, NC, ²Baylor Coll. of Med., Houston, TX

Abstract Body:

Recently, specific genetic loci influencing opiate use disorder (OUD) risk have begun to be identified. Gene expression and network analyses offer additional insight beyond risk. However, even multi-domain investigations rarely focus on producing prioritized targets for translational studies. To address this gap, we constructed a framework to identify biological targets and pharmacotherapies for clinical repurposing studies. This is accomplished by 1) leveraging extant results, 2) performing cross-omic network-based analysis, 3) creating an integrated catalogue of results and biological targets, and 4) generating a prioritized list of existing candidate compounds for future repurposing studies. OUD results were collected including two large ($n=304,507$, $n=79,729$) independent genome-wide association studies (GWAS) and four post-mortem human brain gene expression (285 independent samples). Network analyses were performed using dense module GWAS (dmGWAS) to identify human brain specific protein-protein interaction (PPI) modules. Drug repurposing databases Pharos, Open Targets, Therapeutic Target Database (TTD), and DrugBank were queried for clinical status, safety, and target selectivity. Cross-omic and drug query results were integrated for all genes in the genome, allowing flexible filtering to identify candidate compounds for follow-up review. Gene expression meta-analysis and gene-level GWAS analyses revealed 2698 and 3 genes ($FDR < 0.05$), respectively. Network analysis detected 22 PPI modules containing 71 genes showing enrichment. For pilot drug repurposing analysis, we selected genes showing robust expression differences ($q < 0.01$, $n=605$), suggestive GWAS evidence ($q < 0.16$, $n=115$), and is in an enriched PPI module. 24 of 767 genes showed evidence across more than one domain with 6 of 24 being targeted by approved compounds including known targets OPRM1 and DRD2, as well as four additional actionable genes. Across the four repurposing databases, seven approved compounds were identified targeting the novel genes and represent repurposing candidates for follow-up and expert review. This study leveraged existing results and multiple resources to identify approved compounds that target genes associated with OUD. By querying multiple lines of evidence, the approach allows a) querying many genes of interest, b) detecting candidates missed using a single domain or resource, c) and produces a succinct summary to facilitate efficient expert review. Identifying larger pools of candidate pharmacotherapies and summarizing the supporting biological evidence bridges the gap between discovery and translational studies.

Complex Traits Posters - Wednesday

PB1475. Identifying Developmental Stuttering comorbidities using Electronic Health Records.

Authors:

A. Scartozzi¹, D. M. Shaw¹, H. G. Polikowsky¹, D. G. Pruett¹, H-H. Chen², L. E. Petty¹, R. L. Gordon², Y. Yu³, 23andMe Research Team, R. M. Jones², C. D. Huff³, S. Kraft⁴, J. E. Below²; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³MD Anderson Cancer Ctr., Houston, TX, ⁴Wayne State Univ., Detroit, MI

Abstract Body:

Developmental stuttering is a common speech disorder (studies estimate at least a 5% lifetime prevalence) characterized by prolongations, blocks and repetitions of speech sounds. Although recent population studies have provided evidence for both its heritability and polygenicity, the genetic factors impacting stuttering risk remain largely uncharacterized. We leveraged Electronic Health Records (EHRs) to investigate stuttering comorbidities (i.e., the presence of one or more conditions that co-occur with stuttering). Investigating stuttering comorbidities can help us glean information about shared biology. The current study uses polygenic risk scores (PRS) and a phenome-wide association study (PheWAS) to identify comorbidities of stuttering risk within Vanderbilt's EHR-linked biobank (BioVU). We developed sex- and ancestry-specific PRS models using summary statistics provided by 23andMe, Inc. from a genome-wide association study of self-reported stuttering in more than 1 million individuals. PRS models were tested and validated using an independent and clinically ascertained cohort, the International Stuttering Project (ISP). The European-Male (EUR-M) PRS model was predictive of stuttering status in ISP European males and females. We applied our 23andMe EUR-M PRS model to European participants within BioVU, $N = 71,986$. We then performed a PheWAS to identify phenotypes comorbid with stuttering risk. Although no Bonferroni significant associations were identified, some of the top hits recapitulated previously identified comorbid traits. A prior study identified conditions impacting weight control as common comorbidities, and we also observed a strong association with obesity ($p = 1.33 \times 10^{-4}$). Similarly, prior studies have identified suicidal ideation as notably enriched in individuals with developmental stuttering, and suicidal ideation was among our top findings ($p = 1.14 \times 10^{-3}$). Finally, prior studies have noted that the atopic triad is a common stuttering comorbidity, and we identified respiratory complications as associated with developmental stuttering risk ($p = 4.53 \times 10^{-3}$). Future studies will examine sex-specific comorbidities and examine differences in comorbid phenotypes captured by PRS applied to other sex and ancestry groups. By understanding stuttering comorbidities, we hope to elucidate shared mechanisms that contribute to its pathophysiology and enable improvements in clinical care.

Complex Traits Posters - Thursday

PB1476. Identifying genetic variations that connect iron homeostasis to metabolic disease.

Authors:

S. Praggastis¹, B. Crowell¹, L. Ivanova¹, R. Center², G-R. DiscovEHR Collaboration², A. N. Economides¹, S. Hatsell¹, H. E. Lob¹, J. Bovijn³, L. Lotta⁴, T. De⁵; ¹Regeneron Pharmaceuticals, Tarrytown, NY, ²Regeneron Genetics Ctr., Tarrytown, NY, ³Regeneron Genetics Ctr., Oxford, United Kingdom, ⁴Regeneron, Tarrytown, NY, ⁵Regeneron Genetics Ctr., Chicago, IL

Abstract Body:

Individuals with genetic iron overload diseases, such as hemochromatosis, have a higher incidence of type 2 diabetes. Elevated serum levels of the iron biomarkers ferritin and transferrin are positively correlated with the development of type 2 diabetes in individuals without iron overload diseases. However, the molecular mechanism by which iron overload results in metabolic dysfunction remains unknown. The goal of this study was to identify biomarkers that correlate iron homeostasis with abnormal glucose levels. We performed large-scale exome-wide association analyses and a genome-wide association study (GWAS) of glucose, HbA1c, and iron-related biomarkers (N = 480,389), as well as of type 2 diabetes with 58,379 cases and 530,072 controls. We identified known iron homeostasis pathway members and novel associations to both iron overload and type 2 diabetes. Through our GWAS analyses we identified our top candidate variant associated with reduced iron levels and reduced glucose levels (UIBC, beta (95%CI) = 0.049 SD (0.035,0.063), p = 3.61 x 10⁻¹², Glucose, beta (95%CI) = -0.02 SD (-0.025,-0.017), p = 2.34 x 10⁻²⁴). This variant has been previously associated with iron biomarkers and is a multi-tissue expressed quantitative trait loci (eQTL) in *FADS1/FADS2*, that results in reduced expression of *FADS1*. *FADS1* plays a critical role in long chain polyunsaturated fatty acid synthesis, and hepatic knockdown of this gene improves glucose clearance and limits adipose tissue expansion *in vivo*. *In vitro* studies demonstrated that expression is induced by iron supplementation. Variants in the *FADS1* locus have been strongly associated with blood glucose levels in previous GWAS, however the mechanism underlying this association remains unclear. Our future efforts will be to study the mechanism in a mouse-model. The novel and potentially pleiotropic effect of this variant on iron and glucose levels may help to inform future drug development.

Complex Traits Posters - Wednesday

PB1477. Identifying Metabolomic Signatures of Genetic Liability to Type 2 Diabetes in East Asian Adolescents

Authors:

S. Chen¹, B. He¹, S. Luo¹, S. Lam², M. Kwok¹, C. Schooling^{1,3}, S. Au Yeung¹, G. Leung¹; ¹The Univ. of Hong Kong, Hong Kong, China, ²The Chinese Univ. of Hong Kong, Hong Kong, China, ³City Univ. of New York, New York, NY

Abstract Body:

Background East Asians have a higher burden of type 2 diabetes (T2D) compared to Europeans despite lower rates of obesity. Exploring adolescent metabolomic signatures associated with increased genetic liability to T2D in East Asians can help identify unexplored causal factors of T2D in this population. **Methods** Using the East-Asian-population-based “Children of 1997” birth cohort (n=3,251), we generated a weighted genetic risk score (GRS) based on 128 genetic variants associated with T2D in East Asians ($p < 5 \times 10^{-8}$). Multivariable linear regression was used to assess the association of GRS with Nuclear Magnetic Resonance (NMR)-derived metabolites measured from serum at 16-18 years old, adjusted for batch, age, sex, and 6 top principal components. **Result** Higher genetic liability to T2D (per SD higher GRS) was associated with higher leucine (0.039 standard deviation (SD), 95% CI 0.010 to 0.069), valine (0.035 SD, 95% CI 0.004 to 0.067) and total branched-chain amino acids (BCAA) (0.035 SD, 95% CI 0.004 to 0.065). We also found associations with higher ketones including acetoacetate (0.053 SD, 95% CI 0.018 to 0.087), 3-hydroxybutyrate (0.049 SD, 95% CI 0.015 to 0.083), and acetone (0.051 SD, 95% CI 0.017 to 0.085). However, all these associations did not withstand Bonferroni correction. As expected, genetic liability to T2D was associated with higher glucose (0.048 SD, 95% CI 0.014 to 0.082). **Conclusion** Genetic liability to T2D was potentially associated with higher BCAA and ketones in this East Asian adolescent population. Whether these signals are causal in the etiology of T2D in East Asians requires additional investigation.

Complex Traits Posters - Thursday

PB1478. Identifying novel gene dysregulation associated with opioid overdose death: a meta-analysis of differential gene expression in human prefrontal cortex.

Authors:

J. Carter¹, B. Quach¹, C. Willis¹, D. Hancock¹, J. Montalvo-Ortiz^{2,3}, R. Logan^{4,5}, C. Walss-Bass⁶, B. Maher⁷, E. Johnson^{1,8}, PGC-SUD Epigenetics Working Group; ¹RTI Intl., Research Triangle Park, NC, ²Yale Univ., Orange, CT, ³Clinical NeuroSci.s Div., Natl. Ctr. of PTSD, West Haven, CT, ⁴Dept. of Pharmacology and Experimental Therapeutics, Boston Univ. Sch. of Med., Boston, MA, ⁵Ctr. for Systems NeuroSci., Boston Univ., Boston, MA, ⁶Univ. of Texas Hlth.Sci. Ctr. at Houst, Houston, TX, ⁷Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ⁸Fellow Program, Research Triangle Park, NC

Abstract Body:

Background: Of the U.S. population aged 12 and older, 10.1 million people misused opioids in 2019. In 2020, the U.S. saw the highest 12-month count of opioid overdose deaths (OOD) recorded, >70,000: a 40% increase since 2019. The opioid epidemic continues to be a tremendous burden, but our understanding of the neurogenetics of opioid addiction remains limited. Only recently have human postmortem brain studies of differential gene expression (DGE) associated with OOD been published: Corradin et al. 2022; Mendez et al. 2021; Seney et al. 2021; Sosnowski et al. 2022. The four independent studies had modest sample sizes (N=40-153 decedents), and none tested nominated genes for replication. Here we tested for cross-study replication and conducted a transcriptome-wide meta-analysis.

Methods: All studies had RNAseq data from human prefrontal cortex and DGE analysis by OOD status. We tested top genes reported by each study (gene N=138) for independent replication using a meta-analysis of the other three studies. In a discovery meta-analysis, we combined results from all four studies (Sample N=285 independent decedents; gene N=19,915) to identify OOD-associated genes. We used the weighted Fisher's test for all meta-analyses, and only genes with summary statistics in at least two studies were considered. Results: Among the 138 genes nominated across the four studies, 16 were independently replicated (Bonferroni adjusted $p < 3.6 \times 10^{-3}$), including *ARL4D*, *NPAS4*, and multiple *DUSP* and *EGR* genes. The discovery meta-analysis identified 47 associated genes at a Bonferroni corrected threshold ($p < 2.5 \times 10^{-6}$) and 875 genes with FDR <0.05. Top genes from the discovery meta-analysis included all significant genes from the replication analysis and additional genes both anticipated (e.g. *BDNF*, *FOSB*, *OPRK1*) and not anticipated (e.g., *TENT5A*, *ZNF254*) from model organism and neuroscience studies of addiction. Analyses of all three gene sets (16 replicated, 47 Bonferroni meta-analysis, and 875 FDR meta-analysis) showed significant enrichment for genes in the Orexin receptor pathway, with up to 25 of the 154 genes in this pathway (including *HCRTR2*) being association with OOD. Discussion: Independent replication and greatly increasing sample size through meta-analysis identified numerous robust associations with OOD. This large number of associations suggests broad gene dysregulation in the prefrontal cortex and highlights the Orexin receptor pathway as a key finding in humans, providing a clear link to more than a decade of model organism study of this pathway and addiction, including evidence of the orexin system as a promising target for opioid addiction treatment development.

Complex Traits Posters - Wednesday

PB1479*. Identifying phenotypic signatures of circulatory microRNAs

Authors:

R. Mustafa¹, M. Evangelou², M. Ghanbari³, A. Dehghan^{1,4,5}; ¹Dept. of Epidemiology and Biostatistics, Imperial Coll. London, London, United Kingdom, ²Dept. of Mathematics, Imperial Coll. London, London, United Kingdom, ³Dept. of Epidemiology, Erasmus MC, Rotterdam, Netherlands, ⁴Dementia Res. Inst., Imperial Coll. London, London, United Kingdom, ⁵MRC Ctr. of Environment and Hlth., Imperial Coll. London, London, United Kingdom

Abstract Body:

Plasma circulatory microRNAs (miRNAs) have the potential to act as disease biomarkers or therapeutic targets. Previous investigations on the role of miRNAs in complex phenotypes frequently focused on several well-known miRNAs or traits of interest, while their causality remains poorly understood. Here we investigated the associations of genetically determined miRNAs with a wide range of clinical phenotypes and assessed the causality. We measured 2,083 circulatory miRNAs in plasma samples of 2,178 participants in the population-based Rotterdam Study cohort using HTG EdgeSeq miRNA Whole Transcriptome Assay, followed by identification of cis-variants that affect miRNAs expression (cis-miR-eQTLs). Single cis-miR-eQTL or weighted genetic risk scores (miRNA-GRS) were used as a proxy for plasma circulatory miRNA levels in phenome-wide association studies (PheWAS) comprising 905 clinical diagnoses across 16 disease categories using hospital episode statistics data in the UK Biobank (N=423,419). Two-sample Mendelian randomisation (MR) was implemented to test for causality using the inverse-variance weighted method, with weighted median and MR-Egger as sensitivity analyses to rule out horizontal pleiotropy. Cis-miR-eQTLs were identified for 204 miRNAs consisting of single cis-miR-eQTLs for 85 miRNAs and multiple cis-miR-eQTLs to compute 119 miRNA-GRS. At FDR<0.05, we identified 29 associations for nine miRNAs with single cis-miR-eQTLs and 44 associations for 17 miRNA-GRS. We tested 37 out of 44 associations for miRNA with at least three cis-miR-eQTLs in MR-PheWAS, which suggested their potentially causal role in the observed associations. In particular, the associations between miR-329-3p and miR-543 with obesity-related traits were replicated using large genome-wide association studies consortia data. Collectively, our study highlights phenotypic signatures of circulatory miRNAs and potentially causal roles of several miRNAs in cardiometabolic traits.

Complex Traits Posters - Thursday

PB1480. Identifying Rare Variants Associated with Type 2 Diabetes Mellitus in GENNID: a Multiplex, Multiethnic, Family Study.

Authors:

L. Simon¹, J. Wan¹, K. Edwards¹, B. Wu², T. Norden-Krichmar¹, A. Freedland¹, S. Ferdos¹, S. Sohail¹, American Diabetes Association GENNID Study Group; ¹Univ. of California, Irvine, Irvine, CA, ²Univ. of Minnesota, Minneapolis, MN

Abstract Body:

Objective: Type 2 diabetes mellitus (T2DM) is an important global public health problem. T2DM prevalence differs by ethnicity in the US, with Black and Hispanic individuals having higher T2DM rates compared to Asians and Whites. While T2DM is a complex, polygenic disease with a strong genetic component, the thousands of variants found to be associated with T2DM through genome-wide association studies (GWAS) and linkage studies only explain a small percentage of the total heritability of T2DM. Complex composite risk measures that consider genetics such as common-variant genetic burden polygenic risk scores (PRS) have the potential to enhance T2DM risk prediction. Rare variants may have the potential to further improve the predictive power of PRS. Family-based studies can be especially useful for detecting and identifying rare variants because such variants can be enriched in an extended pedigree and segregate with the phenotype. We will leverage genetic data from the GENNID study to identify rare variants associated with T2DM in this family-based multiethnic population. Given that rare variants that segregate in a family are enriched compared to the general population, we expected to observe a higher frequency of these variants that would allow us to evaluate rare variant associations in T2DM and assess heterogeneity across GENNID ethnic groups.

Methods: The GENNID study target population consists of 1651 subjects in 340 families from African American (AA), European American (EA), Japanese American (JA) and Mexican American (MA) families. We first identified rare variants by parsing through GENNID data and compiling a list of variants present in each GENNID ethnic group that exhibited a MAF of <0.01. Identified novel variants and genes were compared across all GENNID ethnic groups to assess potential overlapping variants and genes of interest. We will use the RVFam statistical package within R to collapse each SNP within their respective genes using a logit-link generalized linear mixed effects model to test for associations between T2DM status and each genotyped SNP.

Results: We identified unique rare variants in each GENNID ethnic group: AA (248 subjects, 68 families, 20835 rare variants), EA (747, 158, 13799), JA (122, 16, 23584) and MA (534, 98, 22033) with minimal overlap.

Conclusions: Rare variants were identified in each GENNID ethnic group with minimal overlap. Future research will assess whether the addition of these rare variants into a traditional common-variant PRS model improves T2DM prediction in the GENNID target population. Collectively, these results and findings will help elucidate the role that rare variants play in T2DM risk quantification.

Complex Traits Posters - Wednesday

PB1481. Identifying shared genetic architecture across eye diseases using electronic health records

Authors:

A. Scalici¹, T. Miller-Fleming², J. HIRBO³, E. Knapik⁴, N. Cox²; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ Med Ctr., Nashville, TN, ³Vanderbilt Univ. Med. Ctr., Clarksville, TN, ⁴Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Vascular and nerve defects in the eye are major contributors to vision loss. Both myopia and glaucoma are associated with optic nerve damage, however, genetic correlation studies of the two diseases have produced varying results. Genetic correlation studies rely on GWAS summary statistics for SNP associations that often have no causal relationship to the phenotype of interest. To gain greater insights into disease mechanisms, we used both transcriptomic and electronic health record (EHR) data to implement gene and phenome-based approaches. To broadly examine shared genetic architecture across all eye diseases, cases were defined as subjects with at least one eye disease phecode in their EHR. For each case, we identified five age-matched controls with no eye disease phecodes in their EHR. Using BioVU, Vanderbilt University Medical Center's DNA biobank, we performed a transcriptome-wide association study (TWAS) to identify shared transcriptional profiles across eye diseases (N=70,493). We identified two genes (GPX7 & AC016590.3) with altered predicted gene expression associated with eye disease status. To identify eye disease comorbidities, we conducted a phenome-wide association study (PheWAS) using age-matched eye disease cases and controls in an independent non-genotyped study population with at least three visits to VUMC in five years (N= 663,228). This PheWAS identified strong associations between eye disease, cerebrovascular disease, neurological and motor deficits, and diabetes. Using the effect estimates from the significantly associated phenome as weights, we constructed a phenotypic risk score (PheRS), representing a weighted sum of a subject's comorbidities. This PheRS is predictive of eye disease status and is associated with altered predicted gene expression (POU1F1, PAK1, AC091100.1, TDRKH, RPL41, NEU2). Employing both the gene and phenome-based approaches used in this study identified both known and novel genes related to eye disease. Further functional analyses of these genes can expand the known pathways involved in eye disease pathogenesis and give greater insight into shared disease mechanisms.

Complex Traits Posters - Thursday

PB1482. Identifying the genetically dysregulated risk and preventative pathways in Alzheimer's disease

Authors:

X. Li¹, Y. Dai¹, A. Liu², B. S. Fernandes¹, Z. Zhao³; ¹UTHealth at Houston, Houston, TX, ²UTHealth., Houston, TX, ³Univ Texas HSC Houston, Houston, TX

Abstract Body:

Background: Alzheimer's disease (AD) is a common neurodegenerative disorder in the elderly, in which genetic factors play an important role. Polygenic risk scores (PRS) could estimate individual AD risk and guide early intervention. A considerable proportion of the older adults (≥ 75) carry a high AD risk but evade AD. On the other hand, many individuals with a low risk for AD eventually develop AD. We hypothesized that unknown counterfactors might be involved in reversing the PRS prediction, which might provide insights into AD pathogenesis and clinical intervention.

Methods: We built a computational framework to identify risk and preventative pathways in AD using PRS-based stratification. First, we calculated PRS (including and excluding the APOE region) based on AD GWAS summary statistics and a genotype dataset of 1,846 individuals from ROS/MAP, MSBB, and Mayo databases. Next, we subgrouped the individuals by their PRS, clinical and pathological diagnosis, including samples with high PRS (high-risk) vs. samples with low PRS (low-risk), and AD subjects vs. controls with similar PRS background. We then conducted a transcriptome-wide association study (TWAS) by differential analysis between groups, adjusting for sex, age, and the first five principle-component of individual genotypes in the following three models with/without the APOE region: 1) full-PRS and full-TWAS; 2) noAPOE-PRS and noAPOE-TWAS; 3) noAPOE-PRS and full-TWAS.

Results: Prediction performance of PRS including the APOE region has the highest AUC (clinical diagnosis: 0.81, pathological diagnosis: 0.73) with a set p-value threshold at 0.5. Excluding the APOE region in the PRS calculation led to a consistent decrease in AUC. Model 1 validated well-known AD-related pathways, including amyloid-beta formation, amyloid-beta clearance (preventative), tau protein binding, and astrocytes response to oxidative stress. In model 2, myelin maintenance (preventative) and microglia antigen processing were detected significantly enriched in the corresponding comparison, suggesting that they are AD-related pathways independent of the effort of APOE. In model 3, we replicated the APOE independent pathways from model 2 and part of the AD-related pathways identified in model 1.

Conclusion: We developed a framework to stratify individuals by their disease risk evaluated by PRS systematically. The TWAS-level comparisons among those groups identify the AD risk and preventative pathways, including many well-known pathways and pathways independent of specific components. The independent validation is warranted. This framework can be extended to other polygenic complex disorders.

Complex Traits Posters - Thursday

PB1483. Impact of Interleukin-4, Interleukin-13 Gene Polymorphisms and HLA-DQ Alleles on Genetic Susceptibility of Type-1 Diabetes Mellitus (T1DM) in Kuwaiti Children

Authors:

M. Haider¹, M. Al-Rushood¹, H. Alsharhan¹, M. A. Rasoul¹, M. Al-Mahdi², H. Al-Kandari³; ¹Kuwait Univ., Faculty of Med., Safat, Kuwait, ²Al-Adan Hosp., Al-Adan, Kuwait, ³Farwania Hosp., Farwania, Kuwait

Abstract Body:

Type-1 diabetes mellitus (T1DM) is a complex multifactorial disease in which both genetic and non-genetic factors are thought to interact and result in the disease onset. Cytokines play a crucial role in pathogenesis of autoimmune diseases possibly due to their effector and regulatory functions in immune and inflammatory responses. *IL4* gene (on chromosome 5q31-33) carries a single nucleotide polymorphism (SNP) in its promoter (-590C/T), which is associated with a number of autoimmune diseases. IL4 and IL13 are key components in induction of the Th2 lymphocyte phenotype and in the downregulation of Th1 lymphocyte phenotype. We determined the genotype frequency of *IL4* gene promoter polymorphism (-590C/T), *IL13* gene polymorphism p.(Arg130Glu) and HLA DQ alleles in Kuwaiti children with T1DM to investigate their role in genetic susceptibility. This study included 244 Kuwaiti children with T1DM and 200 controls. The criteria by International Society for Pediatric and Adolescent Diabetes (ISPAD) was used for the diagnosis of T1DM. The control subjects were healthy Kuwaitis; none had close relative with T1DM and were evaluated by a specialist. The genotypes for *IL4* (-590C/T) and *IL13* p.(Arg130Glu) gene polymorphisms were identified by PCR-RFLP methods. HLA-DQ alleles were determined by sequence-specific PCR method. The frequency of *IL4* (-590C/T) gene polymorphism showed a statistically significant difference between Kuwaiti T1DM patients and controls in the Dominant model of genetic analysis but not in the Co-dominant model in which the significant difference was detected only in the heterozygous state. For *IL13* gene polymorphism p.(Arg130Glu), a statistically significant difference between patients and controls was detected only for heterozygous genotype in the Co-dominant model but not in the Dominant model. In 55% T1DM patients, the HLA genotype was either DQ2/2 or in combination with a DQ8 allele. Collectively, 91% had either DQ2 or DQ8 alleles in different combinations. Amongst the T1DM patients with HLA-DQ2/2 genotype, 80% had CC, 15% had CT and 5% had TT genotype of the *IL4* gene polymorphism. In T1DM patients with DQ2/8 genotype, 69% had the CC, 22% had CT and 9% had TT genotype of the *IL4* gene polymorphism. Our data highlights the role of C-allele of *IL4* (-590C/T) and Q allele of *IL13* p.(Arg130Glu) gene polymorphisms along with HLA-DQ2 and DQ8 alleles in determining the genetic susceptibility of T1DM in Kuwaiti children.

Complex Traits Posters - Wednesday

PB1484. Improving Common Disease Risk Prediction with Clinical Risk Factors using Polygenic Scores in a Primary Care Physician Network of 35,000 Patients Followed for 16 years.

Authors:

R. Mandla^{1,2}, P. Schroeder^{1,2}, B. C. Porneala², J. C. Florez^{2,1,3}, J. M. Mercader^{1,2,3}, A. Leong^{2,1,3}; ¹Broad Inst Harvard & MIT, Cambridge, MA, ²Massachusetts Gen. Hosp., Boston, MA, ³Harvard Med. Sch., Boston, MA

Abstract Body:

Introduction:

Polygenic Scores (PS) can be used for risk prediction of common diseases with complex genetic architecture, e.g., coronary artery disease (CAD), type 2 diabetes (T2D), and chronic kidney disease (CKD). Yet, their informative value above clinical factors has not been established, as studies on their utility in longitudinal risk prediction in real-world clinical settings are limited.

Methods:

We identified patients of European ancestry receiving clinical care in the Mass General Brigham (MGB) Primary Care Physician (PCP) network with genetic data in the MGB Biobank. We extracted their demographic, medical history, and lab data in a 2-year window around their first encounter to construct clinical risk scores (CRS) for coronary heart disease (CHD) (Framingham CHD risk score: age, sex, smoking, total cholesterol, high density lipoprotein [HDL], systolic blood pressure [SBP], hypertension [HTN] treatment), T2D (Framingham T2D risk score: age, sex, family history of diabetes, BMI, SBP, HDL, triglycerides), and CKD (SCORED model - age, sex, and diagnoses of anemia, HTN, diabetes, CHD, congestive heart failure, peripheral vascular disease, and proteinuria). We used the PRS-CS method to calculate genome-wide PS from published genome-wide association study (GWAS) summary statistics. We then performed Cox models to estimate the effects of PS and CRS on incident disease over an observation time of up to 16 years, and compared model performance using hazard ratios per SD (HR (95% CI)), c-index, and likelihood-ratio tests (LRT).

Results:

Among 35,162 patients with genetic data, PS were significantly associated with incident disease (CHD: HR=1.22 (1.16-1.29), $p=2\times 10^{-13}$; T2D: HR=2.78 (2.53-3.05), $p=3\times 10^{-103}$; CKD: HR=1.49 (1.35-1.65), $p=2\times 10^{-14}$). For patients with both genetic and clinical data, adding PS to the CRS model significantly improved model performance (1,047 incident CHD cases: CRS only model -, c-index=0.735; combined model - c-index=0.742; LRT $p=9\times 10^{-12}$; 1,058 incident T2D cases: CRS only model - c-index=0.694; combined model - c-index=0.737; LRT $p=5\times 10^{-58}$; 827 incident CKD cases: CRS only model - c-index=0.819; combined model - CRS c-index=0.824; LRT $p=5\times 10^{-10}$).

Conclusions:

Adding genetic risk to clinical risk models modestly improved longitudinal prediction of genetically complex, common diseases in a large PCP network. The cost effectiveness of such incremental improvement may only be realized where genetic data has already become available.

Complex Traits Posters - Thursday

PB1485. Improving model predictivity and explainability by combining genotypes and immunophenotypes in SARS-CoV-2 infected people.

Authors:

A. Renieri^{1,2,3}, C. Fallerini^{1,2}, K. Zguro¹, G. Brunelli¹, G. Martelloni¹, M. Baldassarri^{1,2}, F. Fava^{1,2,3}, A. Degl'Innocenti¹, S. Furini¹, Y. Mueller⁴, GEN-COVID multicenter study, P. Katsikis⁴; ¹Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Italy, ²Med. Genetics, Univ. of Siena, Siena, Italy, ³Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Siena, Italy, ⁴Dept. of Immunology, Erasmus Univ. Med. Ctr., Rotterdam, Netherlands

Abstract Body:

We first demonstrated that coding variants contribute to COVID-19 severity with different weight and inverse correlation to allele frequency by the so-called post-Mendelian model (Fallerini et al 2021). We then modeled the COVID-19 disease in three different immuno-phenotypes by serum pro/anti-inflammatory, and anti-viral cytokine and anti-SARS-CoV-2 antibody measurements: i) low antibodies, ii) excess of inflammation, and iii) balanced immunotypes (Muller Y et al 2022). We now show here how the immunophenotype can be predicted from genotype and vice versa. “Low antibodies” phenotype patient group is characterised by heterozygosity of ultrarare (loss of function) variants in immunodeficiency genes, while “excess of inflammation” group is characterised by a combination of rare (>1/1000) or low frequency (>1% and <5%) (gain of function) variants in autoinflammatory genes and hypoandrogenic variants in sex development genes in males (not contrasting the excess of inflammation). We also demonstrated that individuals with an apparent “balanced immuno-phenotype” but severe COVID-19 are characterised by variants inducing susceptibility to thrombosis, therefore pointing to a third not immunomediated group of pathogenicity (see also abstract by Fallerini et al at this meeting). Finally, the post-Mendelian model has been improved by collapsing on gene category, according to allele frequency for separating likely gain and loss of function variants (such as immunodeficiency genes, autoinflammatory genes, sex development genes, thrombophilia genes etc) rather than single genes. This approach simplifies and therefore reduces the excess of genetic features, improving model predictivity and explainability.

Complex Traits Posters - Wednesday

PB1486. Improving the selection of variants included in polygenic risk scoring models

Authors:

C. Arehart¹, M. Lin², L. M. Evans¹, C. R. Gignoux²; ¹Univ. of Colorado, Boulder, CO, ²Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Abstract Body:

Polygenic risk scores (PRSs) have become a central tool in human genetics research as they provide a prediction framework in which externally derived genetic associations are used to predict an individual's propensity towards a polygenic trait. The recent growth of large-scale genome-wide association studies (GWASs) has captured many additive weak effect sizes across the genome and has provided ample training datasets for PRS models. However, it has been noted that PRSs tend to differ in accuracy between ancestry groups and these disparities have the potential to exacerbate health inequality, especially for groups of non-European ancestry. There have been many efforts to build software that performs well across a multitude of traits. These software packages, such as PRS-CS and LDpred2 to name a few, provide efficient methods to account for linkage disequilibrium and compute PRSs. The computational costs (in terms of hardware constraints, time, and money), however, have become a primary limitation for these models. To circumvent these costs a simple solution has been to shrink the scope of genetic variants included in the model. As such, many PRS workflows subset the genome to the set of HapMap3 variants which has been noted for its manageable number of variants (1.1M) that are generally well imputed, cover the whole genome, and pass quality controls. This upstream filtering step has yet to be optimized in the context of PRS. We aimed to find a set of representative variants (of a similar size) that leads to better PRS accuracy across traits and ancestry groups. Through principal component analysis of datasets including imputation accuracy, fine-mapping scores, functional scores, GWAS significance, regulation of gene expression, methylation, and linkage disequilibrium, we identified a set of 1.5M "tag" variants for potentially improved PRS accuracy. These variants are generally common, functionally relevant, well-represented by multiple ancestry groups, and only overlap with 265K variants from HapMap3. In development of PRS to predict atrial fibrillation, we have noted an AUC = 0.62 [0.62, 0.63] when using these variants which is +0.01 compared to using HapMap3. We will extend this framework to other traits, across different ancestry groups, and replicate with biobanks outside the Colorado Center for Personalized Medicine. We anticipate improvements in accuracy of PRS prediction at a similar or larger scale to that observed in atrial fibrillation.

Complex Traits Posters - Wednesday

PB1487. In silico analysis to examine COPD genetic associations using scRNA-seq-based CRISPRi approach in iPSC derived alveolar type 2 (iPSC-AT2) cells

Authors:

V. Malik¹, R. Werder², E. K. Silverman¹, X. Zhou¹, A. A. Wilson², M. Cho¹; ¹Channing Div. of Network Med., Dept. of Med., Brigham & Women's Hosp., Boston, MA, ²Ctr. for Regenerative Med., Boston Univ., Boston, MA

Abstract Body:

Background: Genome-wide association studies (GWAS) have identified sets of genetic variants associated with impaired lung function and chronic obstructive pulmonary disease (COPD). Genetic signals are enriched in fetal lung and specific lung cell types, including alveolar type 2 cells (AT2). Prior work has identified AT2 relevant GWAS genes via variant-to-gene mapping and gene expression studies, but the functional implications of these genes is poorly understood. **Methods:** We performed single cell RNA sequencing (scRNA-seq) and data analysis in induced pluripotent stem cell derived alveolar type 2 cells (iPSC-AT2) with CRISPR interference (CRISPRi)-based knockdown of a set of genes identified from variant to gene mapping and AT2 gene expression: *DSP*, *HHIP*, *RBMS3* and *ADGRG6*. We examined patterns of differential expression between control and knockdown, and further examined differential pathway and transcription factor analysis for *DSP* and *ADGRG6* using a Seurat-PROGENy-DoRotheA-CARNIVAL integrated workflow. **Results:** Differential gene expression identified similarities between *DSP* and *ADGRG6* knockdown. Significant pathways included cellular senescence, FoxO signaling, cytoskeletal organization, focal adhesion, ECM receptor interaction, as well as Wnt, mTOR (only in *ADGRG6* knockdown) and TGF-beta signaling pathways. CARNIVAL was used to identify the main signaling pathways using PROGENy and DoRotheA activity scores and identified MAPK, TGF-beta and ATF2 signaling pathway alterations in both *DSP* and *ADGRG6* knockdown. Additionally, in *ADGRG6* but not *DSP* knockdown, CARNIVAL identified alteration of SMAD2/3/4 pathway, Wnt and PS1 pathways, whereas Nrf2-ARE pathway alteration was identified only in *DSP* knockdown. **Conclusion:** In a differential gene expression analysis of four COPD GWAS candidate genes using CRISPRi knockdown in an alveolar type 2 cells (iPSC-AT2) model, we identified transcriptomic similarities between *DSP* and *ADGRG6* knockdown, and identified shared pathways regulating cytoskeletal organization, cell-cell interaction, oxidative stress response and proliferation. These findings require further experimental validation. **Keywords:** scRNA-seq, CRISPR interference, induced pluripotent stem cells, alveolar type 2, COPD

Complex Traits Posters - Thursday

PB1488. Increased burden of pathogenic variants in known aortopathy genes in individuals with early onset thoracic aortic dissection

Authors:

D. Guo¹, I. Marin², M. Boerio¹, S. LeMaire³, S. Prakash¹, A. Estrera¹, H. Safi¹, J. Coselli³, M. Bamshad⁴, D. Nickerson⁵, D. Milewicz⁶; ¹UT McGovern Med. Sch. at Houston, Houston, TX, ²UT Hlth.Sci. Ctr. at Houston, Houston, TX, ³Texas Heart Inst. at St. Luke's Episcopal Hosp., Houston, TX, ⁴Univ. of Washington, Seattle, WA, ⁵Univ Washington Sch Med, Seattle, WA, ⁶Univ Texas McGovern Med Sch, Houston, TX

Abstract Body:

Acute aortic dissections are life-threatening events that are preventable if at-risk individuals are identified. Up to 20% of dissection cases have evidence of a pathogenic or likely pathogenic (P/LP) variant in the 11 genes identified and validated for heritable thoracic aortic disease (HTAD). To identify genetic factors that predispose individuals to early onset, sporadic dissections (ESTAD) in the absence of clinical features of HTAD (family history and syndromic features), we performed whole exome sequencing (WES) on two cohorts of patients with early age of onset of dissection (1st batch ≤56 years old and 2nd batch ≤60 years old; recruited from 2009 - 2019)) and gender, and hypertension-matched controls. In the 2nd batch, WES identified 7.9% of patients have P/LP variants in 11 known HTAD genes, which replicated the finding in the 1st batch of WES (9.3%); no P/LP variants were identified in the controls. *FBNI* had the highest burden of P/LP variants in both cohort despite the fact that the cases were screened for clinical features of Marfan syndrome. Interestingly, when the *FBNI* cases are removed, individuals recruited before 2014 (prior to genetic testing being readily available) had P/LP variants in *COL3A1*, *TGFBR1*, or *TGFBR2*, all genes that cause HTAD associated with syndromic features, while all of the patients recruited after 2014 had P/LP variants in genes that predispose to HTAD in the absence of syndromic features, primarily *ACTA2* (p = 0.027). These data imply that genetic testing is identifying individuals in known HTAD genes associated with syndromic features, whereas the lack of syndromic features leads to a lack of diagnosis of P/LP variants in genes that are not associated with syndromic features.

Complex Traits Posters - Thursday

PB1490. Increased frequency of predicted loss of function variants in *PKP2* in African ancestry individuals

Authors:

A. Winters, D. Smelser, H. S. Rao, E. Carruth, DiscovEHR Collaboration, D. J. Carey, C. M. Haggerty; Geisinger, Danville, PA

Abstract Body:

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is an inherited cardiac condition that frequently presents with sudden death without prior clinical symptoms. Previous studies have found that variants in genes encoding desmosome proteins, including *PKP2*, account for about 50% of ARVC cases, however these variants have variable penetrance and expressivity. While prior studies have mostly been conducted in European ancestry populations, limited prior data have suggested a possible increased prevalence of desmosome variants in African ancestry individuals. To better understand the prevalence of these variants across ancestry groups, we evaluated the frequency of predicted loss of function (LOF) variants in *PKP2* across the MyCode population at Geisinger. Through the DiscovEHR collaboration, exome sequence data linked to longitudinal electronic health records are available for over 175,000 individuals. Consistent with prior findings, in MyCode we saw that *PKP2* LOF variants are more common in Black individuals, with 0.25% of self-reported Black vs 0.04% of self-reported White individuals carrying a predicted LOF variant. In the Genome Aggregation Database (gnomAD), the same difference in *PKP2* carrier frequency is seen; 0.26% of African ancestry vs 0.06% of European non-Finnish ancestry individuals are carriers. The same increased carrier frequency among African ancestry individuals is also observed in the All of Us database, a study of genetic variation across the US, with 0.29% of African ancestry vs 0.07% European ancestry individuals being *PKP2* carriers. We then leveraged the extensive phenotype data available for the MyCode population to compare Black to White *PKP2* LOF carriers on a variety of ARVC-related traits. We observed a non-significant trend of White carriers (18%) being more likely than Black carriers (10%) to have any ARVC-related phenotype including ventricular tachycardia or fibrillation or a history of cardiac arrest, arrhythmia, or cardiomyopathy (OR = 1.7 [95%CI: 0.2 - 15.6] adjusted for age and sex; p = 0.63). A similar non-significant trend was seen in All of Us, with European ancestry *PKP2* LOF carriers being more likely to have an ARVC-related trait compared to African ancestry carriers (OR = 2.9 [95%CI: 0.7 - 11.6] adjusted for age and sex; p = 0.14). This increased genetic prevalence associated with African ancestry without evidence of associated disease suggests phenotypic variation by ancestry. Future work is needed to uncover the genetic and/or environmental factors modifying these associations.

Complex Traits Posters - Wednesday

PB1491. Individuals with GCK-MODY are not at increased risk for common complications associated with T2D as measured by the Diabetes Complications Severity Index (DCSI)

Authors:

L. McEwen¹, E. K. Burrows¹, K. M. Schiabor Barrett¹, E. T. Cirulli¹, A. Bolze¹, J. J. Grzymalski^{2,3}, W. Lee¹, N. L. Washington¹; ¹Helix, San Mateo, CA, ²DRI, Reno, NV, ³Renown Inst. for Hlth.Innovation, Reno, NV

Abstract Body:

Introduction: It's known that those with GCK-MODY have elevated blood glucose measurements, are variably diagnosed with T2D, and may even start insulin or oral hypoglycaemic agents (metformin and glibenclamide) as a result. In GCK-MODY patients, these pharmacologic treatments have little to no effect on HbA1c levels, which are typically used to determine future risk for disease complications related to diabetes.

Objective: The lack of response in HbA1c to treatment supports the thinking that *GCK* carriers with slightly elevated glucose levels may actually not be diabetic in the traditional definition of the disease and may not benefit from being treated as such. To formalize these recommendations, studies comparing rates of complications associated with diabetes, instead of a diabetes diagnosis directly, in relevant *GCK* carriers and those with a T2D diagnosis must first be completed.

Methods: We calculated the Diabetes Complications Severity Index (DCSI) from available ICD-10 codes for over 370,000 UKB participants with baseline glucose measurements. The DCSI is a 13-point scale calculated from patient medical records designed to quantify the severity of diabetes-related complications and can be used to measure the adverse outcomes in people with diabetes. We compared DCSI scores by *GCK* carrier status (defined as carriers of rare coding variants in regions of the gene shown to be statistically associated via the "Power Window" method with high glucose levels), presence of a T2D diagnosis, and glucose levels across both groups.

Results: While we found that the mean DCSI score for T2D cases was significantly higher than that of controls (negative binomial regression $p=6.5e-277$), we found that *GCK* carriers did not have a corresponding increase in the DCSI scores despite their high prevalence of T2D diagnoses (28% of *GCK* carriers vs 10% in non-carriers). Additionally, looking specifically at individuals with glucose measurements falling at least 1 standard deviation above the mean, we see a significant reduction in DCSI scores when comparing *GCK* carriers to non-carriers ($p=0.004$).

Conclusion: Overall, our findings suggest the blood glucose threshold for a diabetes diagnosis may be different for GCK-MODY and that these individuals are not at risk for common complications associated with T2D based solely on carrier status or elevated glucose levels.

Complex Traits Posters - Thursday

PB1492. Insights from HostSeq :Canada's National Whole Genome Sequencing Cohort of 10,000 Canadians Infected with SARS-CoV-2

Authors:

E. Garg¹, S. Yoo^{2,3}, L. Elliott¹, R. Hung^{4,5}, A. Halevy², J. Brooks⁴, S. Bull^{4,6}, F. Gagnon⁴, C. Greenwood^{7,8}, J. Lawless⁹, A. Paterson^{4,2}, L. Sun⁴, J. Lerner-Ellis^{6,4}, R. Abraham¹⁰, J-M. Garant¹⁰, S. Turvey^{11,12}, N. Aziz², B. Knoppers⁷, M. Lathrop⁷, S. Jones¹⁰, S. Scherer^{2,4}, L. Strug^{2,4}; ¹Simon Fraser Univ., Vancouver, BC, Canada, ²The Hosp. for Sick Children, Toronto, ON, Canada, ³Univ. of Ottawa, Ottawa, ON, Canada, ⁴Univ. of Toronto, Toronto, ON, Canada, ⁵Lunenfeld-Tanenbaum Res. Inst., Sinai Hlth., Toronto, ON, Canada, ⁶Sinai Hlth.System, Toronto, ON, Canada, ⁷McGill Univ., Montreal, QC, Canada, ⁸Lady Davis Inst. for Med. Res., Jewish Gen. Hosp., Montreal, QC, Canada, ⁹Univ. of Waterloo, Waterloo, ON, Canada, ¹⁰Canada's Michael Smith Genome Sci. Ctr., BC Cancer Agency, Vancouver, BC, Canada, ¹¹Univ. of British Columbia, Vancouver, BC, Canada, ¹²BC Children's Hosp., Vancouver, BC, Canada

Abstract Body:

While Coronavirus Disease 2019 (COVID-19) is caused by infection with the RNA virus SARS-CoV-2, human genetic variation can modulate susceptibility to infection and the severity of the disease. To characterize these variations, support COVID-19 research across Canada and contribute to the international COVID-19 host genetics effort, Genome Canada and CGEn (Canada's National platform for genome sequencing and analysis) launched the HostSeq project. HostSeq is a national databank of whole genome sequences (WGS) and clinical data from 13 independent clinical and epidemiological studies enrolling SARS-CoV-2-infected participants from across Canada who are broadly consented for health-related research. By summer 2022, the total number of DNA samples received exceeded 9,000. Here, we introduce the HostSeq project and report preliminary results from genetic analysis using polygenic risk scores (PRS) to assess whether contributing variation to COVID-19 severity aligns with other international efforts. Samples were sequenced at 30X depth on Illumina's NovaSeq 6000 at the three CGEn sites and then processed together (including joint calling and data harmonization with detailed case report forms). Rigorous quality control protocols were followed for both sample and data quality. We calculated PRS with PRSice2 and analyzed the association of best-fit PRS with hospitalization status (adjusted by three genetic principal components, age, sex and age-squared). We used the COVID-19 Host Genetics Initiative (HGI v7) GWAS for PRS computation. The genomes that have been joint-called represent a diverse population structure covering 5 populations (N=2,992: 73% European, 7% East Asian, 6% Admixed American, 6% South Asian, 5% African, 3% other). The best-fit PRS (p-value threshold of 4.4e-51 using 1.1e6 SNPs) was significantly associated with hospitalization (p-value < 2.2e-16; N=1,069 after restricting analysis to participants with covariates in the European subset). These initial results provide support for the polygenicity of severe COVID-19 and demonstrate the alignment of the HostSeq databank with international COVID-19 research. HostSeq genomes and their harmonized clinical information are available to the global research community through a Data Access Agreement and summary data are available through publicly accessible data portals both at www.cgen.ca.

Complex Traits Posters - Wednesday

PB1493. Insights into pediatric inflammatory bowel disease subtypes using single-cell RNA-sequencing.

Authors:

M. R. Keever-Keigher¹, L. Harvey², V. Williams², J. J. Johnston¹, D. A. Louiselle², E. Grundberg¹, T. Pastinen¹, C. A. Friesen², C. Smail¹, V. Shakhnovich²; ¹Children's Mercy Res. Inst., Kansas City, MO, ²Children's Mercy Hosp., Kansas City, MO

Abstract Body:

Several chronic, painful, debilitating disorders of the gastrointestinal (GI) tract have origins in autoimmunity, and manifest as tissue damage from inappropriate overactivation of the inflammatory response in genetically predisposed individuals. One example is inflammatory bowel disease (IBD), which can be subclassified as Crohn's disease (CD) and ulcerative colitis (UC). IBD can overlap with other chronic inflammatory disorders associated with infiltration of eosinophils in tissues of the gastrointestinal tract, such as eosinophilic esophagitis (EoE) and eosinophilic duodenitis (EoD). The complex and multifactorial nature of IBD necessitates an approach that can begin to uncover the underlying molecular phenotypes contributing to the altered signaling pathways defining each IBD subtype and its overlap with eosinophilic disorders. In this study, we identified single-cell transcriptomic signatures of IBD subtypes in pediatric patients using peripheral blood mononuclear cells (PBMC). Single-cell RNA-sequencing (scRNA-seq) was performed on PBMCs isolated from the blood samples of treatment-naïve pediatric patients of differing diagnoses, including CD, UC, EoE, EoD, and controls with apparently healthy gastrointestinal tracts. A total of 39,622 cells from 35 individuals (7 CD, 9 EoD, 10 EoE, 3 UC, 6 controls) passed quality control measures. We identified 1,188 (FDR < 0.05) differentially expressed genes (DEGs) between IBD subtypes and controls across six distinct cell populations. We observed downregulation of estrogen Receptor 2 (ERS2) in CD4+ T cells for all IBD subtypes compared to controls, which is supported by previous findings showing ERS2 expression to be inversely correlated with IBD severity. Secretoglobin 3A1 (SCGB3A1) was the most strongly upregulated among all DEGs in CD4+ T cells, with greater expression in EoD individuals compared to controls. In mice, increased expression of Scgb3a1 has previously been associated with pathogenesis of inflammation and infiltration of eosinophils in the lungs. Ongoing work is currently underway to detect robust expression patterns between subtypes of IBD and to identify genes related to disease diagnosis and prognosis. Ultimately, we expect our pediatric IBD scRNA-seq resource will lead to IBD subtype-specific molecular signatures that could be used in the development of minimally invasive diagnostic assays and for drug response applications.

Complex Traits Posters - Thursday

PB1494. Integrating common and rare variants into a genetic risk score for Alzheimer's disease risk prediction.

Authors:

E. Suh¹, G. Lee¹, S-H. Jung², Z. Wen¹, L. Shen^{1,3}, D. Kim^{1,3}; ¹Dept. of Biostatistics, Epidemiology & Informatics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ²Dept. of Digital Hlth., SAIHST, Sungkyunkwan Univ., Samsung Med. Ctr., Seoul, Korea, Republic of, ³Inst. for BioMed. Informatics, Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Background: Polygenic risk scores (PRS) are commonly used to estimate an individual's risk for disease, recently including Alzheimer's disease (AD). However, PRS fails to demonstrate sufficient specificity and sensitivity in risk prediction and does not incorporate rare variants, which also influence AD risk. We generated a novel genetic risk score for Alzheimer's disease which combines a smaller set of more informative common and rare variants using known neuroimaging biomarkers to better predict those with high or low risk of disease, as well as mild cognitive impairment (MCI) conversion to AD. **Method:** Whole genome sequencing data of 1,704 patients were collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Common variants of each patient were aggregated into a gene-based burden to enable smooth model estimation and gene-level interpretation. We then used SAIGE-GENE+ to conduct a gene burden rare variant analysis for each participant (minor allele frequency ≤ 0.01). The risk score was estimated based on common gene burdens exhibiting high correlations with neuroimaging biomarkers, FDG- and AV45-PET, and significantly associated rare loci from the corresponding rare variant analysis. PRS were generated with clumping and thresholding using PLINK. We compared the performances between our risk score and PRS on two tasks: cases vs controls classification and MCI conversion prediction. **Result:** Our risk score demonstrated better prediction performance and improved distinguishability between cases and controls over PRS. Our score also showed lower risk for controls and MCI non-converter patients and, conversely, higher risk for AD and MCI converter patients compared to PRS. Additionally, our score is well-calibrated compared to PRS—the number of predicted AD and MCI converter patients increased smoothly as the risk score increased. **Conclusion:** We proposed a novel approach to calculate an AD-specific genetic risk score that integrates both common and rare variants. The main advantage of this method is that it uses AD-relevant loci for calculation, yielding a score that is a better predictor of AD risk. Experiments demonstrated that this score enhances performance of prediction accuracy, calibration, and interpretability over PRS.

Complex Traits Posters - Wednesday

PB1495. Integrating Genomic Risk into Absolute Risk Estimates for Coronary Heart Disease.

Authors:

M. Hamed, S. Saadatagah, A. Sherafati, H. Bangash, A. Miller, A. Kamzabek, I. Kullo; Mayo Clinic, Rochester, MN

Abstract Body:

Introduction: There is increasing interest in integrating polygenic risk scores (PRS) in the clinical setting. In the electronic medical records and genomics network (eMERGE) IV project, PRS for 8 common diseases in adults, including coronary heart disease (CHD), are being implemented in the clinical setting.

Methods: The Pooled cohorts equation (PCE) is used for assessing the 10-year absolute risk for CHD, with guideline-based recommendations provided for those at elevated risk. PRS can be combined with the PCE, since it is orthogonal to clinical risk factors, to calculate an integrated score (IS) for participants \geq 40 years. The IS will be returned in person to participants who are in the top 5th percentile of the PRS or have monogenic risk, and by letter to those who have a family history of CHD. A validated PRS is available for the three main population groups in the US: European, African American and Latino. The European population PRS will be used for East and South Asian ancestry, as it had comparable performance in our validation analyses. A genome informed risk assessment (GIRA) profile that includes PRS, monogenic risk, family history and 10-year CHD risk derived from PCE will be compiled for each participant. The GIRA provides guideline-based recommendations to manage individuals at high risk for CHD and will be deployed in the electronic health record (EHR) with linkage to clinical decision support. Outcomes such as screening for CHD and initiation of statin therapy will be assessed after the return of results.

Conclusion: We describe a method for integrating PRS into absolute risk estimates for CHD, along with monogenic risk and family history in a GIRA report.

Complex Traits Posters - Thursday

PB1496. Integration Analysis of Multi-Omics Data for Osteoporosis Biomarker Discovery Identification

Authors:

C. Qiu, K. Su, X. Zhang, A. Liu; Tulane Univ., New Orleans, LA

Abstract Body:

Osteoporosis is an increasingly serious public health problem, which is characterized by low bone mineral density (BMD). To date, the functional roles of the most reported biomarker that are associated with BMD remain unclear. To increase the understanding of the biological mechanisms of BMD variation, it is necessary to leverage multiple dimensional datasets to capture the complexity of biological networks. The recent advent of high throughput technologies has created an opportunity for integrating multi-omics data to build a comprehensive and dynamic model of the molecular changes in osteoporosis. In this study, we generated the multi-omics datasets with three data types (mRNA, methylation, and metabolomics) from 903 subjects. Hip BMDs were examined by dual-energy x-ray absorptiometry (DXA). By focusing on potential interesting biomarkers (differentially expressed genes, differentially methylated CpG sites, and differential metabolites) generated from single-layer analysis, we applied a data integration analysis for biomarker discovery using latent components (DIABLO) to identify the key omics biomarkers for BMD variation. We identified a multi-omics signature of 12 mRNA biomarkers, 6 metabolomics biomarkers, 47 methylation biomarkers, respectively. Correlation analysis showed a strong correlation between mRNA and metabolomics omics. In addition, we performed pathway analysis and identified that several biomarkers (e.g., *PCAT1*) have been associated with BMD variation through previous genome-wide association studies. In conclusion, we reconstructed the biological network in a multi-omics level and identified several novel biomarkers that may mediate variation in the risk of osteoporosis. Our results highlighted the advantages of using integration analysis to add value to the traditional single-layer analysis.

Complex Traits Posters - Wednesday

PB1497. Integration network-based analyses and Mendelian Randomization to screen drug repurposing candidates for osteoporosis

Authors:

D-Y. Liu¹, J. Greenbaum², H. Shen³, H-W. Deng³; ¹Tulane Univ., new orleans, LA, ²Tulane Univ., New Orleans, LA, ³Tulane Univ, New Orleans, LA

Abstract Body:

Osteoporosis (OP) is a systemic disease due to an imbalance in bone remodeling and bone resorption, increased skeletal fragility, and fracture tendency resulting in a growing public health issue. However, there are several shortfalls associated with existing therapies available for OP treatment such as long-term safety, side effects, and low efficacy. Given chemical or clinical trials are expensive and inefficient to characterize the therapeutic properties of drugs, drug repurposing (DR) has emerged as a potential tool with low overall development costs and a shorter development timeline. In the present study, a computational network-based integrated mendelian randomization approach was developed to screen 7811 candidate drugs to be effectively repurposed against OP. First, we interrogated transcriptomic and epigenomic biomarkers using publicly available RNA/ DNA methylation datasets of patients (n=121) with bone mineral density (BMD) by Weighted Gene Correlation Network Analysis (WGCNA). WGCNA identified hub gene pathways that described functions related to myeloid cell differentiation, regulation of osteoblast differentiation, ossification, etc. Totally, 157 compounds were screened primarily based on drug-disease proximity distance between osteoporosis modules and drug target modules in a PPI network. To further investigate the drug candidate's efficacy for OP, we utilized targetable expression quantitative trait loci (eQTL) to mimic exposure to the drug candidate and examine the causal effect on BMD. Significant Mendelian randomization results for 50 drug candidates were observed to reduce OP risk, suggesting an underlying mechanism with genetic support. These findings revealed alterations in core genes and pathways for OP, which may be useful for drug discovery and compound prioritization. However, given that computational studies cannot fully recapitulate a clinical trial, further research is needed to evaluate the effectiveness of the identified compounds for osteoporosis treatment.

Complex Traits Posters - Thursday

PB1498. Integration of proteomics quantitative trait loci into genetic association analysis of stroke in the African American population

Authors:

Y. Cai¹, A. J. Molstad², D. Levy³, A. Reiner^{1,4}, S. Wei^{1,4}, C. Kooperberg^{1,4}, L. Hsu^{1,4}; ¹Fred Hutchinson Cancer Ctr., Seattle, WA, ²Univ. of Florida, Gainesville, FL, ³Natl. Heart, Lung, and Blood Inst., Bethesda, MD, ⁴Univ. of Washington, Seattle, WA

Abstract Body:

Stroke is the third leading cause of death among African American (AA) people and the incidence rate is almost double that of European American (EA) people. Genome-wide studies suggest that there is a substantial genetic contribution (estimated heritability ~35%) to stroke risk in African ancestry populations; however, few variants have been identified to date. Recently, substantial efforts have been devoted to studying the association of genetic variation with proteomics, which is then used to inform genetic association for novel discovery. Most of the data in these studies have focused on populations of European ancestry and little has been done in AAs. The objective of this study is to identify variants predicting protein levels in AA participants, and to integrate this functional information into genetic association analysis of stroke in AAs.

We used OLINK Proteomics data on 1472 proteins for 1,537 AA and 883 EA participants from the Women's Health Initiative, and an additional 1062 EA participants from the Framingham Heart Study. We performed genome-wide analyses for each protein and identified proteomic quantitative trait loci (pQTL) using matrix eQTL. We undertook a two-stage discovery and replication design with 1,237 AA and 1,645 EA participants for discovery and 300 AA and 300 EA participants for replication. We adjusted for age, body mass index, smoking status, plate, batch effect, 10 genetic principal components and 20 PEER factors to account for hidden batch or technological effects in proteomic measurements. We set the genome-wide significance threshold 5×10^{-8} in the discovery set (for both cis- and trans- pQTLs), and for replication we used Bonferroni correction based on the number of pQTLs that passed the genome-wide significance threshold in the discovery stage. After pruning out correlated SNPs ($R^2 > 0.2$), a total of 1,678 pQTLs (538 proteins) for AAs were identified in discovery and 65.3% were replicated. For the EA population, a total of 1,153 pQTLs (367 proteins) were identified in the discovery and 50.5% were replicated. Among the 1096 replicated pQTLs in AA, we identified 758 unique pQTLs that were not present in EA results. These AA-specific pQTLs will help inform novel genetic regulations of protein expression in AA people.

Preliminary analysis also shows evidence of stroke-associated SNPs that are linked to discovered pQTLs. Four of 67 known SNPs for stroke were also AA pQTLs. The GWAS analysis of stroke from the MegaStroke study showed a significant enrichment in AA pQTLs. In future work, we plan to assess the genetic association of pQTLs with stroke risk in AA, and provide genetic insights for stroke risk in AA populations.

Complex Traits Posters - Wednesday

PB1499*. Integrative analyses to identify plausible target genes on Alzheimer's Disease genes with genetic evidence

Authors:

Y. Leung¹, B. Kunkle², B. Vardarajan³, E. Marcora⁴, E. Blue⁵, P. Kuksa¹, R. M. Cores⁶, A. DeStefano⁷, R. Mayeux³, L. Farrer⁸, A. Goate⁴, L-S. Wang¹, G. Schellenberg⁹, Alzheimer's Disease Sequencing Project; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Univ. of Miami, Miami, FL, ³Columbia Univ, New York, NY, ⁴Icahn Sch. of Med. at Mount Sinai, New York, NY, ⁵Univ. of Washington, Seattle, WA, ⁶Gladstone Inst.s for Neurological Disease and Data Sci. and Biotechnology, San Francisco, CA, ⁷Boston Univ Sch Med., Boston, MA, ⁸Boston Univ Sch Med, Boston, MA, ⁹Univ Pennsylvania Sch Med, Philadelphia, PA

Abstract Body:

Many publications report genes or loci altering risk or causes Alzheimer's Disease (AD). The quality of evidence presented is highly variable. To clarify which loci are valid versus potential false positives, the Alzheimer's Disease Sequencing Project Gene Verification Committee (GVC) was formed to **review evidence for published GWAS loci, and evidence if a specific gene is causal for AD**. The GVC first performed a literature search to select publications for developing **structured criteria for evaluating genetic evidence**. A **tier system** was developed (tier 1-7) to identify the quality of evidence, with tier 1 loci having the highest quality of evidence supporting an association. We divided published results into **2 phenotype classes**: 1) Class 1 - AD case definitions are based on clinical and/or neuropathology data (most rigorous class); 2) Class 2: both clinical/neuropathological defined AD and dementia are cases, where the phenotype is inferred by family history using questionnaire (e.g. proxy cases from the UK Biobank). We reviewed 176 variants from 7 papers and found 144 tier 1 loci. Of these, 23 are from Class 1 and 130 are from Class 2. For 36 loci, 2 independent association signals highlighted the same locus. We identified causal genes in 2 ways. First, expert review found 20 genes supported by both genetic (GWAS or linkage exists) and functional evidence. These include *APP*, *PSEN1/2*. Then, we performed integrative analyses of association signals using **OpenTargets, AMP-AD Agora and INFERNO** to identify plausible causal genes. OpenTargets performs systematic drug target prioritization via the variant to gene score. The Agora platform hosts evidence (based on AMP-AD brain data) for whether genes are related to AD. INFERNO unbiasedly identifies the tissue/cell types in which the regulation occurs. These **3 gene ranking methods are complementary to each other** as the model and data used vary. By analyzing 176, we found 9 variants targeting the same gene by all 3 methods, suggesting these genes have the highest confidence of being a true causal gene in AD. These include *ABCA7*, *CRI*, *ACE*, *BIN1*, and *CLU*. Besides, 4 genes (*ABCA7*, *RIN3*, *GPC2* and *NDUFS2*) were identified in the brain with proteomics or metabolomics support. These 8 genes serve as potential therapeutic targets for AD. Retrospective analyses show that therapeutic targets supported by genetic evidence are 2-3x more likely to succeed compared to targets without genetic findings support. Thus, genetic gene discovery work for AD is a valid source of therapeutic targets. Yet, before significant resources are invested in drug development, evidence supporting a causal gene needs to be critically evaluated.

Complex Traits Posters - Thursday

PB1500*. Integrative analysis of large-scale multi-ancestry genome-wide association study and single-cell omics data provides high-resolution insight of cell-types in the pathogenesis of type 2 diabetes.

Authors:

K. Suzuki^{1,2,3}, L. Southam⁴, K. Hatzikotoulas⁴, X. Yin⁵, C. Spracklen⁶, N. Rayner⁴, R. Mandla^{7,8}, A. Huerta^{7,8}, K. Lorenz^{9,10}, H. Taylor^{11,12}, DIAMANTE Consortium, VA Million Veteran Program, J. Mercader^{7,8,13}, J. Meigs^{7,13}, M. Boehnke⁵, M. McCarthy¹⁴, M. Vujkovic^{9,15}, A. Mahajan¹⁴, J. Rotter¹⁶, B. Voight^{9,10,17}, E. Zeggini^{4,18}, A. Morris^{1,4}; ¹Ctr. for Genetics and Genomics Versus Arthritis, Ctr. for Musculoskeletal Res., Div. of Musculoskeletal and Dermatological Sci., Univ. of Manchester, Manchester, United Kingdom, ²Dept. of Diabetes and Metabolic Diseases, Graduate Sch. of Med., The University of Tokyo, Tokyo, Japan, ³Dept. of Statistical Genetics, Graduate Sch. of Med., Osaka Univ., Osaka, Japan, ⁴Inst. of Translational Genomics, Helmholtz Zentrum München, German Res. Ctr. for Environmental Hlth., Neuherberg, Germany, ⁵Dept. of Biostatistics and Ctr. for Statistical Genetics, Univ. of Michigan, Ann Arbor, MI, ⁶Dept. of Epidemiology and Biostatistics, Univ. of Massachusetts-Amherst, Amherst, MA, ⁷Programs in Metabolism and Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, ⁸Diabetes Unit and Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁹Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA, ¹⁰Dept. of Systems Pharmacology and Translational Therapeutics and Dept. of Genetics, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, ¹¹Ctr. for Precision Hlth. Res., Natl. Human Genome Res. Inst., NIH, Bethesda, MD, ¹²British Heart Fdn. Cardiovascular Epidemiology Unit, Dept. of Publ. Hlth. and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, ¹³Dept. of Med., Harvard Med. Sch., Boston, MA, ¹⁴Oxford Ctr. for Diabetes, Endocrinology and Metabolism, Radcliffe Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, ¹⁵Dept. of Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, ¹⁶The Inst. for Translational Genomics and Population Sci., Dept. of Pediatrics, the Lundquist Inst. for BioMed. Innovation (formerly Los Angeles BioMed. Res. Inst.) at Harbor-UCLA Med. Ctr., Torrance, CA, ¹⁷Inst. of Translational Med. and Therapeutics, Univ. of Pennsylvania, Philadelphia, PA, ¹⁸TUM Sch. of Med., Technical Univ. of Munich and Klinikum Rechts der Isar, Munich, Germany

Abstract Body:

Type 2 diabetes (T2D) is a common metabolic disorder characterized by impaired glucose homeostasis affecting more than 400 million people worldwide. To gain insight into the pathophysiology of the disease, we integrated summary statistics from the largest multi-ancestry genome-wide association study (GWAS) of T2D (2,535,601 individuals including 428,452 cases) with publicly available single-cell chromatin accessibility (ATAC-seq) and transcriptome (RNA-seq) data from human adult and fetal tissues. We began by conducting joint analysis of the T2D GWAS summary statistics and ATAC-seq data of 222 distinct cell types from DESCARTES and CATLAS using fGWAS. ATAC-seq peaks in seventeen cell types, including adipocytes and endocrine/exocrine cells in the pancreas (alpha, beta, delta, gamma, acinar and ductal cells) and gastrointestinal tract (goblet, chief and enterochromaffin cells), were significantly enriched in T2D associations (false discovery rate (FDR) $q < 0.05$). We identified more than 100 independent T2D signals mapping to cell-type-specific ATAC-seq peaks that were computationally predicted to interact with gene promoters (for example, *HNF4A* and *WNT11* in beta cells, *HK1* and *HSPB7* in delta and gamma cells, and *ADH4* and *ADH1C* in adipocytes). T2D signals in cell-type-specific ATAC-seq peaks were jointly associated with fasting glucose adjusted for body mass index (FGadjBMI) and liver fat ($P = 0.000187$ and 0.0190 , respectively). T2D-risk alleles at signals mapping to ATAC-seq peaks in adult beta cells and fetal islet cells were associated with increased FGadjBMI and liver fat. We then conducted gene-set enrichment analysis using the T2D GWAS

summary statistics and 700 cell-type signature gene sets identified by multiple single-cell transcriptomic studies. Twelve gene sets were significantly enriched in T2D associations (FDR $q < 0.05$), including signatures for pancreatic islet endocrine cells ($P = 7.93 \times 10^{-11}$) and endocrine cells in the gastrointestinal tract such as G cells ($P = 0.000603$) and enterochromaffin cells ($P = 0.000742$). Taken together, these complementary approaches using single-cell chromatin accessibility and transcriptome data highlighted a potential role of secretory cells in the gastrointestinal tract, which is involved in the mechanism of action of widely used incretin-based medications for T2D, as well as pancreatic islet cells in the pathogenesis of the disease. Our study showcases the use of integrated analysis of large-scale GWAS and single-cell omics data to provide high-resolution insights into the pathophysiology of the disease.

Complex Traits Posters - Wednesday

PB1501. Integrative genomic and transcriptomic analyses revealed new driver variants associated with progression of the non-alcoholic fatty liver disease

Authors:

D. Kim^{1,2}, J-H. Choi¹, Y-H. Baek³, J-S. Jeong⁴, S-Y. Han⁵, Y. Lee⁶, G. Park⁷, Y-S. Lee⁷, H. Woo^{1,2}; ¹Dept. of Physiology, Ajou Univ. Sch. of Med., Suwon, Korea, Republic of, ²Dept. of BioMed. Sci., Graduate Sch., Ajou Univ., Suwon, Korea, Republic of, ³Liver Ctr., Dept. of Internal Med., Dong-A Univ. Coll. of Med., busan, Korea, Republic of, ⁴Dept. of Pathology, Dong-A Univ. Med. Ctr., Busan, Korea, Republic of, ⁵Liver Ctr., On Hosp., Busan, Korea, Republic of, ⁶Dept. of Cancer BioMed. Sci., Graduate Sch. of Cancer Sci. and Policy, Natl. Cancer Ctr., Goyang, Korea, Republic of, ⁷Div. of Rare Cancer, Res. Inst., Natl. Cancer Ctr., Goyang, Korea, Republic of

Abstract Body:

Non-alcoholic fatty liver disease (NAFLD) is a broad spectrum of liver diseases ranging from non-alcoholic fatty liver (NAFL) to non-alcoholic steatohepatitis (NASH). In this study, to obtain an integrative insight into genomic and transcriptomic alterations during NAFLD progression, we performed RNA-Seq profiling (n=146) and whole exome sequencing from the biopsied tissue samples (n=132). Cluster analysis of the transcriptome revealed the three subtypes of NAFLD, which showed distinct genetic variants and pathological findings. We report that macrophage activity was associated with the transcriptomic subtypes, showing the altered macrophage repertoire plays a critical role in NAFLD progression. The findings could be verified using independent data of the whole genome sequencing (WGS) from the PBMC samples (n=94). In addition, we identified the candidate genes whose variants were correlated with transcriptional deregulation and macrophage activity across the subtypes. In conclusion, our results could demonstrate a deeper molecular-level understanding, providing novel targets for the diagnosis and treatment of NAFLD progression.

Complex Traits Posters - Wednesday

PB1502. Interactions between Valley Fever and genetic ancestry: does genetic ancestry affect risk of developing disseminated coccidioidomycosis?

Authors:

S. Spendlove^{1,2}, S. L. Jensen^{2,3}, D. Orellana², A. V. Stephens⁴, G. R. Thompson^{5,6,7}, R. H. Johnson^{8,9}, A. Heidari^{8,9}, R. Kuran^{8,9}, M. Butte^{3,4}, V. A. Arboleda^{1,2,3,10}; ¹InterDept.al Bioinformatics Program, David Geffen Sch. of Med., UCLA, Los Angeles, CA, ²Dept. of Pathology and Lab. Med., David Geffen Sch. of Med., UCLA, Los Angeles, CA, ³Dept. of Human Genetics, David Geffen Sch. of Med., UCLA, Los Angeles, CA, ⁴Div. of Immunology, Allergy, and Rheumatology, Dept. of Pediatrics, David Geffen Sch. of Med., UCLA, Los Angeles, CA, ⁵UC Davis Ctr. for Valley Fever, UC Davis, Davis, CA, ⁶Dept. of Med. Microbiol. and Immunology, UC Davis, Davis, CA, ⁷Dept. of Med. and Div. of Infectious Diseases, UC Davis, Davis, CA, ⁸Dept. of Med., David Geffen Sch. of Med., UCLA, Los Angeles, CA, ⁹Valley Fever Inst., Kern Med., Bakersfield, CA, ¹⁰Dept. of Computational Med., David Geffen Sch. of Med., UCLA, Los Angeles, CA

Abstract Body:

Coccidioidomycosis is a fungal infection endemic to the American Southwest. While most patients are asymptomatic after infection or develop only a respiratory infection, a small proportion of individuals develop severe disseminated coccidioidomycosis (DCM), which is associated with significant morbidity and mortality and requires prolonged antifungal treatment. The factors that contribute to an individual's risk for developing DCM are multifactorial, however, the genetic basis of this risk remains unclear. Previous studies have shown that self-identified race and ethnicity (SIRE) is associated with risk of DCM, including a higher risk of DCM in individuals identified as African-American, Latino or Filipino. The previous epidemiological research has used race and ethnicity, which is a social construct and not always equivalent to genetic ancestry. While race is based on an individual's external appearance and how they identify, genetic ancestry is based on where individuals' genetic data cluster with reference populations. In this study we use a merged cohort of 479 individuals from the UC Davis Center for Valley Fever and 87 individuals from the Valley Fever Institute at Kern Medical, who have differing severity of coccidioidomycosis. We show that having 20% or more of one's genetic ancestry clustering with the genetics of individuals from the African continent is associated with a 16.3 odds of DCM. Meanwhile having 20% or more genetic ancestry clustering with the genetics of individuals from Europe gives a decreased odds of 0.2. We also use GWAS and admixture mapping methods to identify any areas of the genome that may be transmitting this risk in some individuals. This project will allow us to leverage genetic biomarkers to identify individuals at the highest risk for DCM and who would benefit from early-treatment with immunomodulatory therapies.

Complex Traits Posters - Wednesday

PB1504. Interplay of age and Alzheimer's disease diagnosis in estimating heritability

Authors:

G. Leonenko¹, E. Baker¹, J. Hardy², V. Escott-Price¹; ¹Cardiff Univ., Cardiff, United Kingdom, ²UCL Inst. of Neurology, London, United Kingdom

Abstract Body:

The genetic contribution to a phenotype is frequently measured by heritability, the proportion of trait variation explained by genetic differences. Both Alzheimer's disease (AD) and ageing have a strong genetic component. Substantial heterogeneity of SNP-based heritability estimates of AD have been reported. We aimed to investigate these estimates in a number of different data cohorts to determine if we can identify a consistent estimate of the AD heritability. We investigate the sources of this heterogeneity in two cohorts a) clinically assessed cases and population-based controls from the UK Biobank and b) pathologically confirmed cases and controls, and calculate heritability with genome-wide complex trait analysis (GCTA) software. The cohorts vary by sample size, age of cases and controls and the definition of AD and control phenotypes. We compute heritability in four settings a) for all common (minor allele frequency >5%) SNPs, b) excluding APOE region, c) excluding both APOE region and genome-wide association study (GWAS) hits, and d) using gene-set specific SNPs that were identified as being associated with AD. We find that heritability estimates vary between 30-47% with 5% population prevalence when accounting for population structure. It drops substantially when additionally adjusting for age in the age-mismatched cohort but remains almost unchanged in the pathology confirmed cohort of cases and controls (46%). The heritability estimates decrease by 5-10% when the APOE region is removed. In conclusion, substantial variability in the AD SNP-based heritability estimates may be caused by incorrect adjustment for age in the age-mismatched cohorts. The heritability remains almost unchanged in the pathology confirmed cohorts independently of age adjustment.

Complex Traits Posters - Thursday

PB1505. Interrogation of shared genetic architecture of infectious disease traits and Alzheimer's disease

Authors:

R. Sale¹, D. Zhou², E. R. Gamazon³; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Genetics Inst., Nashville, TN, ³VUMC Clare Hall, Univ. of Cambridge, Nashville, TN

Abstract Body:

Alzheimer's Disease (AD) is a progressive, neurodegenerative disorder characterized by pathological hallmarks including amyloid beta plaques and neurofibrillary tau tangles that contribute to neuronal death. Accumulation of these neuropathological hallmarks may occur years before symptom onset, making it difficult to determine a causal relationship between AD neuropathology and cognitive decline. Amyloid beta oligomerization and amyloid beta generation are part of the innate immune system and integral to pathogen entrapment and clearance. Furthermore, studies have also shown that elevated brain microbe levels correlate with AD. This has led to the hypothesis that increased microbial burden exacerbates beta amyloid deposition and contributes to AD progression. However, there are many barriers to measuring microbial susceptibility, progression, and burden and linking these measurements to AD in a longitudinal cohort. Independent studies have determined that genetics influence AD development and microbial susceptibility, infection, and adverse outcomes. However, there is limited understanding of how risk loci associated with microbial infection and adverse outcomes may influence neurodegenerative diseases such as AD. The purpose of this study is to determine the genetic architecture of infectious disease traits, and determine shared genetic architecture between infectious disease traits and AD. We have obtained 90,000 individuals from the BioVU database, Vanderbilt University's deidentified electronic healthcare record (EHR) coupled to genomic information, who have been determined as either a case or control for an infectious disease trait (defined using ICD9 and ICD10 billing codes). We conducted GWAS on the individual infectious disease traits and performed LD Score regression analysis and Mendelian Randomization analysis to determine: 1.) the correlation between the genetics of AD and an individual infectious disease trait and 2.) determine causality between the genetics of an infectious disease trait and AD. These analyses determine the relationship between the genetics of infectious disease and AD and provide additional insight into AD risk and susceptibility.

Complex Traits Posters - Wednesday

PB1506. Intragenic loci within *TOMM40* modulate *APOE* expression in human microglia

Authors:

A. M Ramirez¹, O. Oron², L. Wang¹, M. Vasquez², B. DeRosa¹, K. Celis³, j. young⁴, K. Nuytemans⁵, J. Vance¹, D. Dykxhoorn¹; ¹Univ. of Miami, Miami, FL, ²Hussman Inst. for human genomics, Miami, FL, ³Univ Miami, Miami, FL, ⁴Univ. of miami, Miami, FL, ⁵John P. Hussman Inst. for Human Genomics, Miami, FL

Abstract Body:

Previously we demonstrated that the ancestry-related risk for Late Onset Alzheimer's Disease is driven by a local genomic region (termed Local Ancestry; LA) around *APOEε4*. Furthermore, we showed that the expression of *APOE* is higher in the brain of individuals bearing European LA compared to those with African LA. In a follow-up study, utilizing reporter assays and Capture-C data we located, within the European LA, two intronic loci in the *TOMM40* gene (named B10 and B13) that increased *APOE* expression in microglia and astrocytes. Here, we used CRISPR interference/activation (CRISPRi/a) to validate the regulatory role of these two regions in *APOE* expression and identify the specific subregion responsible for the interaction between *TOMM40* and the *APOE* promoter. Human Microglial Clone 3 (HMC3) CRISPRi/a lines were produced by transducing inducible dCas9-VP64 (Activator), dCas9-KRAB (Interferer) or dCas9 (control) using lentiviral vectors. To direct the dCas9 constructs to our regions of interest, we generated multiplex vectors that encode four single-guide RNAs (sgRNAs) targeting either B10 or B13 loci. We used 4 different sgRNAs in each case to ensure full coverage of the tested regions (~850bp/loci). Additionally, we constructed multiplex vectors including only 3 out of the 4 sgRNAs targeted at the B10 locus, as well as a non-targeting vector control bearing polyT tracts to truncate the expression of the scaffold RNAs. We then transduced each of the multiplex vectors into the HMC3 CRISPRi/a lines. We induced expression of the dCas9 constructs for 2 or 6 days with Doxycycline. RNA was extracted and the expression of *APOE* and close by genes (e.g. *TOMM40*) was measured by qRT-PCR. *APOE* expression significantly increased when targeting B10 or B13 with dCas9-VP64 after two days of Doxycycline treatment. Six days after treatment the significance persisted only when targeting B10. No significant changes in *APOE* expression were observed when inducing dCas9-KRAB. The upregulated *APOE* expression was reduced by leaving the first ~200bp of B10 untargeted in the dCas9-VP64 line, suggesting this area is key to modulate *APOE* expression. These preliminary results support our previous findings that regions B10 and B13 may house regulatory elements that modulate *APOE* expression and, show the potential location of a ~200bp cis-active element within B10 that could be driving the modulatory effects of European LA on *APOE*. While dCas9-VP64 increased *APOE* expression, dCas9-KRAB did not show repression presumably due to the already low basal levels of *APOE* in these cells. The expression of *TOMM40* did not vary across cell lines supporting that the effect observed is *APOE* specific.

Complex Traits Posters - Thursday

PB1507. Investigating the cellular and molecular response to hyperglycemia in iPSC-derived cardiomyocytes

Authors:

O. Johnson¹, E. Matthews¹, J. Gutierrez¹, M. Ward²; ¹the Univ. of Texas Med. Branch, Galveston, TX, ²Univ. of Texas Med. Branch, Galveston, TX

Abstract Body:

An estimated 451 million individuals worldwide are diagnosed with diabetes mellitus. Hyperglycemia is both a diagnostic hallmark and promoter of various diabetes-related pathologies. In cardiovascular tissue, prolonged hyperglycemia can lead to a type of cardiovascular disease (CVD) called diabetic cardiomyopathy (DCM) in some individuals. Because most CVDs are irreversible, understanding susceptibility to DCM is critical to early diagnosis and risk mitigation. Individual susceptibility to DCM can be assessed before the onset of cardiac pathology by determining genetic risk. Gene-by-environment (GxE) interactions, differential responses to the same environmental perturbation between cells of different genotypes, can be utilized to find disease-relevant variants that may inform genetic risk to DCM. To identify these interactions and understand the genetic and mechanistic basis for DCM, we have developed an *in vitro* cardiomyocyte model using three genotyped induced Pluripotent Stem Cell (iPSC) lines derived from both male and female healthy individuals to test the effects of hyperglycemia on the molecular and cellular response to glucose in iPSC-derived cardiomyocytes (iPSC-CM). Given that increased blood glucose concentrations are associated with increased risk of the development of DCM in diabetic individuals, we use glucose as an environmental perturbation relevant to DCM to identify functional variants. We can reliably differentiate beating iPSC-CMs at purities greater than 85%, as measured by the expression of cardiac-specific troponin T. Glucose exposure at hypoglycemic, physiologic and hyperglycemic concentrations reveals differences in gene expression within two hours. In the hyperglycemic condition we observe increased mRNA expression of known glucose response genes, such as *TXNIP*, which is also implicated in DCM. Several gene expression changes are maintained following week-long glucose exposure times; however a handful of known glucose response genes, including *SLC2A4*, show the opposite direction of effect over time, suggesting gene-regulatory changes mediating the differences between early and late responses. Early gene expression changes coincide with cellular changes affecting cardiomyocyte function whereby iPSC-CMs exposed to hyperglycemic conditions display arrhythmic effects determined by calcium flux measurements, and increased mitochondrial ROS production. Using this model, we plan to measure the global transcriptional, epigenetic and cellular response to glucose across individuals and correlate these changes with each individual's genetic variants to gain insight into DCM susceptibility.

Complex Traits Posters - Thursday

PB1508*. Investigating the influence of genetic ancestry on gene-environment interactions on polygenic risk score and acculturation: Results from the Hispanic Community Health Study/Study of Latinos.

Authors:

J. Sharma¹, C. McArdle², M. Graff³, C. Isasi⁴, M. Daviglus⁵, G. A. Talavera⁶, K. North³, L. Fernandez-Rhodes², G. L. Wojcik¹; ¹Dept. of Epidemiology, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ²Dept. of Biobehavioral Hlth., Coll. of Hlth. and Human Dev., The Pennsylvania State Univ., University Park, PA, ³Dept. of Epidemiology, Gillings Sch. of Global Publ. Hlth., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁴Dept. of Epidemiology and Population Hlth., Albert Einstein Coll. of Med., Bronx, New York City, NY, ⁵Inst. for Minority Hlth. Res., Univ. of Illinois at Chicago, Chicago, IL, ⁶Dept. of Psychology, San Diego State Univ., San Diego, CA

Abstract Body:

Background: Many analyses of Hispanic/Latino groups in epidemiologic research aggregate participants into an assumed homogenous sample, despite diversity in the genetic makeup, environmental conditions, and sociocultural patterns in Hispanic/Latino groups. Using the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), we examined the role of self-identified background group and genetic ancestry proportions in gene-diet and gene-acculturation interactions on the relationship between BMI and a BMI polygenic risk score (PRS_{BMI}).

Methods: Univariate and multivariable generalized linear models were executed using complex survey weights to understand heterogeneity in the distributions of several environmental variables identified a priori by McArdle et al. (2021). To understand the influence of both indigenous to the Americas (AME) ancestry and background group identity on the relationship between BMI and PRS_{BMI}, we used 3 modeling comparisons: 1) full model with and without AME ancestry proportion, 2) full model stratified by AME ancestry quartile, and 3) full model stratified by both background group and AME ancestry tertile. Key gene-environment interactions were PRS-diet and PRS-age at immigration.

Results: After complex survey weighting, 7,075 analytical samples remained, representing 6 background groups: 602 Central Americans, 1,717 Cubans, 636 Dominicans, 2,560 Mexicans, 1,157 Puerto Ricans, and 403 South Americans. Heterogeneity, assessed through t- and chi-squared tests, persisted in the distributions of key sociocultural, environmental, and health-related variables among participants after stratifying by self-identified background group and in univariate regressions with AME ancestry. The performance of the PRS_{BMI} decreased with increasing AME ancestry proportion. Statistically significant GxE interactions did not retain significance upon stratification by ancestry quartile and by both ancestry tertile and background, except for PRS-age at immigration interactions in the first tertiles of Mexican ($\beta = -1.33$, $p < 0.01$) and Dominican ($\beta = 4.11$, $p = 0.004$) groups and in the second tertile of the Central American group ($\beta = -3.11$, $p = 0.049$).

Conclusions: The BMI - PRS_{BMI} relationship and its interactions with age at immigration and diet varied significantly in both direction and statistical significance between AME ancestry tertiles of various background groups. Future analyses incorporating gene-environment interactions should account for different modifying effects of sociocultural variables between Hispanic/Latino background groups in addition to genetic ancestry.

Complex Traits Posters - Wednesday

PB1509. Irina St. Louis

Authors:

I. St. Louis; Johns Hopkins, Baltimore, MD

Abstract Body:

Recombination-activating gene 2 (RAG2) encodes the protein subunit RAG2 of the RAG1/2 endonuclease complex. This complex is well-known for its critical role in generating the diversity of B and T cell receptors. Mutations that impact the function of RAG2 in vitro, in an animal model, and in humans result in impaired V(D)J recombination activity, and various forms of severe combined immunodeficiency (SCID). In this study, I performed database mining and evaluated the molecular epidemiology of the SNP rs121918573 (a C>T missense mutation) and examined its potential impact on RAG2 function. Rs121918573 was mapped to the putative zinc-binding site within the plant homeodomain (PHD) region of RAG2. The effect of SNP on RAG2 protein structure was analyzed via a three-dimensional structure prediction model and the impact of a Cys478Tyr mutation on zinc-dependent binding to DNA was visualized. The I-TASSER server was used for the structural prediction of the SNP effect on both DNA and peptide binding. Predictive analyses of the effects of rs121918573 mutation alone showed evidence of destabilization on the PHD region of the protein but otherwise showed no impact on RAG2 expression. This study provides extensive data mining and bioinformatics analysis of the genomic location of rs121918573 and the predictive effect on the RAG2 gene and protein structure.

Complex Traits Posters - Thursday

PB1510. Is Whole-Blood Cell Mitochondrial Copy Number Lower in Chronic Obstructive Pulmonary Disease in the NHLBI trans-omics for precision medicine (TOPMED) Program?

Authors:

A. Rocco¹, X. Liu², H. Rossiter³, R. Casaburi³, M. Cho⁴, M. Irvin¹, H. Tiwari⁵, C. Liu², D. DeMeo⁶, M-L. McDonald¹; ¹Univ. of Alabama at Birmingham, Birmingham, AL, ²Boston Univ., Boston, MA, ³The Lundquist Inst. at Harbor-UCLA Med. Ctr., Torrance, CA, ⁴Brigham and Women's Hosp., Duxbury, MA, ⁵Univ Alabama at Birmingham, Birmingham, AL, ⁶Harvard Med. Sch., Boston, MA

Abstract Body:

RATIONALE: Chronic obstructive pulmonary disease (COPD) is the fourth leading cause of the death in the United States. With age, the prevalence of COPD increases while mitochondrial copy number (mtCN) tends to decrease. Lower whole blood mtCN is associated with increased systemic oxidative stress. Cigarette smoking, the primary risk factor for COPD, is known to contribute to oxidative stress. We postulated mtCN would be lower in COPD patients than current or former smoker controls without COPD.

METHODS: Analyses were performed on DNA from whole-genome sequencing of whole blood of 3,785 participants in three studies: SPIROMICS (373 COPD, 291 controls), ECLIPSE (2,006 COPD, 315 controls) and Framingham Heart Study (FHS) (167 COPD, 633 controls). mtCN was estimated using fastMitoCalc. In each study, linear regression models were used to test the relationship between mtCN as predictor for COPD and adjusted for age, sex, and smoking pack-years. In FHS, the model was also adjusted for current smoking. Fixed effect meta-analysis was performed among the three studies using the R library meta.

RESULTS: In SPIROMICS, participants with greater mtCN were less likely to have COPD (OR= 0.996, P=0.02). The relationship direction between mtCN and COPD was consistent in ECLIPSE (OR_{ECLIPSE}= 0.999, P_{ECLIPSE}= 0.88) and FHS (OR_{FHS}=0.887, P_{FHS}= 0.26) but did not reach P<0.05. Meta-analysis of the three studies indicated, although effect directions were consistent, there was moderate heterogeneity in effect and the meta-analysis result was not significant (OR_{meta}=0.999, CI= 0.9965-1.0007, P=0.20, I²=50.8%).

CONCLUSION: We demonstrate that COPD participants in SPIROMICS were slightly more likely to have lower mtCN than current or former controls. However, this finding was not replicated in ECLIPSE and FHS. We analyzed data from three large studies with state-of-the-art whole-genome sequencing and note significant degree of effect size heterogeneity. Additional research will be performed to understand whether factors such as white blood cell counts as well as factors related to COPD pathology or severity such as degree of lung function decline, percent emphysema and chronic bronchitis account for effect size heterogeneity among the three studies.

Complex Traits Posters - Wednesday

PB1511. Isoform usage differences in schizophrenia.

Authors:

P. Giusti-Rodriguez¹, A. Abrantes², N. Ancalade², S. Sekle², F. Memic³, A. Dijkstra⁴, E. Tseng⁵, G. Sheynkman⁶, J. Hjerling-Leffler³, A. B. Smit⁴, P. Sullivan^{7,3}; ¹Univ. of Florida Coll. of Med., Gainesville, FL, ²Univ. of North Carolina, Chapel Hill, NC, ³Karolinska Inst.t, Stockholm, Sweden, ⁴Amsterdam UMC, Amsterdam, Netherlands, ⁵Pacific BioSci.s, Menlo Park, CA, ⁶Univ. of Virginia, Charlottesville, VA, ⁷Univ North Carolina, Chapel Hill, NC

Abstract Body:

The genetic architecture of schizophrenia is dominated by common variants of small effects, similar to most complex human traits. The majority of the schizophrenia GWAS SNPs reside in non-coding regions where they could act through a variety of regulatory mechanisms, including alternative splicing, which could result in differential isoform usage in schizophrenia cases compared to healthy controls. The brain expresses the most diverse repertoire of isoforms, yet our understanding of brain-specific isoform expression and the mechanisms of their regulation by RNA binding proteins (RBPs) are incomplete. Rbfox1 is an RNA binding protein that regulates splicing. The gene that encodes it, *RBFOX1*, is located on a GWAS locus for schizophrenia and major depressive disorder, and mutations of *RBFOX1* have been linked to autism and epilepsy. Furthermore, gene-set analysis of schizophrenia GWAS findings has shown an enrichment for genes with a binding site for Rbfox1 and Rbfox2, another important splicing factor. Long-read RNA-sequencing technologies have made dramatic improvements in throughput, accuracy, access, and cost. Our study aims to generate a comprehensive isoform survey of human dorsolateral prefrontal cortex (DLPFC) and to identify potential isoform level differences in schizophrenia cases and controls. We generated long-read RNA-sequencing data from post-mortem human DLPFC from schizophrenia cases (n=10) and controls (n=15) using the PacBio Sequel II system. We also generated complementary short-read RNA-seq data to examine isoform abundances. We used published and in-house pipelines to generate a comprehensive isoform survey of human DLPFC and to compare differential isoform usage in DLPFC from schizophrenia cases vs healthy controls. We identified differential isoform usage (DIU) in DLPFC from schizophrenia cases compared to healthy controls using established and in-house pipelines (IsoSeq, Cupcake, SQANTI3, IsoAnnot, and DiffSplice) and contextualized these findings at the transcript- and gene-level using pathway and network analyses. Our findings support DIU as a regulatory mechanism at play in the pathophysiology of schizophrenia. DIU may reveal schizophrenia-associated isoforms that impact protein-protein interactions (PPIs) and could nominate actionable targets.

Complex Traits Posters - Thursday

PB1512*. Key discoveries of the genetic basis of stuttering.

Authors:

D. Pruett¹, H. Polikowsky¹, D. Shaw¹, A. Scartozzi¹, H-H. Chen¹, L. Petty¹, R. Gordon², Y. Yu³, 23andMe Research Team, C. Huff⁴, S. Kraft⁵, J. Below², R. Jones¹; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³UT MD Anderson Cancer Ctr., Houston, TX, ⁴U.T. MD Anderson Cancer Ctr, Houston, TX, ⁵Wayne State Univ., Detroit, MI

Abstract Body:

Developmental stuttering is a speech condition characterized by syllable repetitions, prolongations, and involuntarily pauses. Despite a prevalence of ~1%, the cause of stuttering is unknown. Heritability studies have established that a genetic component for stuttering exists; however, the genetic factors impacting stuttering risk remain largely uncharacterized. This study utilizes the results of self-reported stuttering from a sample of more than 1 million individuals to further elucidate the genetic basis of stuttering. We conducted a GWAS in individuals genotyped by 23andMe, Inc. Cases were defined as those who self-reported ever having a stammer or a stutter (N=99,776), with all other individuals considered controls, comprising a total of 1,123,019 participants of African American, East Asian, European, and Admixed American ancestry. We performed ancestry and sex-specific association studies, which assumed an additive model for allelic effects and included age, five principal components, and genotype platforms as covariates. These 8 independent GWAS identified twenty-four loci that reached genome-wide significance. One significant risk variant ($p = 3.81 \times 10^{-16}$) was identified in the European-female (rs13107325, RAF= 8.13%, OR = 1.11) GWAS for a missense variant in *SLC39A8*, a gene encoding solute-carrier proteins critical for cellular transport in mitochondria. Additional loci identified significant variant signals that implicated genes *DCC*, *SRPK2*, *NMUR2*, *TSHZ2*, *MITF*, *VRK2*, *CAMTA1*, *MYO16*, *MMAB*, *CTNND2*, *SEMA6D*, *IRS2*, *COL14A1*, *PTPRQ*, *SHISA2*, *CYTH4*, *SORCSI*, and *RGCC* as contributing to stuttering risk. We conducted an enrichment test for gene modules using genes associated with variant signals. Gene modules comprised groups of functionally related gene modules Gerring et al. identified from GTEx tissue gene expression data. Gene module enrichment analysis identified enrichment for modules in multiple tissue models (e.g. pancreas, heart atrial appendage, esophagus mucosa etc.), implicating stuttering associated genes in biological processes such as metabolism, nervous system development, and cell signaling. These results highlight that stuttering risk factors are complex and polygenic and suggest that genetic variation impacts population-level stuttering risk. This large population study provides novel evidence for a neurological origin of stuttering, which may lead to improved treatment of disordered speech.

Complex Traits Posters - Wednesday

PB1513. Known Alzheimer disease candidate variants influence the severity of neuropathologic lesions underlying Alzheimer disease and related dementias.

Authors:

D. Godrich¹, J. Pasteris¹, E. Martin², G. Schellenberg³, M. Pericak-Vance², M. Cuccaro⁴, W. Scott¹, T. Montine⁵, G. Beecham¹, The Alzheimer's Disease Genetics Consortium; ¹Univ. of Miami, Miami, FL, ²Univ. of Miami Miller Sch. of Med., Miami, FL, ³Univ Pennsylvania Sch Med, Philadelphia, PA, ⁴John P. Hussman Inst. for Human Genomics, Miami, FL, ⁵Stanford Univ., Palo Alto, CA

Abstract Body:

Introduction: The hallmark neuropathologic lesions underlying Alzheimer disease (AD) include neurofibrillary tangles and amyloid plaques. However, these AD lesions seldom appear in isolation; instead, they often co-occur with AD related dementia (ADRD) lesions including lewy bodies (LB), cerebral amyloid angiopathy (CAA), arteriolosclerosis (ARTE), hippocampal sclerosis (HS), vascular brain injury (VBI), and TAR DNA-binding protein 43 (TDP-43). While AD genetics has been well studied, the genetics of the lesions underlying AD/ADRD has been largely overlooked. Herein, we report on the effect of known AD variants on AD and ADRD lesions, instead of a binary clinical diagnosis of AD. **Methods.** We used the National Alzheimer's Coordinating Center dataset, identifying 4,972 autopsied individuals with neuropathology and genetic data. Lesions were ranked ordinally based on increasing severity. We pulled 23 known AD-associated SNPs from a recent AD GWAS of clinically defined AD [Kunkle, 2019]. We used ordinal logistic regression models to test for associations between AD candidate SNPs and lesion severity, adjusting for age at death and sex. Finally, we investigated the combined effect of the independent variants, excluding *APOE*, by constructing genetic risk scores (GRSs). **Results.** Regression models showed 10 of the 23 previously identified variants were strongly associated with severity of tangles: *CR1*, *BINI*, *INPP5D*, *MEF2C*, *TREM2*, *PICALM*, *SORL1*, *ACE*, *APOE*, *CASS4* and of plaques: all the same besides *MEF2C* but including *FERMT2*. The ADRD neuropathologic lesions also showed some associations with known AD variants. We found *APOE* to be highly associated with CAA, LBs, HS, and TDP-43, but not ARTE or VBI. Severity of LBs and ARTE was associated with *BINI*, but of opposite effect in ARTE. CAA was associated with *PICALM* and *ACE*. HS was associated with *TREM2* and VBI with *CASS4*, but both of opposite effect from the AD lesions. TDP-43 showed a strong association with *ACE*. Finally, we found the GRS to be predictive only for the AD lesions and LBs. **Conclusions.** We showed that there was some overlap in the known AD variants that were shared in the ADRD lesions. Interestingly, although the AD lesions and CAA are correlated, there does not seem to be a highly shared genetic architecture, besides *APOE*, in these lesions. Additionally, the lack of association of *APOE* and other AD variants with the vascular lesions (ARTE and VBI) may suggest that even though these lesions frequently co-occur, their genetic profiles may differ. Finally, the lack of predictive value of the GRS in the ADRD lesions suggests that there is a need to improve our understanding of the genetic landscape of ADRD.

Complex Traits Posters - Thursday

PB1514*. Large scale genome-wide association study highlights the causality of vasospastic angina

Authors:

K. Hikino¹, S. Koyama¹, K. Ito¹, M. Koido², T. Matsumura³, R. Kurosawa⁴, K. Tomizuka⁴, Y. Koike¹, S. Ito¹, Y. Momozawa⁴, T. Morisaki⁵, Y. Kamatani⁶, The Biobank Japan Project, T. Mushiroda¹, C. TERAO⁴; ¹RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, ²RIKEN Ctr. for Integrative Med. Sci., Yokohama, Kanagawa, Japan, ³Jichi Med. Univ., Tochigi, Japan, ⁴RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, ⁵The Univ. of Tokyo, Inst Med Sci, Minato-ku, Tokyo, Japan, ⁶The Univ. of Tokyo, Minato-ku, Japan

Abstract Body:

Background: Vasospastic angina (VSA) results from coronary artery spasm and is more common in East Asians. VSA is an important disease that can lead to fatal conditions such as myocardial infarction and sudden death, but the genetic factors that may cause the disease have not yet been fully explored on a genome-wide level. Methods: A genome-wide association study (GWAS) of VSA was performed using the BioBank Japan (BBJ) data with 5,242 cases and 150,352 controls in total. The effect sizes of previously reported coronary artery disease (CAD) -susceptible loci were then compared between those with VSA and non-VSA CADs. Additionally, we compared the effect size of the variants associated with VSA with those of other CADs, which were also analyzed separately based on sex and age. Finally, we analyzed the mortality from AMI in patients who did not originally have CAD by using subsequent follow-up data. Results: We identified that rs112735431, a missense deleterious variant in *RNF213* gene (p.R4810K) in chromosome 17, had strong associations with VSA in all discovery, replication, and combined studies [odds ratio (OR): 2.98, 95% confidence interval (CI) 2.29 - 3.88, $p = 5.3 \times 10^{-16}$ in combined study]. This variant is an Asian-specific, low-frequency variant that has been experimentally linked to reduced angiogenesis and is a well-known cause of Moyamoya disease (a highly prevalent vascular disease in East Asians). Most of the previously reported CAD-associated variants showed higher absolute effect sizes in non-VSA CAD than in VSA, but only rs112735431 showed higher effect size in VSA. The effect sizes of rs112735431 in VSA were also distinctly different when compared to those in other CADs. Furthermore, in VSA patients, homozygous carriers of rs112735431 showed a penetrating association (OR: 13.80, 95%CI: 3.55 - 53.65, $p = 1.5 \times 10^{-4}$) compared to the expected heterozygous carriers in the additive model. In stratified analyses, consistent with epidemiologic findings, the risk allele for rs112735431 was more strongly associated in males (omnibus $p = 0.035$) and in the younger age group. Finally, a strikingly high fatality rate in AMI was observed in follow-up data in risk allele carriers of rs112735431, who did not have CAD at the time of BBJ enrollment (Hazard ratio:2.91, 95%CI: 1.65 - 5.12, $p = 2.1 \times 10^{-4}$). Conclusions: The rare missense variants in *RNF213*, rs112735431, identified in this first large-scale GWAS of VSA may be partially responsible for the high frequency of VSA in East Asian populations. The association between rs112735431 and high mortality from AMI may indicate a strong contribution of vascular cell dysfunction to the progression to AMI via coronary spasm.

Complex Traits Posters - Wednesday

PB1515. Large-scale chromatin accessibility fine-mapping of blood lipids GWAS loci in human liver cell types.

Authors:

B. Wenz¹, S. Ramdas¹, M. Caliskan², M. Dudek¹, K. Creasy¹, R. Aubin¹, J. Judd³, D. Xin¹, K. Olthoff¹, A. Shaked⁴, P. Camara¹, S. Grant⁵, D. Rader⁶, C. Brown¹; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Daiichi Sankyo, Chesterfield, NJ, ³Stanford Univ., Stanford, CA, ⁴Hosp Univ Pennsylvania, Philadelphia, PA, ⁵Children's Hosp. of Philadelphia/Univ. of Pennsylvania, Philadelphia, PA, ⁶Univ of Pennsylvania, Philadelphia, PA

Abstract Body:

Genome wide association studies (GWAS) have identified thousands of genetic loci associated with a variety of common, complex human traits. For the vast majority of these loci, however, the true causal variant, target gene, and mechanism of action leading to these associations remain unclear. Most GWAS variants have high levels of linkage disequilibrium (LD) and/or fall in non-coding genomic regions, complicating fine-mapping and mechanistic studies. The impact of genetic variants that are associated with gene expression has been well-studied, suggesting that gene expression regulation plays a causal role at many complex trait associated loci. However, the upstream impacts of genetic variation on transcription factor binding and chromatin accessibility has not been comprehensively investigated. We assessed the landscape of chromatin accessibility in 189 human liver tissue samples using ATAC-seq. We identify more than two million regions of open chromatin in the human liver, many of which exhibit tissue specific regulatory activity. The depth of these data allowed for nucleotide resolution transcription factor footprinting to infer transcription factor binding sites (TFBS) in our samples. Integration of single nucleus RNA sequencing (snRNA-seq) from a subset of human liver samples has enabled prediction of TF binding, chromatin accessibility, and gene expression in specific liver cell types. Combined with whole genome sequence data from these samples, we identified thousands of cis-regulatory elements whose chromatin accessibility is associated with local genotypes, called chromatin accessibility quantitative trait loci (caQTLs). At dozens of blood lipids GWAS loci, we achieve cell type specific single nucleotide resolution to identify the specific TF binding site disturbed that leads to a co-localized caQTL, eQTL, and GWAS signal, which allows us to generate specific molecular hypotheses about the mechanisms leading to disease. Integration of chromatin conformation capture data generated in liver cell models allows us to test many of these hypotheses. These findings highlight the benefits of integrating multiple cellular traits for the identification and characterization of disease-causing variants and contribute to basic understanding of genetic and epigenetic regulation of gene expression in human liver tissue.

Complex Traits Posters - Thursday

PB1516. Large-scale exome sequencing identifies common and rare coding variants contributing to tinnitus

Authors:

A. Ayer¹, GHS-RGC DiscovEHR Collaboration, Regeneron Genetics Center, M. Drummond², G. Coppola¹, K. Praveen¹; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Regeneron Pharmaceuticals, Tarrytown, NY

Abstract Body:

Tinnitus, defined as the perception of sound without external stimulus, is a highly prevalent condition that affects 10-20% of the adult population worldwide. Chronic or severe tinnitus may negatively impact quality of life and has been associated with depression, anxiety, insomnia, cognitive impairment and psychiatric diagnoses. Current treatment options are limited to symptom management via noise suppressing devices and therapy focused on coping techniques, with little ability to target the causal pathways of the disorder. Identifying genes associated with tinnitus can elucidate relevant biological pathways, which may then be targeted to treat the underlying biological causes of tinnitus. We performed genome-wide and exome-wide association analyses of tinnitus in the largest such study to date in Europeans, including 27,409 cases and 566,160 controls from UK Biobank, Geisinger DiscovEHR, the Malmö Diet and Cancer Study, Mount Sinai's BioMe Personalized Medicine Cohort, Penn Medicine Biobank, and FinnGen. Cases were defined using self-reported and medical record diagnoses of tinnitus and controls excluded individuals diagnosed with hearing loss, which is highly correlated with tinnitus. We identified 13 loci associated with tinnitus, 3 of which are also hearing loss risk loci (*NID2*, *CTBP2* and *CRIP3*). In rare (minor allele frequency < 1%) variant analyses, we identified genome-wide significant associations with intronic variants in *CACNA1A* and *NFE2L2*. We also observed associations with missense variants in *SYNJ2* (Thr656Met; OR=1.42; p=3e-08), *COL11A2* (Phe80Ser; OR=11.29; p=2e-09) and aggregates of predicted loss-of-function variants within *KLHDC7B* (OR=2.06; p=5e-08) and *TRIOBP* (OR=1.89, p=3E-07), all of which have previously been associated with hearing loss in humans. *COL11A2* mutations can cause Stickler syndrome and tinnitus can be a symptom in affected individuals. Our results suggest that *SYNJ2* and *KLHDC7B* variants, which are associated with increased risk for hearing loss, may also confer a risk for tinnitus that likely accompanies the loss of hearing in carriers. Our analysis has identified novel common variant associations and is the first to interrogate the contribution of rare variants to tinnitus.

Complex Traits Posters - Wednesday

PB1517. Leveraging aggregation of juvenile idiopathic arthritis in large pedigrees to investigate susceptibility genes.

Authors:

C. Avery¹, J. F. Bohnsack², A. Hersh², S. Thompson³, M. Sudman³, S. Prahalad⁴, L. Jorde¹; ¹Dept. of Human Genetics, Univ. of Utah Sch. of Med., Salt Lake City, UT, ²Dept. of Pediatrics, Univ. of Utah Sch. of Med., Salt Lake City, UT, ³Div. of Human Genetics, Dept. of Pediatrics, Cincinnati Children's Hosp. Med. Ctr. & Univ. of Cincinnati, Cincinnati, OH, ⁴Dept. of Pediatrics, Emory Univ. Sch. of Med., Atlanta, GA

Abstract Body:

Juvenile idiopathic arthritis (JIA) is a complex rheumatic disease, affecting approximately 1 in 1000 children. While the etiology of JIA is largely unknown, genetic factors have been predicted to contribute to the development of JIA. Familial studies have estimated that the relative risk of JIA for siblings of probands is approximately 11.6x that of the general population. However, clinical heterogeneity and low disease prevalence have impeded investigation of heritable risk factors. Recent genome-wide association studies have suggested that up to 22 genes are associated with JIA susceptibility but are limited by their capacity to detect rare variants with high-effect size. Pedigree-based studies minimize genetic heterogeneity and target variants that are rare in the population but enriched in high-risk families. This study utilizes whole genome sequencing in 22 small multiplex pedigrees to identify genes harboring excess burden, and 10 large multi-generational pedigrees to identify segregating shared genomic segments (SGS) from a common founder to identify coding or regulatory risk variants for a dominant mode of inheritance. In our small multiplex pedigrees, we have identified a candidate coding variant in *ACVRI*, a component of BMP signaling, and a recessive candidate variant in *FOXPI*, recently implicated in a meta-GWAS of JIA. In our large pedigrees, SGS analysis identified several interesting segregating loci shared in cases and longer than one would expect by chance, including a genome-wide significant region flanking the HLA (1.8 Mb) and several genome-wide suggestive loci, including a 2 Mb region on chromosome 12 that includes *PARPBP* (PARP-1 binding protein), which positively regulates PARP-1. Interestingly, PARP-1 is a known regulator of various inflammatory disorders. These initial findings have the potential to refine and build upon previous efforts to identify heritable susceptibility to JIA, highlight key biological processes in JIA progression, and address the contribution of rare variants to JIA risk.

Complex Traits Posters - Thursday

PB1518. Leveraging genetic variation to evaluate risk factors and therapeutic opportunities for aortic valve stenosis: a mendelian randomization analysis.

Authors:

H. Urquijo¹, R. J. Richardson², T. R. Gaunt¹, T. G. Richardson¹; ¹MRC Integrative Epidemiology Unit (IEU), Bristol, United Kingdom, ²Faculty of Life Sci., Sch. of Physiology, Pharmacology and NeuroSci., Bristol, United Kingdom

Abstract Body:

Aims: Aortic valve stenosis (AVS) is an increasingly prevalent disease with no pharmacotherapies available. In this study, we used mendelian randomization (MR) to estimate the genetically predicted effects of established and emerging risk factors on AVS risk, as well as evaluate drug repurposing opportunities.

Methods and Results: Univariable MR analyses using genetic data of up to 462,927 UK Biobank participants supported previous MR findings indicating several established risk factors, such as elevated blood pressure and lipoprotein lipids, increase AVS risk. We additionally identified calcium (OR=1.26 per SD increase, 95% CI=1.10-1.43, $P=7 \times 10^{-4}$) and lifetime smoking (OR=1.92 per SD increase, 95% CI=1.55-2.39, $P=5 \times 10^{-9}$) as risk factors for AVS. Applying the same analysis on coronary artery disease (CAD) risk found comparable estimates with the notable exception of calcium which provided limited evidence of an effect on CAD (OR=1.06, 95% CI=0.97-1.15, $P=0.23$). Multivariable MR provided evidence that apolipoprotein B is the predominating lipoprotein trait increasing AVS risk when conditioned on low density lipoprotein cholesterol (LDL-C) and triglycerides. Finally, we found a genetically predicted effect of inhibition of antihypertensive target *ADRB1* on lower AVS risk. LDL-C lowering target *PCSK9*, lipoprotein(a) lowering target *LPA* and triglyceride lowering targets *ANGPTL4* and *LPL* also were genetically predicted to lower AVS risk.

Conclusions: We provide a comprehensive evaluation of modifiable risk factors driving the increased healthcare burden of AVS, which should be considered in risk stratification and prevention strategies. Furthermore, our findings provide valuable insight into the prioritisation of the pharmacotherapies which may yield clinical benefit towards treating AVS.

Complex Traits Posters - Wednesday

PB1519. Leveraging spatial patterns in the distribution of dental caries to resolve genetic heterogeneity.

Authors:

S. Haworth¹, J. R. Shaffer², A. Esberg³, P. Lundberg⁴, K. Divaris⁵, M. Marazita⁶, I. Johansson³; ¹Bristol Dental Sch., Univ. of Bristol, Bristol, United Kingdom, ²Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh; Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, ³Section of Cariology, Dept. of Odontology, Umeå Univ., Umeå, Sweden, ⁴Section of Molecular Periodontology, Dept. of Odontology, Umeå Univ., Umeå, Sweden, ⁵Div. of Pediatric and Publ. Hlth., Adams Sch. of Dentistry, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁶Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, PA

Abstract Body:

Objective:

Dental caries (i.e., tooth decay) is a common disease with an incompletely understood genetic basis. The distribution of caries-affected tooth surfaces in the mouth follows recognizable spatial patterns, where groups of tooth surfaces have similar caries experience (termed caries clusters). We aimed to test whether analysis of these clusters would yield improved insight into the genetic basis of caries experience.

Methods:

The study included Swedish adults from the SIMPLER and VIKING studies. Quantitative traits were derived for five different caries clusters using electronic dental records and results of previously published cluster analysis. For each trait, genome-wide association statistics were generated and results from the studies were combined using fixed effects meta-analysis. SNP-based heritability estimates and estimates of genetic correlation between pairs of traits were estimated using the SumHer method, implemented in LDAK5.2.

Results:

Analysis included 18,819 adults, with mean age 74 years in SIMPLER and 64 years in VIKING. The caries traits were moderately heritable, with estimates of SNP-heritability between 0.12 and 0.20. At the genome-wide level, the caries clusters captured traits with distinct genetic profiles. As an example, the cluster representing caries in the occlusal surfaces of molar teeth had little genetic overlap with the other caries clusters (rg estimates between -0.39 and +0.11).

Differences in genetic architecture were also seen at the single variant level. As an illustration we examined the effects of rs1122171, a common variant located near *PITX1* which has previously been reported to associate with caries in adults. This variant had strong effects on caries in upper incisor teeth (0.079 standard deviations per T allele, $P=1.3 \times 10^{-11}$) but weak effects on caries in the occlusal surfaces of molar teeth (0.019 standard deviations per T allele, $P=0.071$)

Conclusion:

Different genetic risk loci are associated with caries experience at different tooth surfaces. Leveraging spatial patterns through clustering provides one way to resolve heterogeneity in the genetic architecture of caries. This may be useful to boost statistical power or provide more precise understanding of different disease mechanisms at different sites in the mouth. Genome-wide analysis of these caries clusters will now be included in GLIDE2.0, the second round of the oral and dental health GWAS consortium.

Complex Traits Posters - Thursday

PB1520. Leveraging the nutritional geometry framework to dissect the impact of macronutrient composition on metabolic function and gene regulation in fat tissue

Authors:

K. Farris¹, A. Crean², D. Sobreira¹, R. Barrès³, S. Simpson², M. Nobrega¹; ¹Univ. of Chicago, Chicago, IL, ²Univ. of Sydney, Sydney, Australia, ³Univ. of Copenhagen, Copenhagen, Denmark

Abstract Body:

Changes in macronutrient composition, even in the absence of changes in caloric intake, can have significant impacts on metabolic health and lifespan in mice. However, the functional and molecular underpinnings of these changes remain largely unexplored. Using the nutritional geometry framework, we have measured the impact of ten different isocaloric ratios of proteins, carbohydrates, and fat on metabolic function and gene regulation in fat tissue in mice. We can therefore dissect the impact of individual macronutrients on changes in gene regulation across the diets and identify possible mechanisms underlying the observed metabolic changes. Across the ten diets, we saw significant changes in metabolic measures such as body weight and glucose tolerance and identified 5,644 differentially expressed (DE) genes and 4,308 differentially spliced (DS) exons. The gene regulation changes we observed were predominately correlated with fat content in the diets, although there were still substantial contributions from protein and carbohydrate content. Further, the significant expression and splicing changes occurred in largely different genes, with only 967 genes being acted on by both modes of regulation, indicating that expression and splicing regulation are likely involved in separate processes in response to changes in macronutrient composition. To explore this further, we performed gene ontology (GO) analysis on sets of DE or DS genes that respond to each macronutrient. We found different GO terms enriched in response to the different macronutrients and modes of gene regulation. For example, in the set of DE genes with a significant positive correlation with fat content, the top GO term was cilium organization. In investigating this signal further, we found that it was being driven in part by 9 DE genes that were all associated with Bardet-Biedl syndrome (BBS). BBS is a rare autosomal recessive ciliopathy whose symptoms include obesity. Mouse knockouts of BBS-associated genes have demonstrated an association between some BBS genes and increased food intake. However, to our knowledge BBS-associated genes have not previously been associated with response to diet. We did not find cilium organization as a top GO term for the other macronutrient response groups, demonstrating the power of the nutritional geometry framework to provide novel insights into the gene regulatory response to changes in diet. By disentangling the response to individual macronutrients, we can provide novel insights into potential mechanisms underlying the overarching metabolic changes that are observed in response to changes in macronutrient composition.

Complex Traits Posters - Wednesday

PB1521. Limited effect of Y chromosome variation on coronary artery disease and mortality in UK Biobank.

Authors:

J. Wilson¹, P. R. Timmers²; ¹Univ Edinburgh, Edinburgh, United Kingdom, ²Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract Body:

The effect of genetic variation in the male-specific region of the Y chromosome (MSY) on coronary artery disease and cardiovascular risk factors has been disputed. We systematically assessed the association of MSY genetic variation on these traits using a kin-cohort analysis of family disease history in the largest sample to date.

We tested 90 MSY haplogroups against coronary artery disease, hypertension, blood pressure, classical lipid levels, and all-cause mortality in up to 152,186 unrelated, genomically British individuals from UK Biobank. Unlike previous studies, we did not adjust for heritable lifestyle factors (to avoid collider bias), and instead adjusted for geographical variables and socioeconomic deprivation, given the link between MSY haplogroups and geography. For family history traits, subject MSY haplogroups were tested against father and mother disease as validation and negative control, respectively. Our models find little evidence for an effect of any MSY haplogroup on cardiovascular risk in participants. Parental models confirm these findings.

Kin-cohort analysis of the Y chromosome uniquely allows for discoveries in subjects to be validated using family history data. Despite our large sample size, improved models, and parental validation, there is little evidence to suggest cardiovascular risk in UK Biobank is influenced by genetic variation in MSY.

Complex Traits Posters - Wednesday

PB1522. Longitudinal transcriptomic profiling of IBD patients reveals therapeutic decision-making determinants of drug response.

Authors:

L. Cervera Seco¹, M. Sanchez-Mayor¹, A. M. Corraliza^{2,3}, A. Salas^{2,3}, J. Panes^{2,3}, U. Marigorta^{1,4}; ¹Integrative Genomics Lab, Ctr. for Cooperative Res. in BioSci.s (CIC bioGUNE), Basque Res. and Technology Alliance (BRTA), Bizkaia Technology Park, Derio, Spain, ²Dept. of Gastroenterology, IDIBAPS, Hosp. Clínic, Barcelona, Spain, ³Centro de Investigación Biomédica en Red de Enfermedades Hepática y Digestivas (CIBERehd), Barcelona, Spain, ⁴IKERBASQUE, Basque Fndn. for Sci., Bilbao, Spain

Abstract Body:

Inflammatory Bowel Disease (IBD) is a complex immune-mediated inflammatory disease that triggers a series of recurrent relapse-remission phases, leading to a chronic inflammation of the gut, and ultimately to severe outcomes like surgery. Although association studies have made deep inroads into the genetic component associated with diagnosis, our understanding of the factors contributing to drug response remains especially poor. Primary response to biological drugs is low, and secondary loss of response is prevalent. Even if the remission rates are low (around 30% of patients), the therapeutic repertoire available for clinicians will be expanding in the immediate years. It is therefore pressing to develop an understanding of the determinants of response to individual treatments with advanced therapies in IBD patients.

We have carried out a longitudinal follow-up including 147 patients from whom we obtained 264 intestinal biopsy and 119 blood RNA-seq samples. Our study includes samples from patients treated with the three main biological drugs approved, namely infliximab ("anti-TNF"), ustekinumab ("anti-IL12/IL23") and vedolizumab ("anti- α 4 β 7 integrin"), with 248, 47, and 88 samples, respectively. We also have around 3 timepoints (0, 14 and 46 weeks) for a total of 232, 92, and 59 samples, respectively. A first exploratory analysis using the full dataset confirms that tissue and, to a lesser extent, disease subtype are the main determinants of transcriptomic variability among IBD patients. Regarding drug response across time, ustekinumab response is associated with pathways related to extracellular recognition, whereas infliximab is putatively involved in translation- and general immune-related pathways. Interestingly, ustekinumab alters a particularly low number of genes across time points, suggesting that patients are not efficiently responding, or that gene modulation takes place at a slow rate. To improve our understanding of individual response to treatment, we carried out trend analyses based on connectivity scores that serve to classify patients into responders and non-responders. This allows uncovering new biological pathways associated to drug response profiles in particular patients. Besides providing a renewed understanding of the molecular effects of biological drugs, we propose that the creation of personalized models to explain longitudinal response to treatment is feasible and a promising venue to develop therapeutic positioning guidelines of biological drugs in IBD.

Complex Traits Posters - Thursday

PB1523. Loss of chromosome Y in leukocytes as a risk factor for critical COVID-19 in men.

Authors:

B. Bruhn-Olszewska¹, H. Davies¹, D. Sarkisyan¹, U. Juhas², E. Rychlicka-Buniowska², M. Wójcik², M. Horbacz², M. Jakalski², P. Olszewski², J. Orzechowski Westholm³, A. Smialowska³, K. Wierzba², A. Torinsson Naluai⁴, N. Jern⁴, L-M. Andersson⁴, J. Järhult¹, N. Filipowicz², E. Tiensuu Janson¹, S. Rubertsson¹, M. Lipcsey¹, M. Gisslen⁴, M. Hultstrom¹, R. Frithiof¹, J. Dumanski¹; ¹Uppsala Univ., Uppsala, Sweden, ²Med. Univ. of Gdansk, Gdansk, Poland, ³Stockholm Univ., Stockholm, Sweden, ⁴Univ. of Gothenburg, Gothenburg, Sweden

Abstract Body:

COVID-19 pandemic, which has a prominent social and economic impact worldwide, is characterized by a bias towards higher severity and mortality in males. The loss of chromosome Y (LOY) is a male-specific factor, associated with age-related diseases, cancer, and all-cause mortality in males. Recent studies demonstrated that LOY has a significant impact on the transcription of immune genes in leukocytes. Therefore, in this study, we tested the hypothesis that LOY is a plausible factor responsible for the severity of COVID-19 infection in males. We re-analyzed publicly available scRNA-seq data of PBMC samples collected from critically ill male patients and identified low-density neutrophils (LDNs) as the cell type most affected with LOY. These results, were confirmed by ddPCR in sorted leukocytes from the newly collected cohort, consisting of 211 men with severe COVID-19, showing that LOY above 5% cut-off was present in 46% of LDNs, 32% of granulocytes, and 29% of monocytes samples. In particular, LDNs occur in patients with severe COVID-19 and contribute to the hypercoagulable state of the disease. Remarkably, we demonstrate that LOY level is significantly associated with COVID-19 severity and mortality classified by the WHO grade. The analysis revealed that the level of LOY was higher in the whole blood of patients who died due to COVID-19, compared to the survivors. Similarly, critically ill patients with WHO grade 10 had significantly more LOY than patients with WHO grade 9 and 6. Moreover, the percentage of LOY in sorted subpopulations of leukocytes correlated with clinical variables including increased thrombocyte and decreased erythrocyte counts, as well as increased levels of IL-4 and IL-5. Interestingly, the follow-up analysis of recovered patients revealed a decrease in the number of LDNs and monocytes accompanied by a reduction of LOY level in PBMC, suggesting that LOY could be a consequence of emergency myelopoiesis. Collectively, our results imply that LOY is associated with the dysregulation of immune genes as well as in the shift in the composition of leukocyte subsets and might be involved in the generation of poor outcomes for men with COVID-19. Our results might also be relevant for other common viral infections showing similar male bias.

Complex Traits Posters - Wednesday

PB1524*. Loss-of-function variants association study highlights potential new causal genes implicated in blood pressure regulation

Authors:

E. Lecluze¹, G. Lettre²; ¹Montreal Heart Inst., Montréal, QC, Canada, ²Université de Montréal, Montreal, QC, Canada

Abstract Body:

Hypertension is a widely spread cardiovascular disease, affecting 1.28 billion peoples worldwide and leads to the death of 7.5 million people each year. Blood pressure (BP) is tightly regulated by a complex network relying on the kidney, the brain, the heart, the blood vessels, and the endocrine system, and influenced by environmental and genetic factors. For more than a decade, genome-wide associations studies (GWAS) performed in millions of individuals have associated over 1,400 SNPs with BP variation. However, how these genetic associations can highlight new genes and pathophysiological mechanisms remain elusive. Rarer variants can have large effect size on BP, but the GWAS design tend to ignore them. Among these, many coding variants are predicted to impair gene function by modifying the sequence of the encoded protein, and association of such variants directly incriminates important regulators of the trait. In this study, we focused our analyses on rare loss-of-function variants to highlight novel genes that may be implicated in BP regulation. We performed an association study with rare loss-of-function (LoF) variants in the 500,000 whole-exome sequencing dataset of the UK Biobank. We used additive and recessive models to test for association, as well as a burden gene-based test that aggregates the effect of rare LoF variants (minor allele frequency (MAF) ≤ 5 in the same gene. This led to the association of some well-known regulators of blood pressure, such as *NOS3*, implicated in nitric oxide production, a vasodilator (gene-based test, diastolic blood pressure, P-value= 1.1×10^{-8}). *FES* loss-of-function variants are also associated with BP variation (gene-based test, systolic blood pressure (SBP), P-value= 1.2×10^{-7}). *FES* encodes a tyrosine kinase controlling cell differentiation and adhesion; the gene has never been implicated in BP regulation, but the locus carries common non-coding SNPs associated with BP and coronary artery disease. *CASP9* is also a new gene of interest, as the frameshift variant rs2234723 (MAF=21%) is associated with the trait (additive test SBP, P-value= 3.5×10^{-11} ; $\beta = -0.3$). It lies in a complex locus (chr1:15429787-15677590) of several genes associated with glomerular filtration rate and SBP. Even though they have predicted loss-of-function variants, none of them are significant in our analysis, suggesting that *CASP9* may be the causal BP gene at this locus. Our focus on loss-of-function variants associated with BP directly incriminate potential causal genes that may be implicated in its regulation, setting the stage for future functional studies to better understand these potentially novel pathophysiological mechanisms.

Complex Traits Posters - Thursday

PB1525. Lung-specific alternative splicing is a shared etiological risk for COVID-19 severity and chronic respiratory diseases.

Authors:

T. Nakanishi¹, Y. Farjoun², S. Zhou³, B. Richards⁴; ¹McGill Univ., Montréal, QC, Canada, ²Lady Davis Inst., JGH, Montreal, QC, Canada, ³Lady Davis Inst. of Jewish Gen. Hosp., Montreal, QC, Canada, ⁴McGill Univ., Montreal, QC, Canada

Abstract Body:

Despite the successful development of vaccines and treatments, a striking number of individuals are still suffering from COVID-19, which requires us to identify mechanistic targets for therapeutic development and/or drug repurposing. Alternative Splicing (AS) is an essential mechanism for generating tissue-specific functional diversity in the form of protein isoforms, and it has been implicated that AS plays an important role in the immune response to infections. However, the causal role of AS in COVID-19 pathogenesis and its potential therapeutic relevance are not fully understood. Here, we show the etiological evidence of lung-specific AS for COVID-19 using a Mendelian Randomization (MR) approach and we demonstrate its shared genetic mechanisms with other chronic respiratory diseases. We first obtained the *cis*-regulatory genetic determinants of 8,201 gene splicing (*cis*-sQTLs) of 4,477 genes in lung and whole blood, the two most relevant tissues for acute SARS-CoV-2 infection, from GTEx Consortium. We then performed two-sample MR to assess whether these *cis*-sQTLs were associated with COVID-19 outcomes in the GWASs from the COVID-19 Host Genetics Initiative, encompassing up to 219,692 COVID-19 cases. In addition to the replication of our previous findings with *OAS1*, we identified that AS in *ATP11A*, *DPP9*, and *NPNT* in lung, rather than their total RNA levels, were associated with COVID-19 severity. Specifically, decreased use of alternative splice junctions, chr13:112875941-112880546 (hg38) in *ATP11A* and chr19:4714337-4717615 in *DPP9*, and increased use of alternative splice junction, chr4:105898001-105927336 in *NPNT* afford these associations. Colocalization analyses supported shared genetic signals of COVID-19 severity with idiopathic pulmonary fibrosis in the *ATP11A* and *DPP9* loci, and that with chronic obstructive lung disease in the *NPNT* locus. Taken together, our study suggests the shared etiological importance of AS in lung between COVID-19 and other chronic respiratory diseases. Given the tissue-specific nature of AS, those mechanism could be prioritized for drug discovery which may act in a tissue-specific manner. We anticipate our findings will motivate the ongoing effort to map the disease-relevant AS to the corresponding isoform or protein product, by means of emerging technologies such as long-read sequencing and high-throughput protein quantification.

Complex Traits Posters - Wednesday

PB1526. Machine learning in COVID-19 phenotype prediction

Authors:

G. Novelli¹, S. Casali², V. Colona¹, D. Cocciadiferro³, V. Favalli², R. Giannini¹, A. Latini¹, M. D'Apice⁴, V. Ferradini¹, M. Andreoni¹, P. Rogliani¹, F. Leonardis¹, M. Perrone¹, I. Qunti⁵, A. Novelli⁶, M. Biancolella¹; ¹Tor Vergata Univ, Rome, Rome, Italy, ²Bases Italia S.r.l., Campospinoso, Italy, ³Bambino Gesù Hosp. Rome, Rome, Italy, ⁴Tor Vergata Univ, Hosp. Rome, Rome, Italy, ⁵Sapienza Univ, Rome, Rome, Italy, ⁶Bambino Gesù Hosp., Rome, Rome, Italy

Abstract Body:

Since the beginning of the pandemic, numerous clinical, genetic and social COVID-19-related studies have been performed. Data science plays a key role in the identification of molecular mechanisms underpinning virus-host interaction, which represents a major challenge in a proactive perspective of precision medicine and to promote sustainable and rational public health interventions. In the present study, we trained a Random Forest classification model on target genetic data (variants identified by WES and functional analysis) to predict host response to SARS-CoV-2 infection. The dataset was composed by 128 samples and subsequently split in test and validation datasets, following a stratified k-fold cross-validation approach. The two target classes, representing disease severity (severe disease vs asymptomatic), were slightly imbalanced, with severe disease cases being overrepresented. After data cleaning and basic parameters tuning, model average classification accuracy was 0.77. In a second experiment, we included age and gender as basic clinical features for our machine learning model: average classification accuracy increased to 0.84. However, age distribution turned to be imbalanced, with a high percentage of the samples being seventy-five years old or older. Further data are needed to clarify the possible interplay between genetic and clinical features. However, this study demonstrates the development and application of Machine Learning (ML) to host-genome with integrative analyses, that allows not only to optimize polygenic risk score analysis, but also to refine the stratification of COVID-19 patients, to estimate an accurate prognosis and to set individual and targeted therapeutic approaches. This study was also supported by a grant of Regione Lazio (Italy, Progetti di Gruppi di Ricerca 2020 A0375-2020-36663 GecoBiomark).

Complex Traits Posters - Thursday

PB1527*. Machine learning-based classification of Age-related Macular Degeneration using gene expression profiles

Authors:

R. Ratnapriya¹, H. Nakajima¹, A. Barman²; ¹Baylor Coll. of Med., Houston, TX, ²Rice Univ., Houston, TX

Abstract Body:

Genome-wide association studies (GWAS) in humans have established a strong genetic component of AMD that is mostly driven by common variants. However, the progress in translating genetic findings into understanding disease pathways and mechanisms has been very slow. Gene expression dysregulation is emerging as a dominant model in complex diseases. However, accessing the transcriptome profiles in large disease and non-disease cohorts are challenging, which present significant limitation in identifying genes and pathways involved in the disease process using existing methods such as differential gene expression analysis. In this study, we apply machine learning (ML) methods to uncover gene expression signatures associated with Age-related Macular Degeneration (AMD). We use RNAseq data from 453 donor retina belonging to normal and AMD patients that were used to train and test different ML models for AMD classification. We investigated three feature selection methods and tested seven machine learning models to find the best model. To train the models, the data was randomly split into 70% training data and 30% test data. The training data was then further split for a 10-fold cross validation repeated 5 times. Sensitivity, specificity, accuracy, and Area under the receiver operating characteristics (ROC) curve were used to compare the performance of the machine learning models. We used a combination of biclustering plots, pathways analysis and co-expression regulation networks to identify the candidate genes relevant to AMD biology. Most classifiers achieved 60-80% accuracy in differentiating the advanced AMD from controls based on 81 genes/features. These features were enriched for genes in immune response, complement and extracellular matrix and connected to known AMD genes through co-expression networks and gene expression correlation. Our work demonstrates the merits of ML algorithms for disease classification and suggests the key role of gene expression changes in AMD despite a small patient population. Gene regulatory networks are sufficiently interconnected with individual genes having a small impact on the disease outcome. Thus, our method provides an opportunity to regain the holistic view of the AMD that is lost in experimentally tested reductionist approaches.

Complex Traits Posters - Wednesday

PB1528. Mapping Autism Spectrum Disorder Endophenotypes to Genomic Regions in Thousands of Families

Authors:

N. Stockham, K. Paskov, B. Chrisman, J-Y. Jung, D. P. Wall; Stanford Univ., Stanford, CA

Abstract Body:

The genetic architecture of Autism Spectrum Disorder (ASD) is multifaceted and unresolved despite consistently high heritability estimates. Recent studies with extremely large sample sizes have associated a few common variants and many high-effect ultra-rare variants with ASD diagnosis status but leave the genetic etiology of the vast majority of ASD diagnoses unexplained. A possible explanation for this lack of discovery is that there are several etiologically distinct forms of ASD and grouping these distinct forms under one diagnosis status obscures the genomic elements relevant to each etiology; therefore, ASD must be studied at the level of detailed psychiatric endophenotype while integrating information across the entire genome. In this work we attempt to map to specific genomic regions the psychiatric endophenotypes described by the subdomains (Reciprocal Social Interaction (RSI), Restricted Repetitive and Stereotyped Behavior (RRSB), Communication (COM)) of the Autism Diagnostic Interview-Revised (ADIR) and the closely related Social Communication Questionnaire (SCQ). Due to recombination siblings share varying fractions of chromosomal material that are Identical-By-Descent (IBD) from each parent. We hypothesized that increased sibling IBD fraction would entail decreased sibling differences in the molecular mechanisms relevant to ASD, and therefore decrease the sibling differences in ASD phenotypes. To determine which chromosomes are most relevant to different ASD endophenotypes, we applied a modified Haseman-Elston regression to seventeen hundred sibling pairs with concordant ASD diagnosis status from the two largest family-based ASD cohorts: the Autism Genetic Research Exchange/Hartwell (AGRE/iHART) collection and Simons Foundation SPARK (SPARK). The RSI endophenotype regression yielded the strongest genetic signal in both SPARK and AGRE/iHART cohorts, with chromosome 15 passing conservative permutation tests in SPARK and AGRE/iHART. In contrast, neither the RRSB nor COM subdomains could be mapped to a chromosome in both cohorts. The highlighting of chromosome 15 is particularly interesting given that chromosome 15q11-13 is the known causal loci for Angelman Syndrome and Prader-Willi Syndrome; genetic imprinting disorders that share many social behavior phenotypes with ASD. By focusing on detailed phenotypic and genomic differences between sibling pairs, we shift focus from searching for causal genetic variants in unrelated ASD cases to discovering the molecular mechanisms most relevant to ASD phenotypes.

Complex Traits Posters - Wednesday

PB1529. Metabolomic signatures of body mass index in the Hispanic Community Health Study/ Study of Latinos (HCHS/SOL)

Authors:

V. Buchanan¹, H. Highland², B. Yu³, C. L. Avery¹, S. Buyske⁴, J. Cai⁵, M. Daviglus⁶, A. Howard⁵, C. R. Isasi⁷, R. Kaplan⁷, R. J. F. Loos⁸, Q. Qi⁷, R. Rohde⁵, J. I. Rotter⁹, L. Van Horn¹⁰, E. Boerwinkle³, K. E. North¹¹, K. Young¹¹; ¹Univ. of North Carolina - Chapel Hill, Chapel Hill, NC, ²Univ North Carolina at Chapel Hill, Chapel Hill, NC, ³Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ⁴Rutgers Univ, Piscataway, NJ, ⁵Univ. of North Carolina, Chapel Hill, NC, ⁶Univ. of Illinois Coll. of Med., Chicago, IL, ⁷Albert Einstein Coll. of Med., Bronx, NY, ⁸Univ. of Copenhagen, Copenhagen, Denmark, ⁹Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, ¹⁰Northwestern Univ., Chicago, IL, ¹¹Univ North Carolina, Chapel Hill, NC

Abstract Body:

The public health burden of obesity is substantial, but the underlying etiology and metabolic dysfunction remains unclear, especially in underrepresented Hispanic/Latino populations who may be disproportionately affected. To address this gap, we aim to identify metabolites associated with body mass index (BMI) in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), a large, community-based prospective cohort from randomly selected households at 4 US study centers. Exams, interviews, and fasting blood draws were conducted at in-person baseline visits (2008-2011). A total of 3972 genotyped participants were randomly selected for metabolomic profiling using an untargeted platform. We investigated cross-sectional associations between Blom-transformed metabolites and BMI, adjusting for age, study center, background group, smoking status, and principal components of ancestry, stratified by sex (SUGEN), and meta-analyzed (SAS 9.4) across strata. We then conducted pathway analyses (MetaboAnalyst 5.0) for the significantly associated metabolites. Data for 687 known metabolites and BMI were available for 3667 individuals (mean age: 45.6 years, 57% female). We identified Bonferroni-corrected significant associations ($p < 0.05/687$) for 256 metabolites in the meta-analysis, 147 of which were positively and 109 negatively associated with BMI; to our knowledge, 39 (15%) were novel. Among the most strongly associated metabolites were: tyrosine ($\beta = 2.104$, $SE = 0.093$, $p < 1E-16$), found in high protein foods and previously associated with obesity and insulin resistance; glucose ($\beta = 1.209$, $SE = 0.100$, $p < 1E-16$), associated with obesity and diabetes; and histidine ($\beta = -1.119$, $SE = 0.099$, $p < 1E-16$), associated with decreased insulin resistance and BMI and may suppress inflammatory cytokines in adipocytes. The caffeine metabolism pathway was significantly overrepresented (FDR-corrected $p = 1.82E-2$). Positive and negative associations of caffeine and obesity traits have been shown; caffeine was positively associated with BMI in our stratified and meta-analyzed ($\beta = 0.885$, $SE = 0.101$, $p < 1E-16$) results.

Caffeine consumption may occur via sugar sweetened beverages and weight loss supplements. Using data from multiple visits and analyses that assess for metabolites that mediate the genetic effects of BMI and obesity may help disentangle the complex relationship between obesity and caffeine. Results are expected to facilitate understanding of metabolic dysregulation associated with excess adiposity in diverse, highly admixed populations and may provide novel insights concerning metabolic and physiologic mechanisms underlying obesity.

Complex Traits Posters - Thursday

PB1530. Metabolomics profiling upranks the role of natural steroids and phenylalanine metabolism over oxidative stress in insulin resistance amongst lean subjects.

Authors:

I. Diboun¹, L. Al-Mansoori², O. Albagha³, K. Fakhro⁴, Y. Mokrab¹, M. A. Elrayess²; ¹Sidra Med., Doha, Qatar, ²Qatar Univ., Doha, Qatar, ³Hamad Bin Khalifa Univ., Doha, Qatar, ⁴Sidra Med. and Res. Ctr., Doha, Qatar, Qatar

Abstract Body:

Insulin resistance (IR) is a prediabetic condition characterized by a loss of normal Insulin Sensitivity (IS) whereby an elevated insulin level is required to trigger glucose absorption from the blood. IR is strongly associated with obesity, but it is also occasionally observed amongst lean individuals. This study aims to characterize the metabolic correlates of IR in healthy lean to shed light on the underlying mechanisms. In this cross-sectional study, clinical and metabolic data were obtained for 200 lean healthy females (100 IR and 100 IS) from Qatar Biobank and another set of 107 obese subjects (41 IR and 20 IS) from local hospitals in Qatar. Linear models using metabolite levels comparing IR versus IS revealed upregulation of anabolic steroids, for both lean and obese including Androsterone Glucuronide and Epiandrosterone in addition to derivatives of Phenylalanine including the microbiota-product 1-Carboxyethylphenylalanine. Conversely, interaction analysis looking at differential IR/IS effects between obese and lean revealed that levels of long chain unsaturated fatty acids and markers of oxidative stress, including 2-Hydroxybutyrate, were uniquely elevated amongst obese IR but not lean IR, in comparison to their respective baseline levels in IS. These results were confirmed at a more global level using multivariate analysis whereby an Orthogonal Partial Least Square classifier trained on lean IR/IS was able to only marginally, though significantly, distinguish IR/IS from the obese cohort (Area Under the Curve (AUC) from Receiver Operating Characteristic Curve (ROC) analysis=0.68, pvalue=0.02). Elastic-Net-Regularized Generalized Linear Model was used to select a subset of 20 best predictor metabolites of IR in the lean and ROC analysis suggested a high discriminant capacity (AUC=0.93), in comparison to a similar classifier trained entirely on phenotypic traits (AUC=0.73). Amongst the predictor set of metabolites selected were Phenylalanine derivatives, steroids, xenobiotics (PFOS and Piperidinone), Glucose and metabolic traces of Lysine metabolism. Our data suggest an interplay of potentially genetic and environmental factors underlying IR in the lean population as evident in the enrichment in natural steroids but also food-derived amino acids including lysine and phenylalanine as well as xenobiotics. Importantly, our study suggests a different profile of risk factors to IR in the lean in comparison to obese subjects where IR is thought to be primarily associated with oxidative stress. Our subset of predictive IR metabolites in lean presumably healthy individuals may hold clinical value, hence its validation is well warranted.

Complex Traits Posters - Wednesday

PB1531. Metagenomic analysis of the gut resistome in dyslipidemia and healthy individuals.

Authors:

S. Raza¹, H-L. Kim², H-N. Kim^{3,4}; ¹Med. Res. Inst., Kangbuk Samsung Hosp., Sch. of Med., Sungkyunkwan Univ., S, Seoul, Korea, Republic of, ²Ewha Womans Univ, Seoul, Korea, Republic of, ³Med. Res. Inst., Kangbuk Samsung Hosp., Sch. of Med., Sungkyunkwan Univ., South Korea, Seoul, Korea, Republic of, ⁴Dept. of Clinical Res. Design and Evaluation, SAIHST, Sungkyunkwan Univ., South Korea, Seoul, Korea, Republic of

Abstract Body:

Dyslipidemia is defined as a disruption in lipid metabolism that affects the concentration of lipids in the blood, is a modifiable risk factor for cardiovascular disease. The intestinal microbiota has been reported to play a regulatory role in host lipid metabolism. The human gut microbiome harbors various antibacterial resistant genes (ARG). However, the role of gut resistome in the dyslipidemia patient is still unclear. Here we analyze the antibiotic resistance genes of gut microbiota from 1,026 patients with dyslipidemia and 618 healthy subjects. We used metagenomic approach to address differentially abundant species and the variation in ARGs profiles. Total ARG abundance was significantly higher in dyslipidemia patients than in healthy subjects (Wilcoxon rank-sum test; $p < 0.05$). Overall, tetracycline, multidrug-resistance, and beta-lactamase were the most abundant ARG classes in all subjects. However, aminoglycoside, phenicol, and fluoroquinolones were significantly more abundant in dyslipidemia than in healthy control ($p < 0.05$). The abundance of *tetQ* was the only ARG type significantly higher in dyslipidemia, while *tetW*, *tet40*, and *CblA-1* in healthy subjects were higher than in dyslipidemia. Similarly, *Prevotella* was differentially more abundant in dyslipidemia patients, and *Prevotella copri* and *Bacteroides plebeius* were significantly high in dyslipidemia patients. The *Prevotella* was negatively correlated with most ARGs except the *tetQ*. Similarly, *tetQ* was positively correlated with *Prevotella* genera and negatively correlated with other genera. Overall, more than eight bacterial genera were positively correlated with most ARGs in dyslipidemia, while only four genera were in healthy subjects. In dyslipidemia patients, the level of total cholesterol and low-density lipoproteins (LDL) cholesterol were negatively correlated with total ARGs, while aminoglycoside was positively correlated with LDL cholesterol. In conclusion, the taxonomic and ARGs of gut bacteria in dyslipidemia patients were significantly altered compared with healthy subjects. Further studies are required to investigate the role of *Prevotella* and tetracycline-resistant genes in dyslipidemia patients. This research was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2020R1A2C1012931).

Complex Traits Posters - Thursday

PB1532. Mitochondrial and Nuclear genetic variants demonstrate mitochondrial function determines severity but not risk of amyotrophic lateral sclerosis

Authors:

C. Harvey¹, M. Weinreich², S. Zhang³, P. Hop⁴, R. Zwamborn⁴, K. van Eijk⁴, T. Julian¹, T. Moll⁵, A. Iacoangeli⁶, A. Al Khleifat⁶, J. Quinn⁷, A. Pfaff⁸, S. Koks⁸, J. Poulton⁹, S. Battle¹⁰, D. Arking¹¹, M. Snyder¹², Project MinE ALS Sequencing Consortium, J. Veldink⁴, K. Kenna⁴, P. Shaw¹³, J. Cooper-Knock¹; ¹Univ. of Sheffield, Sheffield, United Kingdom, ²Univ. of Sheffield, German Cancer Research Center and University Hospi, Germany, ³Stanford Univ., Palo Alto, CA, ⁴Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, ⁵Univeristy of Sheffield, Sheffield, United Kingdom, ⁶King's Coll. London, London, United Kingdom, ⁷Dept. of Pharmacology and Therapeutics, Inst. of Systems, Molecular & Integrative Biology, Liverpool, United Kingdom, ⁸Perron Inst. for Neurological and Translational Sci., Perth, Australia, ⁹Univ. of Oxford, Oxford, United Kingdom, ¹⁰Johns Hopkins Univ., Baltimore, MD, ¹¹Johns Hopkins Univ Sch. of Med., Baltimore, MD, ¹²Stanford Univ., Stanford, CA, ¹³Univeristy of Sheffield, Sheffield, United Kingdom

Abstract Body:

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease. Selective vulnerability of energy-intensive motor neurons (MNs) has fostered speculation that mitochondrial function is a determinant of ALS. Previously, the position of mitochondrial function in the pathogenic cascade leading to neurotoxicity has been unclear. We separated upstream genetic determinants of mitochondrial function, including genetic variation within the mitochondrial genome or autosomes; from downstream changeable factors including mitochondrial copy number (mtCN) and MN gene expression. We discovered that functionally validated mitochondrial haplotypes are a determinant of ALS survival but not ALS risk. Loss-of-function genetic variants within, and reduced MN expression of, ACADM and DNA2 lead to shorter ALS survival; both genes impact mitochondrial function. MtCN responds dynamically to the onset of ALS independent of mitochondrial haplotype, and is also significantly correlated with disease severity. We conclude that mitochondrial function impacts ALS progression but not risk; our findings have therapeutic implications.

Complex Traits Posters - Wednesday

PB1533. Mitochondrial *TXNRD2* and *ME3* genetic risk scores are associated with specific primary open-angle glaucoma phenotypes

Authors:

I. Aboobakar¹, T. Kinzy², Y. Zhao¹, B. Fan³, L. Pasquale⁴, J. Cooke Bailey⁵, J. Wiggs⁶, NEIGHBORHOOD Consortium; ¹Mass Eye and Ear, Boston, MA, ²Case Western Reserve Univ., Cleveland, OH, ³Harvard Med Sch, Massachusetts Eye & Ear Infirmary, Boston, MA, ⁴Mount Sinai Hlth. System, New York, NY, ⁵Case Western Reserve Univ, Cleveland, OH, ⁶Harvard Med Sch, MEEI, Boston, MA

Abstract Body:

Purpose: Genetic variants in regions that include the mitochondrial genes *TXNRD2* and *ME3* are associated with primary open-angle glaucoma (POAG) in genome-wide association studies (GWAS). These genes may functionally contribute to POAG development by altering NADPH levels and oxidative stress response. To assess their clinical impact, we investigated whether *TXNRD2* and *ME3* genetic risk scores (GRSs) are associated with specific glaucoma phenotypes. **Methods:** 2617 POAG cases and 2634 controls from the NEIGHBORHOOD consortium were studied. All POAG-associated single nucleotide polymorphisms (SNPs) in the *TXNRD2* and *ME3* loci were identified using GWAS data ($p < 0.05$). Of these, 20 *TXNRD2* and 24 *ME3* SNPs were selected after adjusting for linkage disequilibrium. The correlation between SNP effect size and gene expression levels was investigated using the Gene-Tissue Expression (GTEx) database. GRSs were constructed for each individual using the unweighted sum of *TXNRD2*, *ME3*, and *TXNRD2+ME3* combined risk alleles. Age and gender-adjusted odds ratios (ORs) for POAG diagnosis were calculated per decile for each GRS. Additionally, the clinical features of POAG cases in the top 1, 5, and 10% of each GRS were compared to the bottom 1, 5, and 10%, respectively. **Main Outcome Measures:** POAG OR per GRS decile; maximal intraocular pressure (IOP) and prevalence of paracentral visual field loss among POAG cases with high vs. low GRSs. **Results:** Increased SNP effect size strongly correlated with higher *TXNRD2* and lower *ME3* expression levels ($r = 0.72$ and -0.97 , respectively). Individuals in decile 10 of *TXNRD2+ME3* GRS had the highest odds of POAG diagnosis (OR=1.79, 95% CI 1.39-2.30, $p < 0.001$). POAG cases in the top 1% of *TXNRD2* GRS had higher mean maximal IOP compared to the bottom 1% (19.9 mmHg vs 15.6 mmHg, $p = 0.01$). POAG cases in the top 1% of *ME3* and *TXNRD2+ME3* GRS had higher prevalence of paracentral field loss compared to the bottom 1% (72.7-89% vs 14.3-33%, $p = 0.02$ and 0.01 , respectively). **Conclusions:** Individuals with higher GRSs have higher odds of POAG diagnosis, with a greater effect seen for *TXNRD2+ME3* combined compared to either gene alone. Moreover, POAG patients with higher GRSs have higher IOP and greater prevalence of paracentral field loss. These data suggest that *TXNRD2* and *ME3* functionally contribute to POAG development and disease severity, potentially through a mechanism involving NADPH levels and mitochondrial dysfunction.

Complex Traits Posters - Thursday

PB1534. Modeling cardiac cell developmental trajectories at high temporal resolution.

Authors:

E. McIntire, K. Rhodes, K. Barr, M. DeMille, B. Umans, J. Burnett, N. Gonzales, Y. Gilad; The Univ. of Chicago, Chicago, IL

Abstract Body:

Cardiovascular disease (CVD) is the leading cause of mortality worldwide and has the highest economic impact of all physical noncommunicable diseases. Genetic variants modulate susceptibility in nearly all types of CVD. Most variants are noncoding, suggesting that gene regulation plays an integral role in mediating disease outcomes. However, pathogenic regulatory events are highly context dependent and may only occur during specific developmental stages, in specific cell types, or when certain environmental conditions are met. The specificity of gene regulation has made its characterization challenging, especially in cell types exclusive to early development. Induced pluripotent stem cells (iPSCs), which can be differentiated into cardiomyocytes and other disease-relevant cell types, have provided valuable insights into developmental gene regulation and its impact on CVD. For example, a timecourse study of cardiomyocyte differentiation discovered that even fleeting gene regulatory effects that occur at intermediate developmental stages can be associated with CVD-related outcomes in adults. However, these timecourse experiments are limited to examining a discrete number of collection time points, which are only snapshots of a highly dynamic process. Here, we have developed a new iPSC-derived cardiac organoid model that allows us to study gene regulation at unprecedented temporal resolution. We aggregated iPSCs to form three-dimensional embryoid bodies (EBs) that are guided to a cardiac fate over 9 days. Using scRNA-seq data from 7,286 cardiac EB cells collected from three individuals, we used unsupervised clustering and marker gene expression to uncover 11 known cell types (cardiac progenitor cells, proliferating cells, cardiomyocytes, fibroblasts, myofibroblasts, primitive endoderm, foregut, epicardium, endocardium, ectoderm, and neural crest). Our cell type annotations are consistent with those obtained using external scRNA-seq reference data collected from primary fetal tissue, indicating that cardiac EBs have transcriptional profiles resembling those observed in human fetal cells. Finally, we combined our cardiac EB data with single-cell data collected in a 16-day timecourse study of cardiomyocyte differentiation. All cell types observed in the extended timecourse study are present in a single collection of cardiac EBs, from pluripotent cells to differentiated cardiomyocytes. In other words, cardiac EBs effectively condense a differentiation timecourse experiment into a single culture system.

Complex Traits Posters - Thursday

PB1535. Modeling gene by environment interactions in post-traumatic stress disorder using hiPSC-derived neurons

Authors:

C. Seah¹, T. Rusielewicz², H. Bader³, C. Xu³, H. Young¹, R. Signer¹, A. dePins¹, P. Deans⁴, M. Breen¹, D. Paull², K. Brennand⁴, L. Huckins¹, R. Yehuda³; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²New York Stem Cell Fndn., New York, NY, ³James J. Peters Dept. of VA Med. Ctr., New York, NY, ⁴Yale Univ., New Haven, CT

Abstract Body:

Post-traumatic stress disorder (PTSD) is a debilitating disorder that is underdiagnosed and under-treated. Better identification of genetic and environmental elements of PTSD susceptibility and resilience is important to mitigate disorder burden. Integrating genomic loci with traumatic exposures may further elucidate gene by environment interactions that influence PTSD susceptibility. To test this, we performed an expression quantitative trait loci (eQTLs) analysis to identify SNPs that alter nearby mRNA expression across two PTSD-relevant cell types treated with the synthetic glucocorticoid hydrocortisone (HCort). eQTLs associate SNPs with nearby expression changes in a context-dependent manner and thus capture genetically- and environmentally-regulated expression. Our approach aimed to reveal cell type-specific and stress-dynamic eQTLs that may offer insight into the combined genetically- and stress-regulated contributions to PTSD. Fibroblasts from combat-exposed veterans with PTSD (n=20) and without (n=20), were used to generate two types of human induced pluripotent stem cell (hiPSC)-derived neurons- glutamatergic and GABAergic, and treated with two doses of HCort. Transcriptomic analysis of PTSD-specific HCort-response was conducted, then we identified eQTLs with interactive effects of genetically and HCort-regulated expression across cell types. We performed gene set enrichment of interactive eGenes for GWAS Catalog reported traits. We identified 675 and 374 unique eGenes in GABA-ergic and glutamatergic neurons, respectively, with a SNP by HCort interaction. These genes enrich in GWAS catalog genes including generalized epilepsy (p=9.65E-08), bipolar disorder (p=9.36E-06), and body mass index (p=1.48E-07), traits associated with stress and glucocorticoid response. Genes with particularly strong interactions include *PNPLA2* (p=5.6E-04), *LRP8* (p=6.59E-04), and *MXD4* (p=4.16E-04), which have been associated with chronic defeat stress, learning/memory formation, and amygdala signatures in a PTSD-like mouse model, respectively. eGenes in GABAergic and glutamatergic neurons differentially enrich in inflammatory, metabolic, and neuropsychiatric traits, indicating their differential role in response to acute stressors. Our preliminary data suggests that integration of genetics and stress signatures identifies genes associated with genetically regulated stress response in two major cell types implicated in psychiatric disorders. Further fine-mapping of eQTLs to disorder-relevant GWAS and validation in human cohorts may elucidate disorder relevance of these signatures.

Complex Traits Posters - Wednesday

PB1536. Modification of coronary artery disease clinical risk factors by coronary artery disease polygenic risk score

Authors:

B. Truong^{1,2}, Y. Ruan^{1,2}, S. Haidermota^{1,2}, A. Patel^{2,2}, I. Surakka¹, W. Hornsby^{1,2}, S. Lee^{3,4}, P. Natarajan^{1,2}; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²Massachusetts Gen. Hosp., Boston, MA, ³Australian Ctr. for Precision Hlth., Univ. of South Australia Cancer Res. Inst., Univ. of South Australia, Adelaide, Australia, ⁴UniSA Allied Hlth.and Human Performance, Univ. of South Australia, Adelaide, Australia

Abstract Body:

Importance Clinical risk factors and coronary artery disease (CAD) polygenic risk score (PRS) complementarily predict risk for CAD. The extent to which the relationships between clinical risk factors and CAD are altered by CAD PRS is not well understood. **Objective** Determine whether the interactions between clinical risk factors and CAD PRS further explain risk for incident CAD. **Design, setting and participants** Participants were of European ancestry from the UK Biobank with incident CAD. Externally-trained genome-wide CAD PRS was generated and then applied to participants. Clinical risk factors, including type 2 diabetes (T2D), family history of heart disease, ever and current smoker, systolic and diastolic blood pressure (SDP and DBP), body mass index (BMI), lipoprotein(a), total cholesterol, low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C) and C-reactive protein (CRP), were ascertained at baseline. **Main Outcomes and Measures** Cox proportional hazards models were fitted to examine the CAD effects of CAD PRS, risk factors and their statistical interactions. Next, the PRS and risk factors were stratified to investigate the attributable risk of clinical risk factors. **Results** A total of 357,144 individuals (196,643 [55.1%] female, mean [standard deviation (SD)] age 57.2 [7.9] years) of European ancestry without prevalent CAD were included. During a median of 11.1 years of follow-up (interquartile range, 10.4-14.1 years), 21,971 individuals developed incident CAD (6.2%). CAD PRS was associated with 1.35-fold (95% CI 1.332-1.368) risk per SD for incident CAD. The prognostic relevance of these risk factors is relatively diminished for those with high CAD PRS: type 2 diabetes ($HR_{interaction} 0.91$, 95% $CI_{interaction} 0.88-0.94$, $P_{interaction} 1.14 \times 10^{-7}$), body mass index ($HR_{interaction} 0.97$, 95% $CI_{interaction} 0.96-0.98$, $P_{interaction} 2.91 \times 10^{-5}$) and C-reactive protein ($HR_{interaction} 0.98$, 95% $CI_{interaction} 0.96-0.99$, $P_{interaction} 9.35 \times 10^{-4}$). However, high CAD PRS yielded joint risk increases with low-density lipoprotein cholesterol ($HR_{interaction} 1.05$, 95% $CI_{interaction} 1.04-1.06$, $P_{interaction} 4.04 \times 10^{-13}$) and total cholesterol ($HR_{interaction} 1.05$, 95% $CI_{interaction} 1.03-1.06$, $P_{interaction} 4.38 \times 10^{-12}$). **Conclusion** CAD PRS is associated with incident CAD, and its application improves the interpretation of the prognostic relevance of several clinical risk factors.

Complex Traits Posters - Thursday

PB1537. Molecular and clinical characterisation of Polish Temple syndrome patients with 14q32 alterations

Authors:

D. Jurkiewicz¹, A. Madej-Pilarczyk¹, A. Swiader-Lesniak², K. Fraczak¹, P. Halat-Wolska¹, D. Siestrzykowska¹, D. Piekutowska-Abramczuk¹, M. Pelc¹, B. Chalupczynska¹, E. Ciara¹, K. Chrzanowska¹; ¹Dept. of Med. Genetics, The Children's Mem. Hlth.Inst., Warsaw, Poland, ²Dept. of Anthropology, The Children's Mem. Hlth.Inst., Warsaw, Poland

Abstract Body:

Temple syndrome (TS14) is an imprinting disorder characterised by pre- and postnatal short stature with small hands and feet, muscular hypotonia and feeding problems in early infancy followed by weight gain, premature puberty and in some cases speech delay and mild mental retardation. TS14 is caused by genetic/epigenetic abnormalities within the imprinted chromosomal region 14q32. It may show overlapping phenotype with Silver-Russell syndrome (SRS).

Here, we present clinical and molecular data of three patients with TS14. Molecular investigations were performed on leukocyte DNA and included methylation sensitive multiplex ligation-dependent probe amplification (MS-MLPA) and microsatellite analyses (MSA). In two patients maternal UPD of chromosome 14 [upd(14)mat] was found and in the third patient *MEG3*:TSS-DMR loss of methylation (LOM) was identified. Analysis of several imprinted *loci* in the patient with LOM excluded the presence of Multilocus Imprinting Disturbances (MLID). Patients demonstrated typical clinical TS14 features, however in one patient with upd(14)mat mild intellectual disability was present.

The investigations allowed to define the type of a 14q32 defect and confirm the diagnosis of TS14 in examined patients. A complex molecular approach to analyse 14q32 region is required for accurate diagnosis of TS14 allowing the recurrence risk determination and proper genetic counselling. The study corroborates that SRS should be considered in differential diagnosis with TS14 and shows efficacy of the applied diagnostic approach.

The study was supported by CMHI project S180/2019 and partly by MEiN projects: 7071/IB/SN/2020, 7088/II-KDM/SN/2020.

Complex Traits Posters - Thursday

PB1539. Molecular epidemiologic and family history evidence suggests that myalgic encephalomyelitis (ME)/chronic fatigue syndrome (CFS) may be an autoimmune disorder.

Authors:

R. Moslehi¹, A. Kumar¹, A. Dzutsev²; ¹Univ. at Albany, Albany, NY, ²NCI, NIH, Bethesda, MD

Abstract Body:

Background: Myalgic encephalomyelitis (ME)/chronic fatigue syndrome (CFS) is a complex disabling disorder with no known etiology or approved treatment. Estimates of the prevalence suggest that up to 3.4 million Americans may be afflicted. It has been suggested that ME/CFS may be triggered by an infectious illness including COVID-19, hence the prediction that 10 million new cases of ME/CFS may be diagnosed globally by the end of this pandemic. It is imperative to identify risk factors and underlying biologic mechanisms for ME/CFS. To this end, we conducted a molecular epidemiologic study to explore the link between ME/CFS, autoimmune disease (AID), and cancer. Methods: Our clinic-based case-control study involved 59 carefully selected ME/CFS patients and 54 appropriately matched healthy controls. We compared cases and controls with respect to the following: 1. prevalence of AID and cancer among their first-degree relatives, 2. prevalence of epidemiologic factors, and 3. serum levels of 48 cytokines. Statistical methods used were logistic regression and cumulative incidence analysis to calculate odds ratios (OR), relative risks (RR), 95% confidence intervals (CI) and p-values. We also used machine learning approaches to study the predictive power of serum cytokine levels in ME/CFS. Results: Our analysis revealed that ME/CFS cases were five times more likely than controls to have a family history of AID (OR=5.30, p=0.002). The life-time risk of AID among first-degree relatives of cases was significantly higher compared to the relatives of controls (RR=2.68, p=0.02). First-degree relatives of cases also had a significantly higher risk of early-onset (diagnosed <60 years of age) cancer compared to the relatives of controls (RR=2.24, p=0.03). Comparison of epidemiologic factors identified certain risk factors for ME/CFS such as history of allergies requiring medication (OR=6.00, 95%CI:2.52-14.28, p<0.0001) and exposure to contaminants (OR=4.35, 95%CI:1.96-9.65, p=0.0002). Our analysis identified a cytokine profile of ME/CFS, which classified patients with 84% accuracy (kappa=0.68, p=0.025, sensitivity=0.75, and specificity=1.00) in random forest models. Two cytokines, Interleukin-27 (IL27) and Macrophage inflammatory protein-1 alpha (MIP1- α), were found to be the most predictive biomarkers for ME/CFS. Both are involved in inflammatory processes and MIP1- α has been linked to other AIDs. Conclusions: Findings from our multidimensional analysis of pedigree, epidemiologic, and molecular data provide the most objective evidence to date that ME/CFS may be an AID. Our findings provide etiologic clues and druggable targets for treatment of ME/CFS.

Complex Traits Posters - Wednesday

PB1540. Molecular states during acute COVID-19 reveal distinct etiologies of long-term sequelae.

Authors:

R. Thompson¹, N. W. Simons¹, L. Wilkins¹, E. Cheng¹, D. M. Del Valle¹, G. E. Hoffman¹, C. Cervia², D. Yuan³, The Mount Sinai COVID-19 Biobank Team, J. R. Heath³, O. Boyman², S. Kim-schulze¹, R. Sebra¹, M. Merad¹, S. Gnjatic¹, E. E. Schadt¹, A. W. Charney¹, N. Beckmann¹; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²Univ. of Zurich, Zurich, Switzerland, ³Inst. for Systems Biology, Seattle, WA

Abstract Body:

Post-acute sequelae of SARS-CoV-2 (PASC) infection are debilitating, clinically heterogeneous, and of unknown molecular etiology. While the development of PASC is presumed to be rooted in the acute phase of COVID-19, direct causal links have yet to be established. A transcriptome-wide investigation was performed in a diverse cohort of 165 acutely infected hospitalized patients followed clinically into the post-acute period. We developed a method for modeling cell-type-specific gene expression patterns in bulk RNA-seq data using an interaction term between the phenotype of interest and the estimated cell type fractions. We validated our model by showing significant overlap between cell-type-specific differentially expressed genes for each cell type and published marker genes for that cell type; as well as by applying the model to an independent single-cell RNA-seq data set and showing significant positive correlations between the log fold changes from our model and those obtained by analyzing each cell type separately. We then used this model to identify distinct gene expression signatures of PASC in whole blood during acute infection, with innate and adaptive immune cell types implicated in different sequelae. Two clusters of sequelae exhibited divergent plasma-cell-specific gene-expression patterns. Sleep problems, nausea/diarrhea/vomiting, smell/taste problems, and skin rash were all associated with higher expression of immunoglobulin-related genes, while lung problems and pneumonia were both associated with lower expression of such genes. Furthermore, these expression patterns were dependent on anti-spike antibody titers for 3 of the sequelae (lung problems, pneumonia, and skin rash), directly linking them to the host response to SARS-CoV-2 infection. The downregulation of immunoglobulin genes associated with the other 3 symptoms independently of anti-spike antibody titers was validated in an independent cohort by the observation of lower total immunoglobulin titers in subjects with PASC. Incorporation of molecular data from acute phase enabled the detection of multiple etiologies of PASC beyond what can be discerned using only clinical data, emphasizing that acute phase molecular data is essential for a complete characterization of PASC as well as introducing the possibility for discovery of predictive biomarkers and development of precision treatment and prevention strategies for specific sequelae. Altogether, multiple etiologies of PASC were already detectable during SARS-CoV-2 infection, directly linking PASC with the acute host response to the virus and providing early insights into their development.

Complex Traits Posters - Thursday

PB1542*. Multi-ancestry HLA allele calling using whole-exome sequencing in the UK Biobank reveals 353 novel genome-wide significant HLA alleles associated with 11 auto-immune phenotypes

Authors:

G. Butler-Laporte; McGill Univ., Montréal, QC, Canada

Abstract Body:

The HLA is a highly polymorphic region associated with many immune and inflammatory diseases. SNP-based association studies cannot capture all its genetic variation, making them underpowered at this locus. HLA gene allele imputation can alleviate this problem. However, this can still be imprecise, especially in individuals of ancestries less represented in the imputation panel. Hence, HLA associations might still be missed.

Here, we use the UK Biobank 450,000 whole-exome sequences (WES) to directly call HLA alleles at 19 protein-coding genes and 11 pseudogenes using the HLA-HD software and the IMGT/HLA database (v3.45.0). When comparing previously imputed HLA alleles available in the UK Biobank to our results, imputed class I HLA genes achieved a biallelic 2-digit accuracy match of 82-83%. For class II genes, match rate decreased to 75% at *DPBI*, to as low as 0.2% at *DRB5*, suggesting that HLA sequencing improved HLA allele accuracy considerably.

We next performed HLA association studies of 11 known immune-mediated phenotypes: ankylosing spondylitis, asthma, immune thyroid disorders, coeliac disease, Crohn's, multiple sclerosis, polymyalgia rheumatica, psoriasis, rheumatoid arthritis, type I diabetes, and ulcerative colitis. Firth regression was performed for each allele-phenotype pair using REGENIE, first separately by continental ancestry followed by meta-analysis. At 4-digit HLA accuracy, we found 405 significant associations ($p < 5 \times 10^{-8}/11$), of which 353 were novel. These included 368 alleles in protein-coding genes (316 novel) and 37 in pseudogenes.

We then evaluated the predictive power of HLA calling on 7 of these traits to construct polygenic risk scores (PRSs) by combining LDpred results with imputed or called HLA alleles using XGboost. While adding HLA alleles to the PRS increased accuracy (AUC increased up to 0.157 in coeliac disease), both imputed and called alleles PRSs performed well. However, of the 250 pairs of 6-digit HLA alleles comparisons between equivalent 4-digit alleles, there were 34 with significantly different effect sizes ($p < 0.05/250$), including 20 in opposing effect direction. For example, *DRA*01:01:01* was associated with a 0.88-fold decrease in the odds of psoriasis (95% CI: 0.85-0.91, $p = 4.5 \times 10^{-13}$), but *DRA*01:01:02* increased the odds 2.43-fold (95% CI: 2.30-2.57, $p = 7.3 \times 10^{-206}$).

To conclude, WES HLA allele calling increases power of HLA association studies, especially at higher allele digit accuracy, where non-coding variants play a role in phenotype. This resource will be made available to all researchers through the UK Biobank return of results program.

Complex Traits Posters - Wednesday

PB1543. Multi-ancestry meta-analysis of asthma improves polygenic risk prediction across populations

Authors:

K. Tsuo^{1,2,3}, W. Zhou^{2,3}, Y. Wang^{2,3}, M. Kanai^{3,2,1,4}, S. Namba⁴, R. Gupta^{1,2,3}, L. Majara⁵, L. Nkambule^{2,3}, T. Morisaki⁶, Y. Okada^{4,7}, B. Neale^{2,3}, Global Biobank Meta-analysis Initiative, M. Daly^{2,3,8}, A. Martin^{2,3}; ¹Harvard Med. Sch., Boston, MA, ²Massachusetts Gen. Hosp., Boston, MA, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Osaka Univ. Graduate Sch. of Med., Suita, Japan, ⁵Univ. of Cape Town, Cape Town, South Africa, ⁶The Univ. of Tokyo, Inst Med Sci, Minato-ku, Tokyo, Japan, ⁷RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, ⁸Univ. of Helsinki, Helsinki, Finland

Abstract Body:

Asthma poses a significant global health burden as one of the most common chronic diseases worldwide. Risk prediction models of asthma can help inform prevention, intervention, and management strategies, but current models rely mostly on a limited range of clinical, environmental exposure, and family history factors and have modest clinical value. To investigate the potential utility of applying genetic risk factors for asthma risk prediction, we leveraged the largest genome-wide association study of asthma to date from the Global Biobank Meta-analysis Initiative (GBMI), which included 18 biobanks of diverse genetic ancestry groups and more than 1.8 million participants (153,763 cases, of which 28,492 were of non-European ancestry), to construct polygenic risk scores (PRS) for asthma. We applied a multi-ancestry Bayesian PRS construction method, PRS-CSx, to five different ancestry-specific meta-analyses from European (EUR), African (AFR), East Asian (EAS), Central and South Asian (CSA), and Admixed American (AMR) populations in GBMI. We evaluated the prediction accuracy of the multi-ancestry PRS, measured by variance explained on the liability scale (liability-scale R^2), in independent datasets with individuals of AFR, EUR, EAS, and CSA ancestries. In all target populations, the multi-ancestry PRS derived from GBMI had greater predictive accuracy compared to the PRS derived from a previous multi-ancestry asthma meta-analysis from the Trans-National Asthma Genetic Consortium, which included smaller sample sizes from populations of EUR and non-EUR ancestries (19,954 and 3,994 cases, respectively). In the non-EUR populations specifically, we observed an 8-fold increase in liability-scale R^2 in the AFR population, and a 4- and 3-fold increase in the CSA and EAS populations, respectively. We observed a 2-fold increase in liability-scale R^2 in the EUR cohort. To further interrogate the utility of polygenic risk prediction for different asthma subtypes, which may have both shared and distinct genetic risk factors, we utilize age of onset information available in some participating biobanks in GBMI to assess the performance of the PRS in childhood- and adult-onset asthma cohorts separately. In summary, we demonstrate that leveraging association data from larger and more diverse cohorts, together with a multi-ancestry PRS construction approach, delivers substantial improvements in PRS prediction accuracy for asthma, particularly in populations of non-European ancestries.

Complex Traits Posters - Thursday

PB1544. Multi-ancestry meta-analysis of X chromosome-wide associations for height

Authors:

S. Vedantam^{1,2}, E. Marouli^{3,4}, E. Bartell^{5,1,2}, A. Eliassen^{6,7}, J. Arias⁸, T. Winkler⁹, L. Yengo¹⁰, A. Wood¹¹, S. Berndt⁸, Y. Okada^{12,13,14,15}, J. Hirschhorn^{1,2,5}, The GIANT consortium; ¹Div. of Endocrinology, Boston Children's Hosp., Boston, MA, ²Program in Med. Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, ³William Harvey Res. Inst., Barts and The London Sch. of Med. and Dentistry, London, United Kingdom, ⁴Ctr. for Genomic Hlth., Life Sci., Queen Mary Univ. of London, London, United Kingdom, ⁵Dept. of Genetics, Harvard Med. Sch., Boston, MA, ⁶COPSAC, Herlev and Gentofte Hosp., Univ. of Copenhagen, Copenhagen, Denmark, ⁷Section for Bioinformatics, Dept. of Hlth. Technology, Technical Univ. of Denmark, Copenhagen, Denmark, ⁸Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, ⁹Dept. of Genetic Epidemiology, Univ. of Regensburg, Regensburg, Germany, ¹⁰Statistical Genomics Group, Inst. for Molecular BioSci., The Univ. of Queensland, Brisbane, Australia, ¹¹Genetics of Complex Traits, Coll. of Med. and Hlth., Exeter Univ., Exeter, United Kingdom, ¹²Lab. for Statistical Analysis, RIKEN Ctr. for Integrative Med. Sci., 1-7-22 Suehiro-cho, Yokohama, Kanagawa, Japan, ¹³Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., 2-2 Yamadaoka, Suita, Osaka, Japan, ¹⁴Lab. of Statistical Immunology, Immunology Frontier Res. Ctr. (WPI-IFReC), 2-2 Yamadaoka, Suita, Osaka, Japan, ¹⁵Integrated Frontier Res. for Med. Sci. Div., Inst. for Open and Transdisciplinary Res. Initiatives, Osaka Univ., 2-2 Yamadaoka, Suita, Osaka, Japan

Abstract Body:

The human X chromosome is ~155 million base pairs long and represents 5% of the genome harboring more than 1600 genes, but has been under-investigated in Genome-wide association studies (GWAS). Understanding the contribution of X chromosome variation to polygenic traits and diseases is limited because association results for the X are often omitted due to the analytical technical challenges arising from incomplete inactivation of X chromosome in females and hemizyosity in males. In the GIANT consortium, we aimed to identify the variants on the X chromosome that contribute to adult human height by performing an X-chromosome wide fixed effect meta-analysis. In our initial analyses, we included data from females in 97 studies that included 1.8 million variants and a total sample size of 220K individuals from 5 major population groups (European, African-American, Hispanic, East-Asian and South-Asian). All studies were corrected for relatedness using linear mixed models and markers were filtered for low imputation quality and minor allele counts. We identified 12 independent signals that are more than 500kb apart and reach genome-wide significant p-values (5×10^{-8}). These signals are in the non-PAR region of the X chromosome and located near genes such as ATRX, PGK1 and FAAH2. For comparison, meta-analysis of same studies in chromosome 7 which is similar in size to the X chromosome resulted in 25 genome-wide significant signals that were 500kb apart. In ongoing work we will dramatically expand our sample size by including additional studies and also analyze male specific X chromosome data. This work will help us understand the role of common genetic variation in polygenic traits attributable to the X chromosome and could partially explain sex-specific genetic differences in polygenic traits such as height.

Complex Traits Posters - Thursday

PB1545. Multiple HLA haplotypes and a variant altering immunogenicity of minor histocompatibility antigen epitopes encoded by *CTSH* are associated with age at type 1 diabetes diagnosis.

Authors:

D. Roshandel¹, A. Spiliopoulou^{2,3}, S. McGurnaghan³, S. B. Bull^{4,5}, P. M. McKeigue², H. M. Colhoun³, A. D. Paterson^{1,5}; ¹Genetics and Genome Biology Program, The Hosp. for Sick Children, Toronto, ON, Canada, ²Usher Inst. of Population Hlth.Sci. and Informatics, Univ. of Edinburgh, Edinburgh, United Kingdom, ³Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, ⁴Lunenfeld-Tanenbaum Res. Inst., Sinai Hlth., Toronto, ON, Canada, ⁵Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Type 1 diabetes (T1D) is typically considered a disease with onset in childhood or young adulthood, but it can present throughout the lifespan. Genome-wide association studies (GWAS) have identified 78 loci associated with T1D susceptibility. However, genetics of T1D age at diagnosis (AAD) has received little attention. Here, we performed a large meta-GWAS of T1D AAD including subjects diagnosed later in life compared to previous studies, genotyped on Illumina chips with extensive whole genome coverage and imputed to TOPMed. 7,923 unrelated European subjects with T1D from five studies were included: the Scottish Diabetes Research Network Type 1 Bioresource (SDRNT1BIO, n = 5,349, AAD mean = 23 yrs), Diabetes Control and Complications Trial (DCCT, n = 1,304, AAD mean = 21 yrs), Coronary Artery Calcification in Type 1 Diabetes (CACTI, n = 529, AAD mean = 13 yrs), Pittsburgh Epidemiology of Diabetes Complications (EDC, n = 150, AAD mean = 8 yrs) and Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR, n = 591, AAD mean = 14 yrs). Variants with minor allele frequency >0.01 and high imputation quality (INFO >0.80) were included in the meta-GWAS (n = 8,154,710). HLA imputation was performed using TOPMed multi-ethnic HLA reference panel on Michigan imputation server. AAD was square root transformed and sex was included as a covariate in the model. Meta-GWAS analysis was performed using METAL v1.5 weighting effect sizes using the inverse of the standard errors. Multiple single nucleotide polymorphisms (SNPs) in the HLA region (top SNP: Chr6:32706117 (build GRCh38), T>C, $\beta = 0.28$, $p = 2.9E-37$) and an indel on Chr 15 (Chr15:78943251, TGTT>, $\beta = 0.21$, $p = 3.8E-8$) reached the genome-wide significance threshold ($p < 5E-8$). The indel is in high linkage disequilibrium ($r^2 > 0.8$) with rs2289702 (chr15:78944951, C>T, Gly12Arg) within *CTSH*. *CTSH* encodes two minor histocompatibility antigen epitopes: HLA-A*3101 and HLA-A*3303. The immunogenicity of both epitopes depends on the Arg residue and its substitution with Gly completely eliminates their binding to their corresponding HLA molecules. In the HLA imputation, multiple HLA class II haplotypes including *HLA-DQB1*03:02* ($\beta = 0.27$, $p = 2.3E-25$) and *HLA-A*24:02* class I haplotype ($\beta = 0.22$, $p = 4.9E-10$) were associated with AAD. A few known T1D loci were associated with AAD at nominal level ($p < 0.05$) but the majority of them were not associated at all. The results indicate that most T1D susceptibility loci do not affect AAD and the two identified loci for AAD are both involved in the immune system. Genetic studies can provide insight into heterogeneity within clinically-defined types.

Complex Traits Posters - Wednesday

PB1546. Multi-tissue splicing transcriptome-wide association study identifies 25 new candidate susceptibility genes and a new risk region for Alzheimer's disease

Authors:

G. Gao¹, J. McClellan¹, E. Wu², A. P. Wingo³, T. S. Wingo³, H. Im²; ¹Dept. of Publ. Hlth.Sci., Univ. of Chicago, Chicago, IL, ²Section of Genetic Med., Univ. of Chicago, Chicago, IL, ³Emory Univ. Sch. of Med., Atlanta, GA

Abstract Body:

Genome-wide association studies (GWAS) and transcriptome-wide association studies (TWAS) have discovered numerous genomic loci and genes, respectively, associated with Alzheimer's disease (AD); yet the causal genes are incompletely identified. We performed a joint intron splicing-based TWAS analysis for AD that combined information from multiple excised introns in a gene across multiple brain tissues by an aggregated Cauchy association test (ACAT). Specifically, the ACAT method constructed a test statistic by combining p-values of splicing TWAS for single introns and single tissues. We used splicing prediction models trained in 12 brain tissues in the GTEx (v8) data and summary statistics from genome-wide AD meta-analysis of results from 1) a genome wide association study by proxy (GWAX) (53,042 proxy cases, and 355,900 controls); and 2) AD case-control GWAS of 21,982 cases with diagnosed AD and 41,944 controls.

We identified 65 genes at 17 loci significantly associated with AD at the Bonferroni threshold. Among them, 24 genes at 10 loci were novel, having not been reported by previous TWAS or implicated by previous GWAS. To further determine which of the identified genes more likely to be causal, we performed a gene-based fine-mapping of TWAS by using the software package FOCUS. Among the 24 genes, seven genes at three loci were in credible sets at a nominal confidence level of 90% and had high posterior inclusion probabilities (PIPs) equal to or close to 1.0. The seven genes are: GATS (PIP=1) at locus 7q22.1; five genes HLA-DRB6 (PIP=1), HLA-DRB5 (PIP=1), XXbac-BPG154L12.4 (PIP=1), XXbac-BPG154L12.5 (PIP=1), and COL11A2 (PIP= 0.979) at locus 6p21.32; and UFC1 at locus 1q23.3. These genes are very likely causal. In addition, we performed joint splicing TWAS analysis by using the summary statistics only from the diagnosed AD case-control GWAS (without using proxy cases, see above). We identified a novel gene ELL in a novel locus (19p13.11), which was more than 5Mb away from previous published GWAS index SNPs. In summary, we identified 25 new candidate susceptibility genes for AD at 11 loci; one locus is in a novel risk region. These 25 novel susceptibility genes are promising targets for future mechanistic studies to understand how differentially spliced genes cause Alzheimer's disease.

Complex Traits Posters - Thursday

PB1547*. Multi-tissue transcriptome-wide association study (TWAS) uncovers 10 novel genes associated with multiple ageing outcomes.

Authors:

G. Navoly, O. Giannakopoulou, S. Mueller, L. Partridge, N. Alic, K. Kuchenbaecker; UCL, London, United Kingdom

Abstract Body:

Ageing brings about increased disease susceptibility. The rate at which an individual ages is partly determined by their genetic make-up. Previous genome-wide association studies (GWAS) identified only a few genome-wide significant loci, affecting lifespan. An alternative approach is through a gene-based association analysis called transcriptome-wide association study (TWAS), which integrates GWAS with gene expression data (eQTL), to model the association between gene expression levels and ageing. In this study, we performed a multi-tissue TWAS on human lifespan, using >1 million parental lifespans from the UK Biobank and the LifeGen Consortium, using S-MultiXcan software. We conducted two validation TWAS on two different ageing outcomes: longevity and healthspan (morbidity-free lifespan). We also performed TWAS fine-mapping, using FOGS software, to identify putative causal gene candidates. Finally, to assess the transferability of ageing-related findings from model organisms to humans, we tested for enrichment of genes in hallmark pathways: in the RAS, TORC1 and IIS pathways. We identified 563 genes whose expression was associated with parental lifespan (FDR <0.1), out of which, 137 novel genes were also associated with at least one other ageing outcome: longevity or healthspan (replication threshold $p < 0.05$). 10 of these novel genes, including *TOMM40*, *HTR3B*, *VARS*, *VWA7* and *COASY*, were associated with all the three ageing outcomes. The *HTR3B* gene, encoding a receptor for serotonin, has a function in cognition and memory, and has been linked to pathways in ageing. Furthermore, the 563 genes associated with parental lifespan were located in 109 physically distant (+/- 100kb) clusters. We used FOGS fine-mapping tool to identify likely causal genes for the 109 multi-gene clusters and successfully resolved several loci to one likely target gene (PIP >0.8 of being putatively causal), including *TOMM40*, *FES*, *PSMA4*, *IREB2*. Our analysis assessing the transferability of previous findings from model organisms to humans replicated several key genes of the RAS, TORC1 and IIS pathways, including well-established longevity genes *FOXO3* and *SESNI*. We showed successful application of the TWAS approach to validate previous GWAS findings, and to identify several novel genes associated with multiple ageing outcomes. Finally, our results support the key role of hallmark pathways in ageing in humans.

Complex Traits Posters - Wednesday

PB1548. Multitrait analysis genome-wide association study of atherosclerosis phenotypes

Authors:

N. Hasbani¹, J. Hahn¹, A. Heath¹, D. Meena², M. Bos³, A. Dehghan², M. Kavousi³, B. Voight^{4,5}, T. Assimes^{6,7}, S. Damrauer⁸, P. de Vries¹; ¹Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ²Dept. of Biostatistics and Epidemiology, MRC-PHE Ctr. for Environment and Hlth., Sch. of Publ. Hlth., Imperial Coll. London, London, United Kingdom, ³Dept. of Epidemiology, Erasmus Univ. Med. Ctr., Rotterdam, Netherlands, ⁴Dept. of Systems Pharmacology and Translational Therapeutics and Dept. of Genetics, Perelman Sch. of Med., Philadelphia, PA, ⁵Inst. of Translational Med. and Therapeutics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ⁶Stanford Univ Sch Med., Palo Alto, CA, ⁷VA Palo Alto Hlth.Care System, Palo Alto, CA, ⁸Dept. of Surgery, Perelman Sch. of Med., Univ. of Pennsylvania,, Philadelphia, PA

Abstract Body:

While over 300 genetic loci associated with coronary artery disease (CAD) have been discovered, genome-wide association studies (GWAS) of other atherosclerosis phenotypes have been less successful. Measures of atherosclerosis, such as coronary artery calcification (CAC) and carotid intima media thickness (cIMT), relate more closely to the specific pathophysiological mechanisms that lead to CAD and are highly correlated with CAD and other cardiometabolic risk factors. Our objective is to leverage these correlations to detect shared genetic loci for atherosclerosis phenotypes. Summary statistics from GWAS of 7 atherosclerosis phenotypes (CAD, cIMT, CAC, arterial stiffness index(ASI)) or cardiometabolic risk factors (type 2 diabetes, low density lipoprotein cholesterol (LDL-C), systolic blood pressure (SBP)) were combined to perform 24 separate multi-trait GWAS using N-GWAMA. A total of 1,183 pleiotropic genetic variants met the GWAS significance threshold of 5×10^{-8} and 925 met a Bonferroni-corrected significance threshold of 2.1×10^{-9} . Of the 925 variants, 575 were novel for at least one atherosclerosis phenotype (CAD: 146, CAC:217, cIMT:307, ASI:172). Most variants detected for atherosclerosis phenotypes were identified in loci with known associations with SBP, LDL-C, and type 2 diabetes. Using HyPrColoc, newly identified atherosclerosis variants that met the multi-trait significance threshold were further evaluated to identify clusters of traits with evidence of a shared causal variant. Sixteen novel variants for CAD and atherosclerosis measures also colocalized ($P_{RPA} > 80\%$) with clusters of cardiometabolic risk factors near or in genes whose known function could lead to excess CAD (*ABCB11*, *ADRB1*, *ASTN2*, *ATP1B3*, *BCL2*, *E2F3*, *FAIM2*, *HBSIL*, *HMGN2P25*, *KDM4B*, *PDGFC*, *PPARA*, *RPS6KB1*, *SLC24A3*, *SRRMI*, *Y_RNA*). For example, rs853777 near *ABCB11* was significantly associated with SBP, cIMT, CAD, and identified as a shared causal variant across CAD and SBP. While *ABCB11* has never been associated with atherosclerosis or SBP before, it is a known LDL-C and glucose locus, with a potential role in regulating liver function. Similarly, rs7776054 in *HBSIL* was significantly associated with CAD and ASI, and colocalized with LDL-C. Per GTEx, rs7776054 is a significant eQTL and sQTL of *HBSIL* in arterial and musculoskeletal tissue, and the gene has an important role in hemoglobin concentration. Thus, multi-trait GWAS of atherosclerosis identified loci not yet detected in single-trait GWAS with potential insights into the shared biological mechanisms leading to atherosclerosis and CAD.

Complex Traits Posters - Wednesday

PB1549. Multi-trait rare variant analysis of cardiometabolic traits

Authors:

W. Bone, A. Verma, T. G. Drivas, Y. Bradford, Penn Medicine Biobank, Regeneron Genetics Center, B. F. Voight, M. D. Ritchie; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Rare variant association tests on biobank cohorts have shown enormous promise to characterize the contribution of genetics to human traits. These experiments have shown strong evidence that rare protein-coding variants contribute to the risk of many common complex diseases. However, tests for rare variants carry several technical challenges: owing to the variant frequency, they suffer from a lack of power, and are quite sensitive to population stratification which can generate false positives. To address these challenges, we have implemented a framework to perform multi-trait association test of rare variant burden. Multi-trait methods using correlated traits have been previously shown to increase statistical power for discovery, and the use of burden tests and including common variant, rare variant, and local variant principal components as covariates can mitigate the impact of population stratification. This framework is capable of performing these experiments using both quantitative and binary traits. We have performed a proof-of-concept multi-trait rare variant analysis of lipid levels, using high-density lipoprotein, low-density lipoprotein, and triglycerides, as well as a liver enzymes multi-trait rare variant analysis, using alanine transaminase, aspartate transaminase, and alkaline phosphatase in the Penn Medicine Biobank (PMBB) 45K cohort. For circulating lipid level, we replicated five genes (*ABCA1*, *ANGPTL3*, *APOC3*, *LDLR*, and *PCSK9*) previously identified in an analysis of 200K exomes of UK Biobank (UKBB, Jurgens et al., 2021) and from 300K samples from the Global Lipids Genetics Consortium cohort (Lou et al., 2020). In the liver enzymes analysis, we detected three genes previously identified as likely causal genes at loci identified by genome-wide association studies (GWAS; *ALPL*, *GPLDI*, and *GPT*) as well as one gene not previously reported, *ANKRD36C*. We plan to extend this effort to perform a series of multi-trait scans for cardiometabolic rare variant in PMBB and UKBB to identify genes that modify coronary artery disease risk, type 2 diabetes risk and the quantitative traits that are associated and potentially causal for these diseases, including lipid levels and anthropometric traits. These experiments will give us a better understanding of the contribution of rare-coding variation to cardiometabolic disease and the pleiotropy between cardiometabolic traits. These experiments could also identify novel therapeutic targets. An advantage of these multi-trait rare variant association analyses over GWAS is that the results point directly to the potential therapeutic target gene.

Complex Traits Posters - Thursday

PB1550. Multivariate GWAS of 3D cranial vault shape in multi-ethnic and admixed children

Authors:

s. Goovaerts^{1,2}, H. Hoskens^{1,2}, R. J. Eller³, A. M. Musolf⁴, M. Yuan^{1,2}, N. Herrick³, S. Naqvi⁵, M. Lee⁶, H. L. Szabo-Rogers⁶, P. A. Romitti⁷, S. A. Boyadjiev⁸, J. R. Shaffer⁶, M. D. Shriver⁹, J. Wysocka⁵, S. Walsh³, S. Weinberg⁶, P. Claes^{1,10,2}; ¹KU Leuven, Leuven, Belgium, ²Med. Imaging Res. Ctr., MIRC, UZ Leuven, Leuven, Belgium, ³Indiana Univ. Purdue Univ. Indianapolis, Indianapolis, IN, ⁴Rutgers Univ., Piscataway, NJ, ⁵Stanford Univ. Sch. of Med., Stanford, CA, ⁶Univ. of Pittsburgh, Pittsburgh, PA, ⁷Univ. of Iowa, Iowa City, IA, ⁸Univ. of California Davis, Sacramento, CA, ⁹Penn State Univ., University Park, PA, ¹⁰Murdoch Childrens Res. Inst., Melbourne, Australia

Abstract Body:

Genome-wide studies have now identified several hundred loci associated with quantitative measures of human facial shape. Other regions of the craniofacial complex, by comparison, have received relatively little attention. The cranial vault - the portion of the skull that surrounds the cerebral cortex and cerebellum - is highly variable, has strong clinical relevance, and shows high heritability. Nevertheless, very little is known about the genetic basis of normal-range vault shape. Moreover, most genetic studies of craniofacial traits lack ethnic diversity, which can hamper discovery efforts. Here, we conducted a multivariate GWAS of 3D cranial vault shape extracted from magnetic resonance images in 4198 European ancestry and 2570 non-European ancestry (mostly admixed) children. By adjusting 3D cranial vault shape for both global and local genetic ancestry components, individuals that consist of up to a 3-way admixture component (having European, African, and American ancestry) could be included in a single genome-wide scan with boosted sample size and power while keeping Type I Error regulated. This yielded 30 independent genome-wide significant genetic loci, of which 20 were replicated in a dataset of 16,947 individuals from the UK Biobank. Amongst these loci, include transcription factors and growth factors known to play a role in craniofacial development (TBX15, DLX6, ALX1), brain development/morphology (ZEB2, RSPO2, RSPO3), and cranial suture morphogenesis (RUNX2, FGF10, FGF18, SHH). Collectively, our GWAS hits were enriched for processes related to skeletal development and showed elevated levels of transcription in cranial neural crest cells, suggesting a role during early craniofacial development. Most of the discovered loci were also implicated in disorders that result in abnormal craniofacial development in humans and/or model systems. More specifically, we found that our signals near BMP2, BBS9, and ZIC2 were also significantly associated (after correcting for multiple testing) with non-syndromic sagittal craniosynostosis in a case-parent trio dataset of 189 probands. These results affirm that common variation near genes implicated in abnormal craniofacial development also contribute to normal variation with regards to cranial shape and the craniofacial complex in general.

Complex Traits Posters - Wednesday

PB1551. Natural language processing and modelling of clinical brain disease trajectories

Authors:

I. Holtman¹, N. Mekkes¹, M. Groot², A. Rozemuller³, I. Huitinga⁴; ¹Univ. Med. Ctr. Groningen, Groningen, Netherlands, ²The Netherlands Brain Bank, Netherlands Inst. for NeuroSci., Amsterdam, The Netherlands, Amsterdam, Netherlands, ³Amsterdam Med. Ctr. location VUmc, Amsterdam, Netherlands, ⁴Netherlands Inst. for NeuroSci., Amsterdam, Netherlands

Abstract Body:

The brain is a highly complex organ that consists of many intricately linked substructures and cell types and is susceptible to a wide range of brain disorders. Many brain disorders show considerable overlap in clinical manifestations, genetic risk factors, and pathophysiological mechanisms, suggesting a complex relationship between clinical symptoms, neuropathology, and genetic susceptibility both within and between clinical diagnosis. Brain diseases are often wrongly diagnosed, as up to 30% show a mismatch between a clinical and neuropathologically based diagnosis. A better delineation of the relationships between these disorders might enable improved diagnosis, better prognosis, and personalized therapeutic interventions. This clearly illustrates the need for a cross-diagnostic, data-driven approach. To this end, we recently established the Netherlands Neurogenetics Database which aims to integrate the extensive clinical, and neuropathological data of the Netherlands Brain Bank (NBB) with genetics data from a large number of donors. A first aim of this project is to convert a large corpus of clinical history data, consisting of medical summaries made by medically trained staff of the NBB, into standardized clinical disease trajectories. For this, we aimed to identify and define clinical signs and symptoms associated with brain disorders. We identified 90 signs and symptoms, divided over 14 groupings, within 5 domains, such as ‘memory impairment’ and ‘hallucinations’. Different Natural Language Processing (NLP) models, including Google BERT, were trained and optimized with Optuna to identify these signs and symptoms in individual sentences in the clinical history data. After Optuna optimization, BERT and T5 were able to reliably identify most signs and symptoms with a micro-F1 score above 0.9. Out of the 90 attributes, 81 reached an F1 score of 0.8 or higher in the best performing BERT model. The final model was used to predict signs and symptoms in the full corpus. We subsequently compiled the attributes according to the main neuropathological diagnoses from the NBB and identified clear disease-associated signs and symptom profiles, supporting the validity of this approach. In addition, we were able to identify signs and symptoms that differ between disease subtypes, which might help to distinguish brain-disorder subtypes. Finally, these clinical attributes, in conjunction with polygenic risk scores were used to model the temporal aspects of disease progression, and to train machine learning models to more reliably predict neuropathological diagnosis.

Complex Traits Posters - Thursday

PB1552. Neurobiological pathways underlying opioid addiction identified by systems biology multi-omic gene integration.

Authors:

K. Sullivan¹, D. Kainer¹, M. Lane², M. R. Garvin¹, A. Townsend², B. C. Quach³, C. Willis³, N. C. Gaddis³, R. Mathur³, O. Corradin⁴, B. S. Maher⁵, P. C. Scacheri⁶, S. Sanchez-Roige⁷, A. A. Palmer⁷, V. Troiani⁸, E. Chesler⁹, D. B. Hancock³, E. O. Johnson³, D. A. Jacobson¹; ¹Oak Ridge Natl. Lab., Oak Ridge, TN, ²Univ. of Tennessee-Knoxville, Knoxville, TN, ³RTI Intl., Research Triangle Park, NC, ⁴MIT, Cambridge, MA, ⁵Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ⁶Case Western Reserve Univ., Cleveland, OH, ⁷UCSD, La Jolla, CA, ⁸Geisinger Hlth.System, Danville, PA, ⁹The Jackson Lab., Bar Harbor, ME

Abstract Body:

Recent genome-wide association studies (GWAS) and omic-wide gene dysregulation studies in postmortem human brains have begun to robustly identify variants and genes associated with opioid addiction (OA). Linking OA genes to neurobiological pathways is critical to elucidating biological drivers underlying the observed associations and providing hypotheses for biological context-aware functional experiments. We applied a novel systems biology approach to connect OA-associated genes using a multiplex network with network layers from distinct types of biological experimental evidence, including six brain region-specific gene expression networks constructed using explainable artificial intelligence and RNA-seq data. Our initial analysis focused on 15 OA-associated genes: five GWAS-derived genes (*OPRM1*, *FURIN*, *KDM4A*, *PPP6C*, *PTPRF*), five differentially expressed genes (DEGs; *DUSP4*, *DUSP6*, *EGR4*, *ETV5*, *NPAS4*) from postmortem dorsolateral prefrontal cortex (dlPFC), and five genes from opioid-related epigenetic alterations to the dlPFC, including H3K27 hypoacetylation by ChIP-seq (*ASTN2*, *DUSP4*, *ENOX1*, *GABBR2*, *KCNMA1*) and DNA hypermethylation (*NTN1*). Using network mining algorithms, we identified a strong level of functional connection among these genes, with a high level of gene recovery (AUC=0.93) and a tight network of only 96 additional genes needed to connect all 15 OA-associated genes. Novel connections between genes were observed, including a scaffolding protein (*FLNA*) binding to both *OPRM1* and *FURIN*. Results also included multiple functional links between genes that modulate MAPK signaling (*DUSP4*, *DUSP6*, and *PPP6C*), which responds to morphine exposure. We then expanded our analysis to 182 OA-associated genes, including genes from recently characterized GWAS loci, differentially expressed proteins from the dlPFC using LC/MS proteomics, and additional DEGs from postmortem nucleus accumbens (NAc) and dlPFC. We used our recently developed systems biology tool GRIN (Gene set Refinement using Interacting Networks) to identify the genes that were most biologically interrelated among OA-associated genes in the dlPFC and NAc. Using gene set enrichment analysis from genes retained by GRIN with the larger 182 gene set, we confirmed the involvement of the MAPK signaling pathway in OA and identified a novel enrichment for transcriptional repression that was not identified prior to GRIN. Integrating these multi-omic datasets using AI-derived networks, we linked seemingly disparate single omics findings to neurobiological pathways in multiple brain regions, supporting a new conceptual model of the interplay of genes underlying OA.

Complex Traits Posters - Thursday

PB1553. Neuroimmunogenic architecture in brains of Alzheimer's disease at single-cell resolution

Authors:

Y. Chen¹, R. Dai², C. LIU³, H. Huang⁴; ¹broad Inst., Cambridge, MA, ²SUNY Upstate Med. Univ., Syracuse, NY, ⁴MGH, Boston, MA

Abstract Body:

Recent studies emphasized the crucial role of neuroinflammation in Alzheimer's disease (AD) pathogenesis. However, the expression of immune-related genes (IRGs) and possible regulatory mechanisms involved in major brain cell types of AD remain unclear. With the development of single-cell technology, we are equipped to dissect neuroinflammation at the cell type level in AD. For this study, we investigated IRGs changes in the major brain cell types using single-nucleus RNA-seq data from the Religious Orders Study and Memory and Aging Project (ROSMAP) study with 80,600 cells in the cortex across 48 individuals (24 patients and 24 matched controls).

We curated the IRGs from eight databases using uniform criteria. In total, we characterized 1,821 IRGs expressions across five major cell types in brains: excitatory neurons (Ex), inhibitory neurons (In), astrocytes (Ast), oligodendrocytes (Oli), and microglia (Mic). 1,698 IRGs were retained after data preprocessing and covariates adjustment. Differential gene expression was calculated using the linear mixed-effect model. We identified differentially expressed IRGs in each cell type (Ex: 32, In: 182, Ast: 12, Mic: 32, Oli: 78). Notably, CD81 showed significant differential expression across cell types, including microglia, astrocyte, and oligodendrocytes.

To decipher cell-cell interactions and communication changes in AD patients, we calculated the intercellular communication probability using Cellchat. We found all the brain cell types contained active cell-cell interaction of the immune pathways (IR-CCI). Furthermore, only patients' inhibitory neuron showed less IR-CCI than controls. We also ranked all the significant immune-related signaling pathways based on their differences between AD and controls. For example, we found that IL-4 signaling is more active in patients with AD than in controls.

Lastly, we conducted the single-cell eQTL analysis and identified variants affecting gene expression of IRG in these five major brain cell types. The GWAS signals were highly enriched in the microglia IRG-eQTL SNPs (enrichment = 2.3728; p.value = 3.71E-9) and neuron IRG-eQTL SNPs (enrichment = 1.7514; p.value = 0.0002).

In conclusion, we found that immune gene dysregulation events are widely present in major brain cell types in AD patients. Microglia and intercellular communication of neuroimmunity were highlighted in the AD brain for the neuroimmune dysregulation. Our work uncovers the neuroimmunogenic architecture in major human brain cells of AD patients, which will inform our understanding of the underlying mechanism of AD.

Complex Traits Posters - Thursday

PB1554. Neuronal Assay for Investigating Circadian Rhythm Defects in Prader-Willi and Schaaf-Yang Syndrome

Authors:

A. Victor¹, C. Ramanathan², A. Liu³, Y. Shen³, L. Reiter¹; ¹Univ Tennessee HSC, Memphis, TN, ²Univ. of Memphis, Memphis, TN, ³Univ. of Florida, Gainesville, FL

Abstract Body:

Prader-Willi syndrome (PWS) is a neurodevelopmental disorder defined by a range of phenotypes, including developmental delay, intellectual disability, sleep disorders and increased autism risk. PWS is an imprinting disorder caused by a loss of paternal expression of critical genes in the 15q11.2-q13 regions, including *MAGEL2*, *SNRPN/SNURF*, and *SNORD116*. PWS patients are known to suffer from various sleep disorders, including sleep-disordered breathing and central hypersomnolence. The suprachiasmatic nucleus (SCN), within the hypothalamus, controls the circadian rhythms of many biological processes. Mouse models of PWS have alterations in their circadian rhythms, including REM sleep alterations and defects in rhythmic DNA methylation. While in cultured cells, *Magel2* was found to regulate *Bmal1* and *Per2*, both critical genes in the circadian pathway. Schaaf-Yang syndrome (SYS), a disorder with overlapping phenotypes in common with PWS, results from truncating mutations in the *MAGEL2* gene. Here, we investigated the expression of key Clock gene, *Per2*, across time in dental pulp stem cell (DPSC) patient-derived neurons from PWS, SYS and neurotypical control subjects. We have established the largest collection of DPSC lines from the deciduous teeth of PWS subjects (46 total). We regularly differentiate these stem cells into cortical-like neurons for molecular and functional studies. To investigate the circadian rhythms of PWS and SYS patients, we established DPSC cell lines harboring a *Per2* promoter-driven luciferase reporter (*Per2:Luc*) cell line to assess in vitro circadian bioluminescence rhythm. The reporter stem cells were differentiated for 4-weeks to achieve a mature neuronal reporter cell line. Using a Lumicycle instrument, which measures luciferase activity across time, we were able to observe the circadian bioluminescence rhythms of these patient derived neuronal cell lines across several days. We observed a shorter period length of bioluminescence rhythms from the PWS and SYS neurons compared with the neurotypical control neurons. We aim to use this assay to further examine the defective circadian rhythms in PWS and SYS and to perform drug screens in order to rescue these circadian defects.

Complex Traits Posters - Wednesday

PB1555. NGS testing and risk stratification of malignant cardiac arrhythmias

Authors:

N. Kokalj Vokac^{1,2}, Š. Stangler Herodež^{1,2}, D. Krgović^{1,2}, D. Vokač^{1,2}; ¹Univ. Medical Ctr. Maribor, Maribor, Slovenia, ²Med. Faculty, Univ. Maribor, Maribor, Slovenia

Abstract Body:

Cardiac arrhythmias are frequent in patients with nonischemic cardiomyopathy (CMP) and can precipitate sudden cardiac death (SCD) before symptoms set up. In majority malignant cardiac arrhythmias are nonspecific consequence of advanced hearth disease although in many cases could present a primary onset and proceed overt heart before structural and electrical remodeling. In the clinical practice risk stratification criteria for prophylactic implantable cardioverter defibrillator implantation (ICD) are mostly dependent on structural heart diseases exclusion and arrhythmogenic substrate approval by use of standard invasive and noninvasive clinical tests on the other hand nonspecific ECG changes in asymptomatic patients or patients with modest clinical signs many times are overlooked in these a routine new generation genetic testing (NGS) could be important for better clinical decision. Using NGS 202 patients age 51,2±19,4 Y patients were tested 83 with syncope or after aborted sudden death; 55 with converted ventricular tachyarrhythmia or ventricular fibrillation without structural heart disease and patients suspected for development of overt cardiomyopathy with modest symptoms or borderline cardiac diagnostic tests for recognition of arrhythmogenic cardiomyopathy; 63 patient with overt or high degree probability for relapse VT an ICD implantation and 21 patient with RF ablation procedure. NGS analysis was performed using the TruSight Cardio Sequencing Kit (Illumina, Inc., San Diego, CA, USA) on Illumina MiSeq platform. Data analysis was performed with MiSeq Reporter software 2.5.42.5 according to BWA Enrichment disc workflow. Variant Studio (Illumina, Inc., San Diego, CA, USA) software and open-access bioinformatic tools and databases were used for analysis and interpretation of variants obtained in VCF file. In study, we established molecular genetic disorder in 33% of the study population. Significant proportion 12,4 % of pathogenic and 20,8 % of VOUS variants were analyzed. In the group of patients with VOUS, 16.1% were highly clinically significant. In our presentation genotype - phenotype correlation is discussed for clinically pathological cases which according to NGS analyzes were classified as VOUS. Emphasis is placed on the importance of the results of NGS analyzes in defining the clinical condition and risk stratification with awareness of early cardiomyopathy detection, aggressive treatment and SCD prevention is of great clinical significance.

Complex Traits Posters - Thursday

PB1556. Non-desmosomal genes in Arrhythmogenic Cardiomyopathy: genetic variants rating

Authors:

M. Bueno Marinas, R. Celeghin, M. Cason, R. Bariani, M. De Gaspari, S. Rizzo, G. Thiene, M. Perazzolo Marra, D. Corrado, C. Basso, B. Bauce, K. Pilichou; Dept. Cardiac-Thoracic-Vascular Sci. and Publ. Hlth.. Univ. of Padua, Padua, Italy

Abstract Body:

Background. Arrhythmogenic Cardiomyopathy (AC) is an inherited disorder of the myocardium with a highly heterogeneous clinical presentation including life-threatening arrhythmias at risk of sudden death. Genetic testing impacts greatly in reaching AC diagnosis, but gene-disease associations has yet to be determined for the increasing number of genes included in clinical panels. Methods. Genetic variant re-appraisal was undertaken in most relevant non-desmosomal disease genes, based on current adjudication guidance, identified in 320 unrelated Italian AC patients who did not carry pathogenic/likely pathogenic (P/LP) variants in desmosome-coding genes and reported literature data. Results. In our cohort, 28 rare genetic variants in non-desmosomal genes were identified in 30 patients, of which 17 *FLNC* (Filamin C), 7 *DES* (Desmin), 2 *TMEM43* (Transmembrane protein 43), and 2 *CDH2* (Cadherin-2). No P/LP variants were found in *PLN* (Phospholamban) and *TJPI* (Tight junction protein-1) genes. Gene-based burden analysis, including P/LP variants reported in literature, showed significant enrichment only for *TMEM43* (3.52-fold), *DES* (9.55-fold), *PLN* (117.8-fold) and *FLNC* (93.22-fold). Evolutionary conservation analysis made evident a positive selection pressure (Ka/Ks ratio >1) for *CDH2* and *TJPI*, indicating that missense variants impact less the protein structure. Genotype-phenotype correlation highlighted 71% and 89% of left-dominant AC in *FLNC* and *DES* carriers, respectively. Conclusion. Genes lacking robust clinical and genetic evidences impact greatly the number of variants-of-unknown-significance detected and should be removed from clinical AC-targeted genetic panels since the findings cannot drive clinical decision-making. About two thirds of non-desmosomal P/LP variants occur in *FLNC* leading to fully-penetrant left-dominant AC.

Complex Traits Posters - Wednesday

PB1557. Non-Mendelian inheritance patterns and extreme deviation rates of CGG repeats in autism

Authors:

F. Kooy¹, D. Annear¹, A. Sanchis-Juan², L. Raymond³, G. Vandeweyer⁴; ¹University of Antwerp, Antwerp, Belgium, ²Univ. of Cambridge, Cambridge, United Kingdom, ³University of Cambridge, Cambridge, United Kingdom, ⁴Univ. of Antwerp, Antwerp, Belgium

Abstract Body:

Short tandem repeats (STRs) are tracts of DNA where short nucleotide motifs are repeated in a head-to-tail fashion. To date, greater than 30 STR-associated disorders have been described. Here, we explored the inheritance patterns of CGG trinucleotide STRs and the inheritance differences between autism-affected probands and their unaffected siblings. We utilized ExpansionHunter, to determine the genomic CGG-repeat length genotypes across a total cohort of 1978 trios and 114 quads.

Across the 6063 analysed CGG loci, 11,885,983 individual parent-to-child repeat genotypes were determined, and a total of 23,391,708 CGG-repeat transmissions were observed. Of all STR genotypes, most were monogenic across the proband, mother, and father. 15.6% of genotypes were informative, where the parental genotypes were different or heterozygous, but the proband genotype was reflective. However, 4.4% of CGG STR transmissions resulted in a deviating proband genotype. Deviation events were witnessed at 5881 of the CGG loci with an expansion-to-contraction ratio of 1.65.

This research not only solidifies previous findings surrounding STRs but specifically demonstrates new characteristics of CGG STRs. The mutation rate was strongly dependent on both the genomic location as well as intergenic positioning. While the largest portion of CGG-repeats is localised within genes, the regions that displayed the highest rates of CGG repeat mutation were the intergenic and immediate 5'-upstream regions. We solidify the idea that repeat variation rate is proportional to repeat length, however, we refine this concept and demonstrate how smaller repeats more readily demonstrate a greater degree of size variation and take on additional repeat units.

Interestingly, we observed that CGG STRs did not segregate based on Mendelian principles. The shorter repeat allele length is typically selected and reflected in offspring. Furthermore, this trend appears to magnify as the repeat length difference increases between the two parental repeat alleles. While almost all CGG STRs were determined to be polymorphic, we see how the repeats fall within an increasingly variable continuum, with a distinct subset of hypermutable CGG STRs that cluster in specific genetic regions and locations linked to neurological function and development. Some hypermutable CGG repeats then appear to be involved in autism spectrum disorder as we observe significantly higher rates and degrees of CGG repeat expansion in these regions among autism affected individuals versus their unaffected siblings. This may suggest that CGG STRs are explicitly linked to neurodevelopmental function and disorders.

Complex Traits Posters - Thursday

PB1558. Not one and done: A rare finding of two deleterious variants contributing to the onset of rhabdomyolysis.

Authors:

A. Kunovac¹, N. Berman², C. Munro², R. O. Rosti¹, X. Wang³, L. A. Berenbrok⁴, M. Massart²; ¹UPMC Genome Ctr., Pittsburgh, PA, ²UPMC Primary Care Precision Med., Dept. of Family Med., Pittsburgh, PA, ³AiLife Diagnostics, Pearland, TX, ⁴Univ. of Pittsburgh Sch. of Pharmacy, Pittsburgh, PA

Abstract Body:

Introduction: In the United States, approximately 26,000 cases of rhabdomyolysis are reported annually. Many genetic factors have been identified as contributing to the onset of rhabdomyolysis including metabolic myopathies, channelopathies, and mitochondrial disorders.

Case Presentation: This is a case of a 29-year-old male with recurrent episodes of rhabdomyolysis, resulting in multiple hospitalizations. Laboratory and metabolic evaluations did not reveal a cause. A traditional rhabdomyolysis panel determined that the patient was heterozygous for *RYRI* (AD) c.1589G>A (p.R530H), a likely pathogenic variant, and heterozygous for *PFKM* (AR) c.352G>C (p.G118R), a variant of uncertain significance. Cascade testing of his family members determined that his mother and brother were found to carry the same *RYRI* variant but did not report any symptoms. The patient returned to clinic eight months later. He continued to have more frequent episodes with additional hospitalization with CK levels ranging from 140 to >22,000 U/L. Notably, he also had low associated blood glucose levels that ranged from 69 to 89 mg/dL. The decision was made to advance to Whole Exome Sequencing (WES) due to incomplete explanation of the patient's recurrent symptoms. In addition to confirming that the patient was heterozygous for the *RYRI* and *PFKM* variants, a deep intronic variant, c.936+218G>A, in the *PFKM* gene was identified. This led our team to hypothesize that the severity of the patient's rhabdomyolysis could be the product of the *PFKM* pathway or a combination of the compounded mutations in both the *RYRI* and *PFKM* genes as an alternative explanation. The two *PFKM* variants were not found through targeted genetic testing in the mother, indicating that the patient either inherited these variants from the father in cis or possibly one or both variants are *de novo*. Testing of the father is pending. To the best of our knowledge, this is the first documented case of a patient with mutations in two different genes in a reported case of rhabdomyolysis.

Discussion: The results substantiate the idea that additional genetic testing, such as WES, should be considered when an initial genetic diagnosis is not aligning with the patient's symptoms. With approximately 5% of WES cases resulting in a dual diagnosis, individual panel tests could miss such a finding. Furthermore, in challenging cases such as the one presented herein, we propose that it is more cost-effective and therefore beneficial to both the patient and the healthcare system to perform WES rather than symptomatically treating patients over multiple hospital admissions.

Complex Traits Posters - Wednesday

PB1559. Novel Insights into Pediatric Scoliosis Revealed by Genome-wide Association Study and Whole Exome Sequencing

Authors:

H-Q. Qu¹, J. Qu¹, C. Vaccaro², X. Chang³, F. Mentch⁴, J. Garifallou¹, F. Mafra⁵, K. Nguyen⁶, M. March¹, R. Pellegrino⁷, J. Glessner¹, H. Hakonarson⁸; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²The Children's Hosp. of Philadelphia, Philadelphia, PA, ³CHOP, Bryn Mawr, PA, ⁴Children's Hosp. of Philadelphia, Philadelphia, PA, ⁵The Children's Hosp. of Philadelphia, Philadelphia, PA, ⁶Children's Hosp. of Philadelphia, Phila, PA, ⁷The Children's Hosp. of Philadelphia, Philadelphia, PA, ⁸Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

The etiology of pediatric scoliosis remains obscure. Etiologically, pediatric scoliosis has been linked to central nervous system development with the theory of asynchronous neuro-osseous growth. To investigate the complex nature of pediatric scoliosis, we sought to acquire insight into the molecular mechanisms of pediatric scoliosis pathogenesis through genomic approaches, including genome-wide association study (GWAS) and whole exome sequencing (WES). The GWAS was performed on an African American cohort including 1,386 unrelated scoliosis cases and 8,942 controls, and a European American cohort with 2,833 scoliosis cases and 11,672 controls. We uncovered 25 genes showing replicable genome-wide significance by gene-based association study in both cohorts. Shared genetic susceptibility of pediatric scoliosis and autistic disorder/mental disorders was observed from common genetic variants, suggesting there is enrichment of genes in scoliosis patients with significant roles in the development and function of the central and peripheral nervous systems. Our pediatric scoliosis results also suggest adrenocortical role based on common variants of the phosphodiesterase 11A gene (*PDE11A*), involved in cAMP and cGMP signaling in adrenal cortex. The WES was performed in an African American cohort including 310 unrelated cases and 4,032 unrelated controls, and a European American cohort with 489 cases and 3,392 controls. Mutation burden was tested using the Test Rare Variants with Public Data (TRAPD) algorithm, which we have optimized with normalized genome coverage and internal controls, as a powerful approach for rare causal variants with effects in the same directions. A total of 20 genes showed replicable genome-wide significance by mutation burden study in both cohorts. Genes related to muscular disorders are highlighted by the burden of rare coding variants in scoliosis patients, with or without the diagnosis of neuromuscular scoliosis. In summary, this study unveiled a new view of the panorama of the molecular mechanisms of pediatric scoliosis pathogenesis, represented by 25 genes identified by GWAS, and 20 genes identified by WES, with replicable genome-wide significance. As genetic susceptibility of pediatric scoliosis appears to overlap with central and peripheral nervous system disorders as well as neuroendocrine dysfunctions, causal effects from rare coding variants related to genetic scoliosis may also contribute to muscular, neurologic, and neuroendocrine defects, suggesting significant heterogeneity of the resulting phenotypes.

Complex Traits Posters - Thursday

PB1560. Obesity genomic loci are heterogeneously associated with lipid profiles in ancestrally diverse *Population Architecture using Genomics and Epidemiology (PAGE)* study.

Authors:

D. Kim¹, M. Graff¹, H. M. Highland¹, R. A. J. Smit², Z. Wang², Y. Cai³, M. Kim³, C. L. Avery¹, P. Gordon-Larsen¹, R. C. Kaplan⁴, J. E. Manson⁵, E. Marouli⁶, M. Fornage⁷, C. Gignoux⁸, C. A. Haiman⁹, S. Buyske¹⁰, T. C. Matise¹⁰, C. Kooperberg³, U. Peters³, R. J. F. Loos², K. E. North¹; ¹Gillings Sch. of Global Publ. Hlth., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ²Icahn Sch. of Med. at Mount Sinai, New York, NY, ³Fred Hutchinson Cancer Ctr., Seattle, WA, ⁴Albert Einstein Coll. of Med., New York, NY, ⁵Harvard T. H. Chan Sch. of Publ. Hlth., Boston, MA, ⁶Barts and The London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, ⁷Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ⁸Univ. of Colorado, Denver, CO, ⁹Univ. of Southern California, Los Angeles, CA, ¹⁰Rutgers Univ., Piscataway, NJ

Abstract Body:

Obesity has a heterogeneous impact on cardiovascular disease (CVD), in particular for lipid traits and across diverse populations. In this study, we utilized genetic variation to define the heterogeneous impact of obesity on CVD and to gain insight into the underlying biology that connects the metabolic disturbances of obesity to CVD. We estimated local genetic correlations between obesity (assessed by body mass index (BMI)) and lipid levels to identify loci associated with obesity and low dyslipidemia risk (Ob/LDR loci) and loci associated with obesity and high dyslipidemia risk (Ob/HDR loci) in 56,681 (34% European, 30% Hispanics, 28% African, and 8% others) *Population Architecture using Genomics and Epidemiology (PAGE)* study participants. We conducted pair-wise local genetic correlation analyses between BMI and three lipid traits (high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides (TG)) across 2,495 approximately LD-independent segments of the genome using *LAVA*. We utilized publicly available UK Biobank GWAS summary statistics (<http://www.nealelab.is/uk-biobank/>) of BMI (N~355K) and lipid traits (N~300K). We filtered the loci with statistically significant ($p < 0.05/2,495$) local heritability for both BMI and a lipid trait and identified loci with statistically significant ($p < 0.05/\text{number of tested loci}$) local genetic correlation coefficients (r_g). Ob/LDR loci displayed $r_g > 0$ for BMI-HDL and $r_g < 0$ for BMI-LDL and BMI-TG. Ob/HDR loci displayed $r_g < 0$ for BMI-HDL and $r_g > 0$ for BMI-LDL and BMI-TG. To further investigate the discovered loci, we constructed Ob/LDR and Ob/HDR loci-based BMI polygenic scores (PGS) - using SNP weights modeled from PRS-CS - and assessed the association between the PGS and BMI or continuous lipid levels in the PAGE samples. We identified 19 (14 unique) Ob/LDR loci (BMI-HDL: 5, BMI-LDL: 5, and BMI-TG: 9) - 6 were previously identified - and 83 (66 unique) Ob/HDR loci (BMI-HDL: 58, BMI-LDL: 4, and BMI-TG: 21). In the ancestry-combined results, Ob/LDR loci-based BMI PGS were significantly associated ($p < 0.05$) with both BMI and the lipid trait in a hypothesized way (β s for BMI > 0 for all three pairs, β for HDL > 0 , β s for LDL and TG < 0). Ob/HDR loci-based BMI PGS were significantly associated with BMI ($\beta > 0$) and HDL ($\beta < 0$). The identification and replication of genetically correlated obesity-lipid loci further support the importance of using genetics to define the heterogeneous impact of obesity on dyslipidemia and downstream CVD. Our study findings support and extend previous findings from European studies. Further analyses will be conducted to identify the causal variants and genes underlying these loci.

Complex Traits Posters - Wednesday

PB1561. Opportunities in national SARS-CoV-2 genomic surveillance programs for accelerating study on human genetic determinants of differential COVID-19 severity

Authors:

T. Szemes^{1,2}, T. Sladeczek¹, T. Adamik¹, M. Gaziova³, M. Nemeč¹, J. Turna¹, J. Radvanszky¹, A. Kalinakova⁴, D. Rusnakova^{2,1}, T. Sedlackova¹, M. Bohmer^{1,4}, J. Budis^{1,5}; ¹Comenius Univ. Sci. Park, Bratislava, Slovakia, ²Comenius Univ., Faculty of Natural Sci., Bratislava, Slovakia, ³Comenius Univ., Faculty of Mathematics, Physics and Informatics, Bratislava, Slovakia, ⁴Publ. Hlth. Authority of the Slovak Republic, Bratislava, Slovakia, ⁵Slovak Ctr. of Scientific and Technical Information, Bratislava, Slovakia

Abstract Body:

In March 2020 the World Health Organisation announced global pandemic of SARS-CoV-2 virus with major impact on healthcare and economy in virtually all countries in the world. SARS-CoV-2 is a cause of COVID19 which, based on pathogen-host interaction, can lead to highly variable outcomes. Along with known risk factors for severe forms like age, comorbidities, genetic factors were studied but with limited success.

Unprecedented efforts in pathogen surveillance of clinical COVID-19 cases by genome sequencing of the virus in almost all countries started to aid pandemic management. More than 10 million viral genomes have been published so far. In our country, weekly genomic surveillance of SARS-CoV-2 genomes was launched in March 2021.

Due to inclusion of new NGS laboratories Public Health authority in this surveillance programme, as well as our core sequencing laboratory with 10 years experience, there were different levels of experience with laboratory procedures, but more importantly with NGS bioinformatics and result reporting. Due to this obstacle and a growing number of sequenced samples, we developed a centralized information system for organizing the national COVID-19 sequencing effort. The system with abbreviated name NarCoS is utilized to transfer id and metadata for all planned samples, to define sequencing runs for all labs, to analyze sequencing runs and to batch evaluate and report results to national register as well as to GISAID, ENA/GenBank and Tessa. Currently all facilities use ARCTIC PCR protocol based genome enrichment of the virus and sequencing on Illumina NGS platforms. Our core laboratory contributed by sequencing approx. half of all SARS-CoV-2 sequences.

Unfortunately, only in a very limited number of our cases, the human genome was also a target for analysis. Exome sequencing was carried out in two cases of in utero lethality of the fetus after COVID-19 induced in late stages of pregnancy to rule out genetic causes. We propose to extend pathogen genomic surveillance programs to focus also on human genomic variability as this can lead to valuable knowledge on genetic risk factors and polygenic risks to stratify population groups according to their genetic risks. Different sample types for SARS-CoV-2 tests are suitable for analysis of different levels of human genome variation. Adding the exome analysis could be done by modification of library preparation protocols, which are carried out within SARS-CoV-2 sequencing, without time penalty. It is also technically possible to extend integrated information systems like ours to allow a comprehensive analysis of human exomes. Naturally, ethical and data privacy issues need to be addressed.

Complex Traits Posters - Thursday

PB1562. Osteoarthritis has high genetic heritability in 488,421 multi-ancestry participants from MVP and the UK Biobank.

Authors:

A. Wilson¹, V. Srinivasasainagendra², A. Nair¹, A. Rocco¹, J. Chiles¹, J. Richman³, S. Pyarajan⁴, H. Tiwari², D. Speed⁵, M. Bamman⁶, J. Singh³, M-L. McDonald¹; ¹Univ. of Alabama at Birmingham, Div. of Pulmonary, Allergy, and Critical Care Med., Birmingham, AL, ²Univ. of Alabama at Birmingham, Sch. of Publ. Hlth., Biostatistics, Birmingham, AL, ³Birmingham VA Hlth.care System (BVAHS), Birmingham, AL, ⁴Partners Hlth.Care, Lexington, MA, ⁵Aarhus Univ., Aarhus, Denmark, ⁶Florida Inst. for Human and Machine Cognition, Pensacola, FL

Abstract Body:

Introduction Osteoarthritis (OA) is a progressive joint disease with a poorly understood etiology. Disease management is limited primarily to symptom management (e.g., pain, inflammation). The societal and patient-centered impacts of OA among United States Veterans are profound with healthcare costs for treatment exceeding \$880 million annually. To establish more comprehensive insights into the allelic spectra of OA, we estimated heritability using summary statistics from ancestry stratified GWASes of 488,421 United States Veterans who participated in the Million Veteran Program (MVP) and participants from the UK Biobank (UKB). **Methods** Genotype data for MVP and UKB was generated from DNA typed on customized Affymetrix Axiom biobank arrays for common and rare variants. Standard quality control procedures were implemented to ensure high quality SNP and DNA samples were analyzed. HARE was used to classify subjects into major ancestry groups. OA heritability was estimated in MVP and UKB ancestry groups with $\geq 10,000$ subjects (MVP: European (EUR), African American (AFR), Hispanic (HIS)); and UKB: EUR). Heritability was calculated using the baseline linkage disequilibrium-linkage disequilibrium adjusted kinships (BLD-LDAK) method which relies on a restricted maximum likelihood algorithm. In addition, the BLD-LDAK model provides estimates of heritability stratified by categories of annotated SNPs (i.e., intronic or promotor regions of the genome) that can be used estimate the unique influence for each category on heritability. **Results** In MVP, heritability of OA was highest among Veterans of HIS descent ($h^2= 51.3\%$, $SD= 8.7\%$), followed by AFR ($h^2= 29.0\%$, $SD= 3.2\%$), and EUR ($h^2= 19.0\%$, $SD= 1.4\%$). In UKB, the heritability of OA in EUR was 7.8% ($SD= 0.8\%$). In MVP, heritability of OA was most heavily influenced by SNPs annotated to the hyper-sensitive site 2 (HS2) super-enhancer, transcription factor coding, and DNAase 1 hypersensitive site regions, in EUR, AFR, and HIS, respectfully. In MVP EUR, heritability was most heavily influenced by SNPs annotated to intronic regions. Among all ancestry groups in MVP and UKB, heritability was enriched for SNPs annotated to HS2 super-enhancer regions. **Conclusions** We demonstrate OA is highly heritable among MVP and UKB participants, with heritability estimates greater than 30%. SNPs annotated to HS2 super-enhancer regions, enriched among all MVP and UKB ancestry groups, could serve as potential biomarkers or drug targets to improve health outcomes in OA. Future directions include the investigation of SNPs in the most heavily influential annotation groups to determine whether they are of direct functional significance in OA.

Complex Traits Posters - Wednesday

PB1563. Overexpression screen of chromosome 21 genes reveals modulators of Sonic hedgehog signaling relevant to Down syndrome.

Authors:

A. Moyer¹, F-X. Fernandez², Y. Li³, D. Klinedinst³, L. Florea³, S. Thyme¹, R. Reeves³; ¹Univ. of Alabama at Birmingham, Birmingham, AL, ²Univ. of Arizona, Tucson, AZ, ³Johns Hopkins Sch. of Med., Baltimore, MD

Abstract Body:

Dysregulation of Sonic hedgehog (SHH) signaling may contribute to multiple Down syndrome-associated phenotypes, including cerebellar hypoplasia, congenital heart defects, craniofacial and skeletal dysmorphologies, and Hirschsprung disease. However, despite evidence that activation of SHH signaling rescues Down syndrome-associated phenotypes in cellular and mouse models, chromosome 21 does not encode any canonical components of the SHH pathway. To identify chromosome 21 genes that inhibit SHH signaling, we overexpressed 163 chromosome 21 cDNAs in a series of in vitro and in vivo screens. Chromosome 21 cDNAs that affected SHH signaling in two luciferase reporter lines (Shh-LIGHT2 and SmoA1-LIGHT) and in developing zebrafish embryos were further screened in a SHH-dependent cell-based assay of osteoblast differentiation. Next, cDNAs that consistently up- or downregulated SHH signaling were overexpressed in primary cerebellar granule cell precursors using lentiviral transduction. Our series of screens identified four candidate genes (*B3GALT5*, *ETS2*, *HMGNI*, and *MIS18A*) that inhibit the proliferation of granule cell precursors when overexpressed, which is a finding that is directly relevant to Down syndrome-associated cerebellar hypoplasia. We also confirmed overexpression of trisomic candidate genes with bulk RNA-seq of postnatal day 6 (P6) cerebellum in four Ts65Dn pups (stock #001924), four TcMAC21 pups (stock #035561), and eight euploid littermates. Our study indicates that overexpression of some chromosome 21 genes, including *DYRK1A*, activates SHH signaling whereas overexpression of other genes, such as *HMGNI* and *MIS18A*, inhibits SHH signaling. Moreover, overexpression of genes involved in chromatin structure and mitosis, but not genes previously implicated in ciliogenesis, regulates the SHH pathway. Because identifying which chromosome 21 genes modulate SHH signaling in vivo may suggest new therapeutic avenues for ameliorating Down syndrome phenotypes, we are currently assessing overexpression of top candidate cDNAs using zebrafish reporters of SHH activity.

Complex Traits Posters - Thursday

PB1564*. Parsing Transcriptomic and Variant Signatures in Tetrapartite Brain Regions Uncover a Collection of Novel Genes in the Neuropeptide-Neurotransmitter Axis Conferring Addiction Risk

Authors:

A. Veerappa, C. Guda; Univ. of Nebraska Med. Ctr., Omaha, NE

Abstract Body:

Chronic substance abuse is a neuropsychiatric disorder involving persistent craving, pleasure, and reward, ultimately progressing to addiction. The midbrain controls hunger, reward, and pleasure traits, whereas the dorsolateral prefrontal cortex (DLPFC) controls craving, decision making, and tolerance traits. In contrast, Nucleus Accumbens (NAc) controls feeding, sexual, reward, stress-related, and drug self-administration behaviors, while amygdala regulates emotion and memory. To understand these complex and dynamic events in the background of substance use disorders (SUDs), we performed transcriptome and variant calling approaches combined with a multi-pronged strategy involving transcriptome clustering, classification, and variant filtering. Distinct transcriptomic signatures were observed that were unique and shared across groups. Significant upregulations of genes *CSF3*, *GADD45B*, *SOCS3*, and *NPAS4* were observed in all three regions of midbrain, DLPFC, NAc and amygdala highlighting the presence of their respective traits (motivation, reward memory, and tolerance) in long-lasting maladaptations of neurocircuitry due to chronic substance abuse. Distinct mutational spectrum in genes *CCKAR*, *NPAS4*, *TACR1*, *TENM2*, *EGF*, *GRM4*, and *NPY2R* were identified in the regulatory axis conferring susceptibility towards experimentation, substance use, reward, tolerance, and dependency. By unraveling the transcriptomic and mutational signatures between midbrain, DLPFC, NAc and amygdala, our study advances the understanding of the biology of substance use and its progression towards addiction, providing potential stage-specific testable targets.

Complex Traits Posters - Wednesday

PB1565. Partitioned polygenetic risk scores of adipocyte marker genes and sex-specific GWAS variants explain sex-specific differences in abdominal obesity

Authors:

H. Huang^{1,2}, A. Kar¹, M. Deal¹, M. Alvarez¹, K. Mohlke³, K. Pietiläinen^{4,5}, M. Laakso⁶, J. Sinsheimer^{1,7}, P. Pajukanta^{1,2,8}; ¹Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ²Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, ³Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁴Obesity Res. Unit, Res. Program for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, ⁵Obesity Ctr., Endocrinology, Abdominal Ctr., Helsinki Univ. Central Hosp. and Univ. of Helsinki, Helsinki, Finland, ⁶Dept. of Med., Univ. of Eastern Finland and Kuopio Univ. Hosp., Kuopio, Finland, ⁷Dept. of Computational Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ⁸Inst. for Precision Hlth.at UCLA, Los Angeles, CA

Abstract Body:

Abdominal obesity is a common sex-stratified complex disorder, investigated using waist-hip ratio adjusted for body mass index (WHRadjBMI) as a proxy. Previous genome-wide association studies (GWAS) have identified sex-specific loci and shown that WHRadjBMI GWAS loci are enriched for genes expressed in adipose tissue, suggesting that partitioning polygenetic risk scores (PRS) by sex based on both genetic and adipose cell-type-specific genes can inform us about the key genomic loci and genes driving the sex differences in abdominal obesity. To this end, we compared sex-stratified WHRadjBMI PRSs built using lassosum with split-validation on 3 sets of SNPs in the UK Biobank with 50% partitioning of the unrelated Europeans: 1) SNPs in cis-regions of sex-specific adipocyte marker genes, 2) sex-specific WHRadjBMI GWAS SNPs, and 3) full genome. To first identify sex-specific adipocyte marker genes, we assigned adipocyte marker genes in adipose single-nucleus RNA sequencing (snRNA-seq) data (n=15 Finns; 60% female), followed by replications (n=6 Finns; 50% female). We discovered 31 replicated sex-specific adipocyte marker genes, and built sex-specific adipocyte (SSA) PRSs using all SNPs (MAF>1%) in the cis-regions (+/-500kb) of these 31 genes (~117k SNPs), which explain 0.377% and 0.0642% of WHRadjBMI variance in females vs males. Next, we retrieved WHRadjBMI GWAS SNPs that are significant in females ($p < 5 \times 10^{-8}$) but not even nominally significant in males ($p \geq 0.05$) (4,002 unclumped SNPs), and vice versa (516 unclumped SNPs), from the GIANT Consortium GWAS summary statistics. These sex-specific GWAS (SSG) PRSs explain 4.46% of WHRadjBMI variance in females and 0.442% in males. When comparing these two partitioned PRSs, SSAs and SSGs, to the full genome (~8M SNPs), that explains 9.53% of WHRadjBMI variance in females and 2.77% in males, the female SSA PRS explains 2.77 times more of WHRadjBMI variance than expected by the number of covered loci, and the female SSG PRS, built with only 4,002 SNPs, explains almost half of the WHRadjBMI variance in the full genome PRS. The majority (81%) of the 31 SSA genes are correlated with insulin resistance (Q-value<0.05 for the Matsuda index) in the METSIM adipose bulk RNA-seq cohort (n=335). The adjacent and overlapped genes of the clumped SSG SNPs are enriched for a rheumatoid arthritis pathway (FDR=0.0294) by Webgestalt, and the SSG SNPs for 8 TF motifs by HOMER. Overall, we discover that the 31 adipocyte marker gene regions and sex-stratified GWAS loci explain more of abdominal obesity variance in females (4.84%) than males (0.51%), suggesting a larger role of environment and gene-environment interactions in male abdominal obesity.

Complex Traits Posters - Thursday

PB1566. Patterns of human germline hypermutability identified with whole-genome sequencing

Authors:

S. Dong¹, A. Ljungdahl², G. B. Schwartz³, J. An⁴, D. Werling⁵, S. Sanders²; ¹Univ California San Francisco, San Francisco, CA, ²Univ. of California San Francisco, San Francisco, CA, ³Ann & Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, ⁴Korea Univ., Seoul, Korea, Republic of, ⁵Univ. of Wisconsin-Madison, Madison, WI

Abstract Body:

Germline mutations generate the substrate of evolution, however, a subset of these mutations lead to human disorders, including neurodevelopmental delay. Thus, factors such as parental age that influence germline mutation rate can have widespread impacts on childhood morbidity and mortality.

Here we report the detection of three individuals exhibiting germline hypermutability in an independent cohort including 3,736 offspring from 1,868 quartet families (two parents, two offspring) with whole-genome sequencing of blood-derived DNA, validated independently with PCR and Sanger sequencing. Two brothers both exhibit a three-fold excess of germline *de novo* single nucleotide variants (SNVs) enriched for paternal origin (96%) with similar mutational profiles and signatures (COSMIC signature SBS37) though no specific mutations are shared. This observation strongly suggests the presence of a risk factor in the father's germline, however phenotypic data did not report prior cancer diagnoses or use of chemotherapeutic medications. Analysis of the paternal WGS data did not identify a causal variant in a gene related to DNA repair of cancer risk.

The third offspring was found to have a novel pattern of postzygotic germline hypermutability. Unlike all previously reported cases, the excess mutations were predominantly small insertions and deletions (indels) with no clear bias in parental origin. The indels are enriched by COSMIC signature ID2 (T/A homopolymers). Analysis of the allele balance reveals a higher-than-expected fraction of both *de novo* SNVs and indels arising postzygotically, suggesting that the mechanism was not exclusive to indels. Finally, by leveraging the family-based two offspring design of the cohort to compare mutation rates between siblings, we demonstrate that, aside from paternal age, shared genetic or environmental factors have minimal contribution to population-wide germline mutation rates, suggesting that the germline may exhibit relative protection from mutagenesis.

Complex Traits Posters - Thursday

PB1567. Pediatric Manifestations of Genetic Risk Factors for Polycystic Ovary Syndrome

Authors:

J. Zhu^{1,2,3}, A. U. Eliassen^{4,5}, I. M. Aris⁶, S. E. Stinson⁷, J-C. Holm^{7,8}, T. Hansen⁷, M-F. Hivert⁶, K. Bønnelykke⁴, J. N. Hirschhorn^{1,2,3,9}, Y-M. Chan^{1,2,3}; ¹Div. of Endocrinology, Boston Children's Hosp., Boston, MA, ²Programs in Metabolism and Med. and Population Genetics, The Broad Inst. of MIT and Harvard, Cambridge, MA, ³Dept. of Pediatrics, Harvard Med. Sch., Boston, MA, ⁴Copenhagen Prospective Studies on Asthma in Childhood Res. Ctr. (COPSAC), Copenhagen, Denmark, ⁵Dept. of Hlth. Technology, Section for Bioinformatics, Technical Univ. of Denmark, Kongens Lyngby, Denmark, ⁶Div. of Chronic Disease Res. Across the Lifecourse, Dept. of Population Med., Harvard Med. Sch., Harvard Univ. and Harvard Pilgrim Hlth. Care Inst., Boston, MA, ⁷Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Faculty of Hlth. and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark, ⁸The Children's Obesity Clinic, Dept. of Pediatrics, Copenhagen Univ. Hosp., Holbæk, Denmark, ⁹Dept. of Genetics, Harvard Med. Sch., Boston, MA

Abstract Body:

Background: Polycystic ovary syndrome (PCOS) is a common polygenic disorder of reproductive-age women that is associated with ovulatory dysfunction (irregular periods), hyperandrogenism (increased levels and/or action of androgens such as testosterone), and cardiometabolic dysfunction (e.g., obesity and abnormal lipid levels). However, it is not known whether genetic risk factors for PCOS have manifestations in children. Proposed childhood precursors to PCOS include obesity and premature adrenarche, a common but poorly understood condition with early production of androgens from the adrenal glands. Indeed, girls with premature adrenarche often have a maternal history of PCOS, suggesting shared causal factors across these conditions. A polygenic risk score (PRS) from PCOS-associated genetic variants can be used to identify childhood precursors of this disorder. **Methods:** We used PRS-CS (continuous shrinkage) to optimize a PRS for PCOS in women from the UK Biobank, then to calculate PCOS PRSs for 5,968 children in the Avon Longitudinal Study of Parents and Children (ALSPAC). We used linear regression to assess for associations of the PCOS PRS with a proxy for premature adrenarche (age at pubarche, i.e., first appearance of pubic hair) and with measures of adiposity (body mass index [BMI] and total and trunk fat-mass indices) at ages 4 to 19 years. We sought to replicate associations between PCOS PRS and BMI in Copenhagen Prospective Studies on Asthma in Childhood (COPSAC; N=450), Project Viva (N=512), and The HOLBAEK Study (N=4,005). **Results:** Children who carried a high PCOS PRS had a younger age at pubarche (estimated effect size per 1 SD increase in PRS=-1.0 month, $p=0.001$), higher BMI SD score (SDS) as early as 7 years of age (estimated effect size per 1 SD increase in PRS=+0.04 BMI SDS, $p=0.003$), and higher total-body and trunk fat mass indices at 9 years of age (estimated effect size per 1 SD increase in PRS=+0.017, $p=7 \times 10^{-8}$ and +0.009, $p=3 \times 10^{-8}$, respectively). The association between PCOS PRS and BMI SDS replicated in Project Viva and HOLBAEK but not in COPSAC (estimated effects per 1 SD increase in PRS = +0.15, +0.15 and -0.05, 1-tailed $p=0.0023$, 5×10^{-10} and 0.17, respectively). **Conclusions:** PCOS genetic risk factors are associated with cardiometabolic and androgenic phenotypes in children. Thus, genetic risk factors for PCOS, a disorder of reproductive-age women, can result in early perturbations in yet-to-be-determined biological pathways that underlie PCOS in children. Future dissection of the underlying molecular pathways will inform strategies to prevent, halt, and reverse these perturbations before the development of PCOS and its related conditions.

Complex Traits Posters - Wednesday

PB1568. Perceived benefits and barriers to implementing polygenic risk scores in primary care: results of a national physician survey

Authors:

J. Vassy^{1,2,3}, B. Kerman⁴, A. A. Lemke⁵, M. Clayman^{6,7}, K. MacIsaac¹, E. Harris¹, A. Antwi⁸, T. Yi⁸, C. A. Brunette⁹; ¹Harvard Med. Sch. at VA Boston Hlth.care System, Boston, MA, ²Brigham and Women's Hosp., Boston, MA, ³Precision Population Hlth., Ariadne Labs, Boston, MA, ⁴Brigham and Womens Hosp., Boston, MA, ⁵Norton Children's Res. Inst., Affiliated with the Univ. of Louisville Sch. of Med., Louisville, KY, ⁶UMass Chan Med. Sch., Worcester, MA, ⁷Ctr. for Hlth.care Organization & Implementation Res., Bedford, MA, ⁸VA Boston Hlth.care System, Boston, MA, ⁹VA Boston Hlth.care System, Mission, KS

Abstract Body:

Background: Implementation of polygenic risk scores (PRS) into routine preventive medicine will depend in part on the clinical utility and barriers primary care physicians (PCPs) perceive to their use. Methods: We used a national online survey to ask PCPs about 1) their perceived utility of PRS for individual preventive actions; 2) their perceived benefits of PRS; and 3) their perceived barriers to PRS implementation. We performed latent class analysis (LCA) of the responses to identify unique subgroups of PCPs based on similarities and differences in response patterns. Results: Among 369 PCP respondents nationally (open rate 10.8%; participation rate 96.3%; completion rate 92.9%), mean (SD) age was 55.1 (13.0) years; 137 (37.3%) were women; mean (SD) time since medical school graduation was 27.3 (13.3) years; 73 (19.9%), 12 (3.2%), and 232 (63.2%) reported Asian, Black/African American, and White race, respectively; and 15 (4.1%) reported Latinx/Hispanic ethnicity. The majority (63%-93%) agreed they would use PRS for all 3 classes of preventive actions queried (disease screening procedures, preventive medications, and lifestyle modification), except for using a low-risk PRS to identify patients who might be able to delay/discontinue lifestyle modification (41% agreeing). Across composite scores associated with the 3 classes of preventive actions, respondents endorsed stronger preference for taking earlier/more intensive action for high-risk patients than with delaying/discontinuing action for low-risk patients (signed rank $p < 0.001$). Respondents most often chose out-of-pocket costs (48%), lack of clinical guidelines (24%), and patient insurance discrimination (22%) as extreme barriers to PRS implementation. LCA identified 3 subclasses of PCP respondents. Sceptics ($n=83$, 22.6%) endorsed less agreement with individual clinical utilities, saw patient anxiety and insurance discrimination as significant barriers, and agreed less often that PRS could help patients make better health decisions. Learners ($n=134$, 36.5%) and believers (150, 40.9%) generally expressed similar levels of agreement that PRS had utility for preventive actions and that PRS could be useful for patient decision-making. Compared with believers, however, learners perceived greater barriers to the clinical use of PRS and were most likely to see the lack of clinical guidelines for PRS as a significant barrier. Conclusion: Variation in how PCPs perceive the benefits and barriers for the clinical implementation of PRS suggests tailoring implementation strategies to specific needs and concerns.

Complex Traits Posters - Thursday

PB1569. Performance of externally developed polygenic risk scores in the All of Us Research Program Database

Authors:

L. Hull¹, B. Truong², P. Natarajan¹; ¹Massachusetts Gen. Hosp., Boston, MA, ²Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract Body:

Background: Polygenic risk scores (PRS) have been developed and validated in diverse datasets globally. The All of Us (AoU) Research Program is a novel biobank recruiting diverse U.S. adults. We described the performance of PRS in AoU, using coronary artery disease (CAD), celiac disease, and body mass index (BMI) as examples.

Methods: The Polygenic Score (PGS) Catalog contains a repository of PRS developed to date, including metadata allowing for these scores to be replicated. We computed 26 CAD scores, 10 celiac disease scores and 22 BMI scores for the AoU participants with genetic data available in the V5 data release. We evaluated common predicted genetic ancestries available in AoU with probability higher than 90% and the PGS Catalog including European, African, East Asian and Latino/Admixed American.

For each ancestry, we performed quality control for the genotype data in AoU, we removed variants with genotype rate <95%, Hardy-Weinberg equilibrium p-value <10⁻¹⁴ and minor allele frequency <1%. CAD and celiac cases were defined using diagnostic codes; CAD cases also included those with relevant procedural codes or self-reported personal history of MI or CAD. BMI came from direct measurements of non-pregnant participants. The prediction accuracy was estimated by Nagelkerke R² for disease phenotypes and partial R² for continuous phenotypes adjusted for age, sex and 10 principal components of ancestry.

Results: The prediction performance for CAD and celiac disease, respectively, were 0.026 (95% CI 0.02-0.031) and 1x10⁻⁴ (95% CI 0 - 2.8x10⁻⁴) for European, 0.006 (95% CI 0.001-0.012) and 9.2x10⁻⁵ (95% CI 0 - 2.4x10⁻⁴) for African, 0.019 (95% CI 0-0.044) and 7.4x10⁻³ (95% CI 4.1x10⁻⁴ - 1.4x10⁻²) East Asian, and 0.015 (95% CI 0.006-0.025) and 1.74x10⁻⁴ (95% CI 0 - 5.3x10⁻⁴) for Latino/Admixed American. For body mass index, we observed that the partial R² for European, African, East Asian, and Latino/Admixed American were 0.118 (95% CI 0.113-0.123), 0.022 (95% CI 0.018-0.026), 0.07 (95% CI 0.054-0.087) and 0.075 (95% CI 0.067-0.082), respectively.

Conclusion: We highlighted the utility of PGS catalog to estimate PRS in a novel independent sample, demonstrating a variety of performances by different methods. Benchmarking PRS allows for direct comparison of performance measures in diverse populations.

Complex Traits Posters - Wednesday

PB1570. Phenome-wide analyses with nonsynonymous variants in *SOS2* demonstrate remarkable pleiotropic associations in the UK Biobank

Authors:

S. Bruse, D. Jakubosky, G. McInnes, B. Gonzalez, Z. Romer, N. Karra, A. Leverentz, B. Cajés, O. Gottesman; Empirico Inc., San Diego, CA

Abstract Body:

Large-scale biobanks linked with rich phenotypic information enable the unbiased evaluation of associations between sequence variants and thousands of phenotypes. Such phenome-wide association studies (PheWAS) can improve our understanding of the biological functions of a given gene and facilitate the prioritization of potential therapeutic targets by revealing pleiotropic associations across comorbidities and with diseases previously considered distinct. Utilizing data from 452,401 individuals in the UK Biobank, we identified two independent missense variants (P191R, AAF=0.01 and L183F, AAF=0.001) located in the same domain of the *SOS2* protein, that individually and collectively demonstrate protective associations across a broad range of cardiometabolic diseases and traits in a PheWAS. A burden test aggregating carriers of these two variants is significantly associated with decreased urate ($p=6.17E-14$, $\beta=-0.009$), increased eGFR ($p=3.94E-12$, $\beta=0.066$), decreased systolic ($p=4.82E-18$, $\beta=-0.006$) and diastolic ($p=8.82E-20$, $\beta=-0.006$) blood pressure, decreased ALT ($p=8.99E-24$, $\beta=-0.019$) and AST ($p=1.35E-31$, $\beta=-0.013$), decreased BMI ($p=3.13E-08$, $\beta=-0.050$) and body fat % ($p=5.57E-12$, $\beta=-0.007$), and with decreased HbA1C ($p=3.63E-12$, $\beta=-0.003$). We observe several significant, suggestive and nominal protective associations with related clinical diagnoses, including gout, chronic kidney disease, hypertension, cerebrovascular disease, non-alcoholic fatty liver disease and type 2 diabetes. The *SOS2* burden was also associated with protection from a range of ophthalmic phenotypes, including decreased intraocular pressure, decreased risk of glaucoma and decreased risk of macular degeneration. Region-wide conditional analyses confirmed that P191R is the lead variant across the multitude of associations observed. *SOS2* is a ubiquitously-expressed guanine nucleotide exchange factor that activates small GTPases such as RAS and RHO, implicating multiple signal transduction pathways that are plausibly related to these pleiotropic associations. Based on traits characteristic of Noonan syndrome (caused by *SOS2* gain-of-function mutations) and on *Sos2* KO mouse phenotypes, we hypothesize that P191R and L183F result in *SOS2* loss-of-function. Ongoing work seeks to experimentally characterize the effects of these variants on *SOS2* function, downstream signal transduction pathways and disease pathophysiology; with the aim of mimicking these protective genetic effects with a therapeutic molecule. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Complex Traits Posters - Thursday

PB1571*. Phenome-wide association studies of deleterious variants in the Han Taiwanese people

Authors:

W-Y. Ko¹, Y-H. Io¹, Y-H. Yeh¹, M-L. Kang¹, H-H. Lee¹, W-C. Tseng², J-H. Loo³, D-C. Tarn²; ¹Natl. Yang Ming Chiao Tung Univ., Taipei, Taiwan, ²Taipei Veterans Gen. Hosp., Taipei, Taiwan, ³Mackay Mem. Hosp., New Taipei City, Taiwan, Taiwan

Abstract Body:

The success of precision medicine relies on a comprehensive collection of interpreted genetic profiles and other health records. However, ~80% of genome-wide association studies of common diseases focused on the populations of European ancestry. It is uncertain to what extent can these findings be translated into the healthcare system for the non-European populations. Here, we examined the allele frequency distributions (AFD) of single nucleotide variants (SNVs) in 23,991 genomes of the Han Taiwanese people for a total of 14,555,421 imputed SNVs that were pre-grouped based on their CADD (Combined Annotation-Dependent Depletion) scores. We next performed phenome-wide association analyses (with 27 cardiovascular metabolic-related traits) for a subset of SNVs (36,575) that were identified as deleterious mutations by comparing their AFD with those of synonymous and nonsynonymous mutations. As a result, 161 significant associations in 110 SNVs were identified in 24 traits. Among them, 84 SNVs are newly identified from this study. In particular, we have identified six variants with strong effect sizes ($|\beta| > 0.7$), located at *APOB*, *OR51L1*, *SLC22A12*, *MAP3K11*, *PGAP6* and *LDLR*. Among them, rs144467873 and rs200990725 are associated with cardiovascular-related traits, rs147341630 and rs375498857 are associated with hematologic traits, and rs200104135 and rs202052634 are associated with chronic kidney-related traits. Our results facilitate genetic profiling of several common disease susceptibilities particularly for the populations of Han ancestry.

Complex Traits Posters - Wednesday

PB1572. Phenome-Wide Association Study (PheWAS) investigates the role of long-read discovered structural variants in cardiovascular disease risk in the UK Biobank using an imputation framework

Authors:

A. Basile¹, M. Byrska-Bishop¹, W. Clarke^{1,2}, U. S. Evani¹, A. Corvelo¹, J. Ebler³, T. Marschall³, G. Narzisi¹, M. Zody¹; ¹New York Genome Ctr., New York, NY, ²Outlier Informatics Inc, Saskatoon, SK, Canada, ³Heinrich Heine Univ., Düsseldorf, Germany

Abstract Body:

Structural variants (SVs) are chromosomal rearrangements larger than 50bp and include deletions, insertions, duplications and inversions. SVs account for a 3-10 times greater genetic difference between individual human genomes than single nucleotide variants (SNVs) and are hypothesized to have higher impact on gene function, regulation and phenotypic change. SVs are 3-fold more likely to associate with a genome-wide association study signal as compared to SNVs. SV characterization has lagged behind other genetic variation due to prohibitive sequencing costs and challenges in discovery and genotyping. While costs have dropped, reference imputation panels still provide an effective and commonly used means of increasing variant density and diversity in an array study. However, the majority of reference panels are restricted to SNVs. Herein, we use an imputation framework to investigate the cardiovascular disease burden of common SVs, which have largely been missed by traditional SNV-based association studies. First, we generated an imputation panel using previously released SNV ($n > 59,000,000$), INDEL ($n > 650,000$) and SV ($n > 20,000$) calls from the 1000 Genomes Project (1kGP) for 3,202 samples. SNV genotype calls were generated with GATK HaplotypeCaller using the high coverage Illumina 1kGP WGS data, and SVs/INDELs were genotyped in the same short-read data using variants discovered in 34 long-read sequenced samples with PanGenie. PanGenie is a graph-based method that improves short-read genotyping by leveraging assembly-based reference haplotypes and sequencing reads and achieves better genotype concordance compared to mapping-based approaches. To generate an integrated panel, we merged high-quality SV/SNV/INDEL calls and performed phasing using SHAPEIT. This panel was then used to impute high-quality genotype array data in 487,908 UK Biobank samples using Eagle2 and Minimac4. Imputation performance evaluations with an SV truth set show that SVs are imputed with an accuracy comparable to SNVs at common allele frequencies, $AF > 5\%$, but with lower accuracy at rare sites. Finally, we present results from a cardiac trait PheWAS examining the association between well-imputed SVs ($r^2 > 0.3$) and cardiovascular disease using SAIGE. Assessed cardiac traits include ischemic heart disease, myocardial infarction, atrial fibrillation, cholesterol levels and others. Overall, this work presents a framework for assessing a historically under-studied form of genetic variation. We demonstrate the accuracy of imputing SVs from a diverse panel and assess the implications of structural variation on cardiovascular disease risk.

Complex Traits Posters - Thursday

PB1573*. Phenome-wide perspective into the role of Finnish Y chromosome variation in complex diseases

Authors:

A. Preussner¹, J. Leinonen¹, S. Luo¹, P. Palta¹, FinnGen, M. Pirinen^{1,2,3}, T. Tukiainen¹; ¹Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, ²Dept. of Publ. Hlth., Faculty of Med., Univ. of Helsinki, Helsinki, Finland, ³Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland

Abstract Body:

The Y chromosome (chrY) is routinely excluded from genetic association studies due to the peculiar biology and analytical challenges specific to chrY. Consequently, potential impacts of chrY genetic variation on complex disease remains largely uncharacterized. To broadly investigate the impacts of chrY variation on complex disease, we examined FinnGen project data (R7) for 72,198 Finnish male samples, genotyped on arrays with sufficient number of chrY markers. We studied associations for 517 diverse health registry endpoints, both on the chrY haplogroup-level (including Finnish main haplogroups N1a1, I1, R1a, and R1b) as well as using other available chrY variants (N=347 after quality control), with MAF ranging from 7.6e-5 to 0.39. At the haplogroup-level, we observed 35 suggestive associations ($p < 0.05/4$), of which 16 involved cardiovascular diseases. These included risk-increasing effects for haplogroup I1 and cardiovascular diseases, such as myocardial infarction (MI) (OR=1.10, $p=0.006$), an association which replicated in a meta-analysis with the UK Biobank (OR=1.07, $p=0.0026$) regardless of the vast differences in haplogroup distributions between Finland and the UK. In the variant-level analysis, we observed one significant association ($p < 0.05/27/517$) to glaucoma (OR=1.58, $p=3.0e-6$) for a low-frequency variant correlating with haplogroup I1 ($R^2=0.12$). Moreover, we observed 345 suggestive associations ($p < 0.05/27$) for various disease categories. To further understand the contribution of the chrY variation to complex disease, we used these data and MI as an example to estimate the expected number of discoveries for chrY. Although with our sample size we were well-powered to detect effects for GWAS top loci, the small size of chrY (estimated 27 independent loci in our data) makes it unlikely to discover significant associations for MI or any other single endpoint, assuming the same density of associations and effect sizes in chrY as in autosomes. Given the 517 endpoints examined, we extrapolate we nevertheless should have detected at least five chrY associations in the PheWAS. As we find fewer, it is possible that the potential phenotype effects in chrY are smaller and require larger sample sizes to be detected. We are therefore extending the study to FinnGen R9 (N>100,000 males) in the coming months. These results highlight that chrY may harbor genetic variation related to complex disease risk in Finland, yet validation in larger data sets and with further chrY variants (particularly related to N1a1) are required. Overall, our work serves as the first step towards comprehensive understanding of the role of chrY variation in complex diseases.

Complex Traits Posters - Thursday

PB1574. Phenome-wide PGS portability in the Colorado Center for Personalized Medicine biobank suggests overlooked challenges in diverse populations

Authors:

M. Lin¹, C. H. Arehart¹, N. Rafaels¹, K. Crooks¹, N. Pozdeyev¹, A. Hendricks², S. Raghavan¹, C. R. Gignoux¹, CCPM Clinical PRSUIT Working Group; ¹Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ²Univ. of Colorado Denver, Denver, CO

Abstract Body:

The potential of polygenic scores (PGS) to improve health-related personalized risk prediction is highly promising. Yet, their transportability across populations remains largely poor. With increasingly powered association results available as an abundant resource for PGS, limited studies have empirically and comprehensively examined the landscape of current PGS transportability across a phenome-wide range of conditions. Here, we leverage >600 scores from the PGS Catalog to assess systematically the performance of PGS prediction on >1k electronic health record (EHR) phenotypes in the Colorado Center for Personalized Medicine Biobank, totaling 661,265 predictions. To analyze heterogeneity across populations we stratified 33,863 individuals from six genetically defined ancestry groups (European, African/African American, Hispanic/Latino, East Asian, South Asian, Middle Eastern). Some diseases are found reasonably well predicted by PGSs together with demographic covariates, such as type 2 diabetes (T2D) and hypertension ($P=2.7e-170$ and $1.1e-164$, $AUC=0.77$ and 0.81), with the top scoring ~4% and ~3% of individuals in the biobank having $OR>3$, respectively. However, the majority of predictions have considerable cross-group heterogeneity in performance (average $I^2=0.18$ in phecode~PGS pairs with $FDR<0.1$). This can greatly affect potential use in personalized medicine as well as bias downstream interpretation in such frameworks as Mendelian Randomization. Additionally, we found that the choice of PGS unit of measure, when comparing association statistics per SD or using the top decile against the remainder, yielded highly discordant estimates of heterogeneity ($r=0.29$). Using multilevel nested mixed models, we found primary influences on heterogeneity include the distribution of disease prevalence between ancestry groups in the test cohort, in addition to other features of both the test and training sets, such as methods of score development and allele frequency distribution in CCPM. Our results suggest that there are important consequences of applying PGS to downstream applications that assume causality that only stem from homogeneous predictions across ancestry groups. We describe some of these efforts particularly in the space of metabolic syndrome phenotypes across ancestry groups. The overlooked attributions can be determined empirically and are due to both characteristics of the training and test settings. This provides a description of hopes and pitfalls in the current efforts of applying PGS resources to diverse populations.

Complex Traits Posters - Thursday

PB1575. PLA2G6 associated late onset Parkinson's disease in a Sudanese family: A case report.

Authors:

Y. Bakhit^{1,2}, K. Eltom³, O. Seidi³, Sudan Neuroscience Projects (SNPs)- Parkinson study group, U. Wuellner¹; ¹Univeristy of Bonn, Bonn, Germany, ²Univ. of Khartoum, Khartoum, Sudan, ³Univeristy of Khartoum, Khartoum, Sudan

Abstract Body:

Phospholipase A2 gene encode a class of enzyme that catalyse the hydrolytic release of fatty acids from phospholipids. It is expressed in the Brain and its protein product; iPLA2- β is abundant in the nerve terminals and dendrites of the brain. In its early stage, PLA2G6 associated Autosoma recessive early onset Parkinson's disease patients exhibit Bradykinesia and lower limb tremor, hypomimia and gait disturbance. In this study, two siblings were born to first-degree consanguineous parents were recruited. Symptoms appeared first of 58 and 60 years, respectfully and started as a resting tremor in his right arm, which regressed in three years and became bilateral, with a generalized slowness of movement and rigidity. Saliva samples were donated for targeted next generation sequencing. We found two compound heterozygous mutations in PLA2G6; onr deletion (NM_003560: c.2070_2072del) and one missense (NM_003560: c.956C>T), where the in-frame mutation is known to be pathogenic and is associated with Infantile Neuroaxonal dystrophy. This is considered the first case of late onset Parkinson's disease associated with PLA2G6, adding to the growing complex phenotypes of PLA2G6 associated neurodegeneration.

Complex Traits Posters - Wednesday

PB1576. Polycystic Ovarian Syndrome Physiologic Pathways Implicated Through Clustering of Genetic Loci.

Authors:

M. Stamou¹, K. Smith¹, H. Kim¹, R. Balasubramanian¹, K. Gray², M. Udler¹; ¹Massachusetts Gen. Hosp., Boston, MA, ²Brigham and Women's Hosp., Boston, MA

Abstract Body:

Introduction: Polycystic ovarian syndrome (PCOS) is a phenotypically and genetically heterogeneous disorder. Despite several genomic loci associated with PCOS, the relationship of these loci to disease pathogenesis remains largely unknown. In this study, we utilized soft clustering to identify genetic pathways of PCOS. **Methods:** To identify novel genetic clusters, variant-trait association clustering was performed for 51 PCOS-associated genetic variants and 50 GWAS traits using bNMF clustering. We then generated PCOS partitioned polygenic scores (pPS) using the strongest weighted loci in each cluster and tested for associations with clinical features in participants of the Mass General Brigham Biobank [MGBB, N=28,282 (14,855 female and 13,427 male)]. Additionally, associations with metabolic/cardiovascular clinical outcomes were assessed using GWAS for type 2 diabetes/T2D (DIAMANTE, N=898,130) and coronary artery disease/CAD (CARDIOGRAM/UKBB, N=547,261). **Results: Four variant-trait clusters of PCOS were identified with three having clear physiological mechanisms:** (i) Cluster 1/**obesity** (top loci include *FTO*, *ERBB3*); (ii) Cluster 2/**hormonal** changes (top loci include *FSHB*, *WT1*); (iii) Cluster 3/**inflammatory** markers (top loci include *ATXN2/SH2B3*); (iv) Cluster 4/**Unknown** (top loci include *MAF*, *SLC38A11*). The 51-SNP full pPS was associated with increased risk of PCOS (OR 1.17, p=0.003) in MGBB females. **Cluster pPS's were associated with distinct clinical traits in MGBB participants:** Cluster 1 with increased BMI (p=3.20x10⁻⁷), triglycerides (p=0.003), and systolic blood pressure (p=0.009); Cluster 2 with reduced FSH levels (p=0.009); and Cluster 3 with increased platelet count (p=0.017). Finally, **PCOS genetic clusters were associated with disease outcomes in GWAS and MGBB:** Cluster 1 pPS with increased the risk of T2D (OR 1.04, p=8.48 x 10⁻³⁰) and CAD (OR 1.01, p=0.012) in GWAS, with replication of the former in MGBB (OR 1.11, p=8.99x10⁻⁶). In contrast, Cluster 2 pPS was associated with decreased T2D risk (OR 0.98, p-value=3.37x10⁻⁴) in GWAS, with directional replication in MGBB (OR 0.98, p=0.4). **Conclusions:** Our analysis identified distinct genetic clusters with PCOS-associated biologic pathways. Cluster 1 pPS was associated with increased the likelihood of T2D and CAD, highlighting the genetic predisposition of patients with PCOS and obesity to cardiometabolic disease. In contrast, pPS linked to hormonal changes led to a decreased risk for T2D, suggestive of the protective nature of these genetic loci. Thus, distinct genetic backgrounds in PCOS individuals may underlie the clinical heterogeneity and long-term outcomes in PCOS.

Complex Traits Posters - Thursday

PB1577*. Polygenic burden of major depressive disorder and suicidality in children of diverse ancestries: a US-population-based study

Authors:

P. Lee^{1,2}, M. D. Silberstein³, J-Y. Jung⁴, T. Ge^{1,2}, R. C. Kessler², R. Perlis^{1,2}, J. W. Smoller^{1,2}, M. Fava^{1,2}; ¹Massachusetts Gen. Hosp., Boston, MA, ²Harvard Med. Sch., Boston, MA, ³Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, ⁴Stanford Univ., Stanford, CA

Abstract Body:

Background: The increasing rate of suicide among racial/ethnic minority children is a serious public health concern in the US. The aggravating role of several environmental factors, including structural racism, has become evident. In contrast, little is known of whether and if so, how genetic risk contributes to increasing suicide risk among children. Here we examined whether polygenic risk of major depressive disorder (MDD) is associated with suicidality among US children, especially those in underrepresented racial/ethnic groups. **Methods:** Our cross-sectional study investigated year 2 data from the Adolescent Brain and Cognitive Development (ABCD) study. Lifetime experiences of suicidality, including suicidal ideation, planning, and attempts, were assessed using the Kiddie Schedule for Affective Disorder and Schizophrenia. Children's ancestry-specific polygenic risk scores (PRSs) for MDD were calculated using a newly developed Bayesian PRS construction method, PRS-CSx, and four independent GWASs representing diverse ancestral groups (total sample N=1,186,500). Logistic regression analysis was used to examine the association of genetic risk and suicidality. **Results:** After rigorous quality control of genotype data, we identified 1,675 ABCD participants of African ancestry (AA), 1,630 of Hispanic ancestry (HA) and 4,344 European ancestry (EA). All children were un-related (total N=7,649, mean age 11.920.61, 47.33% female). Suicidality was most common among AA children (N=392, 24.0%), followed by EA (N=1,004, 23.1%) and HA (N=367, 21.9%). After multiple testing correction, we found statistically significant associations between MDD PRSs and suicidality in EA children (OR=1.2, 95% CI=1.12-1.29, p -value= 3.40×10^{-7} , FDR $q=3.06 \times 10^{-6}$), with notably stronger genetic effects among females compared to males (Nagelkerke's R^2 1.62% vs. 0.50%). We also found a statistically significant association between MDD PRSs and suicidality in female HA children (OR=1.26, 95% CI=1.05-1.51, $R^2=1.30\%$, p -value= 1.33×10^{-2} , $q=2.99 \times 10^{-2}$). These associations were mediated in part by children's internalizing problems, and remained significant after adjusting for children's puberty, sociodemographic backgrounds, and family history of mental health problems. **Conclusion:** In the largest genetic sample of US children, we observed a significant genetic association between MDD and childhood suicidality, particularly among females of EA and HA ancestry. Along with thorough ethical considerations, future efforts are warranted to investigate the clinical utility of various risk factors, including genetic PRSs, for mitigating suicide risk in children.

Complex Traits Posters - Wednesday

PB1578. Polygenic risk and complex trait prediction for East Asians using the Taiwan Precision Medicine Initiative and Taiwan Biobank datasets (N = 500k).

Authors:

P-Y. Kwok^{1,2}, C-Y. Chou¹, T. G. Raben³, **C-H. Chang**⁴, E-C. Yeh¹, E. Widen³, C-H. Chen¹, S. Hsu³; ¹Inst. of BioMed. Sci., Academia Sinica, Taipei, Taiwan, ²Cardiovascular Res. Inst. and Inst. for Human Genetics, Univ. of California, San Francisco, San Francisco, CA, ³Michigan State Univ., East Lansing, MI, ⁴Dept. of Med., Univ. of California San Francisco, San Francisco, CA

Abstract Body:

The Taiwan Precision Medicine Initiative (TPMI) and the Taiwan Biobank (TWB) are nation-wide projects which have, in aggregate, recruited, phenotyped, and genotyped (using a custom Thermo Fisher Scientific Axiom SNP array) over 500k individuals of Han Chinese ancestry, and plan eventually to have a sample size in the millions. This is the largest non-European cohort for common disease risk prediction. Here, we introduce the TPMI and TWB projects and report preliminary results on polygenic risk prediction of major diseases and complex traits. PRS were trained using TPMI and TWB data. The strength of prediction is validated in holdback samples from this population, using sibling pairs as well as unrelated individuals. Further validation is performed using Chinese ancestry individuals from UK Biobank and in distant ancestry groups such as Europeans. We examine conditions such as Asthma, Gout, Hypertension, CAD, Hyperlipidemia, Diabetes, and more. We report AUCs above 0.6, using only genetic information, which are comparable to those obtained from state-of-the-art PRS for Europeans trained and validated in European populations. Incorporating other risk factors (age, sex, etc.) leads to AUCs above 0.75. The top 3-5% PRS lead to a 5x odds ratio for some conditions and rival even large effects from single gene mutations. These are the first examples of parity between PRS in European and non-European populations, and hence represent an important advance toward Equity and Diversity in genomic science. We also discuss the performance of Taiwan-trained PRS in distant populations and compare with the European-trained case. We explore the similarities and differences in genetic architecture between Taiwan- and European-trained PRS. This study serves as a model and establishes a path for PRS development in non-European populations.

Complex Traits Posters - Thursday

PB1579. Polygenic risk burden is associated with early psychosis and endophenotypes of psychotic disorders

Authors:

T. Warren¹, J. D. Tubbs², M. B. Corona¹, P. Singh³, V. Zarubin³, S. Morse⁴, T. A. Lesh³, C. S. Carter³, P. C. Sham², A. S. Nord¹; ¹UC Davis Ctr. for NeuroSci., Davis, CA, ²Dept. of Psychiatry, Univ. of Hong Kong, Hong Kong, Hong Kong, ³UC Davis Imaging Res. Ctr., Sacramento, CA, ⁴Washington Univ. in St. Louis, St. Louis, MO

Abstract Body:

Understanding the links between genetic risk, neurological phenotypes, and clinical symptom presentation is a primary goal for psychiatric genetics researchers and has motivated efforts such as the NIMH's Research Domains Criteria initiative. To this end, we conducted a study on polygenic risk in patients being treated at the UC Davis Medical Center for early psychosis who had undergone psychiatric and neurological phenotyping as part of an ongoing research effort. For this study, we included 225 cases (22% female, average age 19.6) from this cohort who progressed to a diagnosis of a psychotic disorder, as well as 126 matched controls (42% female, average age 19.7) with no neurological or psychiatric diagnoses. Both cases and controls represented the population demographics of the area served by the UC Davis Medical Center. For all subjects, we analyzed DNA as well as global functioning, IQ, and task-based fMRI data. For psychosis subjects, we also looked at clinical ratings for positive symptoms, negative symptoms, and cognitive functioning. DNA was extracted from blood and genotyped on the Illumina Infinium PsychArray-24 kit, which tests for 271,000 variants and is enriched for psychiatric markers. Using the lasso method, we calculated schizophrenia (SZ) polygenic scores (PGSs) based on recent GWAS data from the Psychiatric Genomics Consortium. SZ PGS explained a significant portion of the variance in psychosis case versus control status in our cohort. SZ PGS was also significantly associated with IQ in all subjects and negative symptoms in bipolar disorder (BD) subjects. fMRI association analyses are ongoing. As a secondary, exploratory outcome, we defined pathways of genes involved in different neurotransmitter systems and calculated PGSs for these pathways. In these preliminary data, GABA PGS was associated with case versus control status and with IQ in psychosis cases. Glutamate PGS was associated with IQ in psychosis cases and with negative symptoms in BD subjects. Our findings extend PGS as a risk factor for early first episode psychosis, illuminate links between genetic burden and psychotic disorder endophenotypes, and build a case for further study of pathway-specific polygenic burden in the context of psychiatric symptoms.

Complex Traits Posters - Wednesday

PB1580. Polygenic risk for body mass index is associated with atypical antipsychotic induced weight gain.

Authors:

A. Justice¹, G. Chittoor¹, N. Josyula², R. A. J. Smit³, R. J. F. Loos³, R. Birnbaum^{4,5}, K. E. North⁶, K. L. Young⁶, O. A. Andreassen^{7,8}, A. E. Locke⁹, G. Nicol¹⁰, GIANT Consortium, DiscovEHR Collaboration, ORAcLE Consortium; ¹Dept. of Population Hlth.Sci., Geisinger, Danville, PA, ²Dept. of Population Hlth.Sci., Geisinger, Fremont, CA, ³The Charles Bronfman Inst. for Personalized Med., Mount Sinai Sch. of Med., New York, NY, ⁴Psychiatry, Mount Sinai Sch. of Med., New York, NY, ⁵Genetics and Genomic Sci., Mount Sinai Sch. of Med., New York, NY, ⁶Dept. of Epidemiology, Univ. of North Carolina, Chapel Hill, NC, ⁷NORMENT Ctr., Div. of Mental Hlth.and Addiction, Oslo Univ. Hosp., Oslo, Norway, ⁸Inst. of Clinical Med., Univ. of Oslo, Oslo, Norway, ⁹Regeneron Genetics Ctr., Tarrytown, NY, ¹⁰Dept. of Psychiatry, Washington Univ. Sch. of Med., St. Louis, MO

Abstract Body:

Obesity and its metabolic consequences contribute substantially to accelerated aging and reduced lifespan in psychiatrically ill individuals. The risk for antipsychotic induced weight gain (AIWG) is thought to be highly genetic, but our understanding of these contributors remains limited. Characterizing genetic variation influencing AIWG has the potential to identify those at greatest risk for weight gain and ensuing cardiometabolic dysregulation. Thus, we examined the relevance of polygenic risk scores (PRS) for cross-sectional body mass index (BMI) for predicting AIWG, measured as change in BMI, in the MyCode study population. We derived a PRS using the current interim GWAS of BMI by the GIANT consortium including >1.1 million European participants. PRS weights were estimated in SBayesR. The final PRS included 1,080,060 variants. To define AIWG, we restricted BMI measurements to those assessed during use of atypical antipsychotic medications, and used a linear mixed model where BMI was regressed with intercept and time (years) as both fixed and random effects to derive the slope of AIWG, stratified by sex and ancestry. Analyses were conducted for overall change in BMI and also restricted to gainers (i.e. positive AIWG slope). Associations between standardized PRS and AIWG were done using generalized linear models while accounting for families using GEE. All models were adjusted for baseline age, age², sex, baseline BMI, follow-up duration, and five principal components of ancestry. Incremental R² values were estimated in an unrelated subset of participants using linear regression. We also evaluated the association between PRS and baseline BMI. We identified 8,493 participants (97% European ancestry, 68% female, 61% gainers) with ≥ two BMI measures, with a mean (SD) of 33 observations (58), 3.7 yrs (3.7) of follow-up, and AIWG of 0.18 BMI/yr (0.93). The PRS was associated with baseline BMI (P=1e⁻³⁰⁵, β=4.03, R²=14.4%). We identified a significant association between the cross-sectional BMI PRS and overall AIWG (R²=0.3%, P=3.9e⁻⁸, β=0.069 BMI units/yr) and AIWG in gainers only (R²=0.3%, P=6.5e⁻⁶, β=0.051 BMI units/yr). Also, we observed a small, but significant difference (0.125 BMI/yr, P=2e⁻⁴) in adjusted mean AIWG between the lowest and upper quintiles of the PRS. Our study demonstrates that PRS derived from BMI GWAS can explain variation in AIWG. The finding that PRS is significantly associated with AIWG, shows a clear heritable component of this important side effect. While the gain in % variance explained in AIWG by the cross-sectional BMI PRS was small, it shows that genetic variation inducing AIWG exclusive of overall BMI may improve AIWG prediction.

Complex Traits Posters - Thursday

PB1581. Polygenic Risk score based phenome wide association study for Tourette Syndrome

Authors:

P. R. Jain¹, A. Topaloudi¹, T. Miller-Fleming², L. Davis², P. Paschou¹; ¹Purdue Univ., West Lafayette, IN, ²Vanderbilt Univ Med Ctr., Nashville, TN

Abstract Body:

Tourette Syndrome (TS) is a complex neurodevelopmental disorder characterized by vocal and motor tics lasting more than a year. It is highly polygenic in nature with studies identifying numerous genes and variants associated with the disorder. Epidemiological studies have shown TS to be correlated with various phenotypes and environmental factors but large-scale phenome wide analyses in biobank level data have not been performed for the disorder. Polygenic risk score (PRS) based Phenome Wide Association studies (PheWAS) have become an increasingly common method to identify different factors associated with the genetic risk of a complex diseases without requiring a prior assumption. In this study, we use the summary statistics from the latest meta analysis of TS to calculate the PRS of individuals in the UK Biobank data. We then use a PheWAS approach determine the association of disease risk with a wide range of phenotypes ranging from physical and mental health to lifestyle and sociodemographic factors. A total of 57 traits were found to be significantly associated with the Polygenic risk of TS risk. Multiple psychosocial factors and mental health conditions like anxiety disorder and depressive episode were significantly associated with TS risk. Additionally, significant associations were observed with other complex disorders like Type 2 Diabetes, heart palpitations and respiratory conditions. Cross disorder comparison with other Neurodevelopmental disorders like ADHD, ASD and OCD indicated an overlap in associations between TS and these disorders. ADHD and ASD had a similar direction of effect with TS while OCD had an opposite direction of effect for all traits except mental health factors. Sex specific PheWAS analysis identified differences in the associations with TS between males and females. Type 2 Diabetes and heart palpitations are significantly associated with TS risk in males and not in females, whereas diseases of the respiratory system are associated with TS risk in females and not in males. The results of these analyses identified multiple factors associated with the genetic risk of TS and sheds light on the shared genetic and phenotypic architecture of the different Neurodevelopmental disorders.

Complex Traits Posters - Wednesday

PB1582. Polygenic risk scores affected by different strategies of selection of relevant genomic positions and equations used for calculations

Authors:

J. Radvanszky^{1,2,3}, Z. Pös^{1,3}, M. Hlavacka⁴, A. Hurtuk⁵, O. Kubicka⁵, I. Lojova^{1,2}, J. Minarik², A. Zatkova¹, L. Kadasi^{1,2}, J. Gazdarica^{3,6}, J. Budis^{3,6}, T. Szemes^{3,2}; ¹Inst. of Clinical and Translational Res., BioMed. Res. Ctr. of the Slovak Academy of Sci., Bratislava, Slovakia, ²Dept. of Molecular Biology, Faculty of Natural Sci., Comenius Univ., Bratislava, Slovakia, ³Comenius Univ. Sci. Park, Bratislava, Slovakia, ⁴Dept. of Applied Informatics, Faculty of Mathematics, Physics and Informatics, Comenius Univ., Bratislava, Slovakia, ⁵Inst. of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies of Slovak Univ. of Technology in Bratislava, Bratislava, Slovakia, ⁶Slovak Ctr. of Scientific and Technical Information, Bratislava, Slovakia

Abstract Body:

After welding the ever increasing datasets of large scale genome-wide association studies (GWAS) for complex phenotypes with genomics options offered by massively parallel sequencing, especially whole-genome sequencing (WGS), studies describing polygenic risk score (PRS) calculations for different diseases started to appear rapidly. Effective and reliable PRS calculations, however, still pose several challenges, specifically when considering the vast variability and complexity of GWAS datasets and possible equations. We aimed, therefore, to characterize some of the practical challenges for potential implementation of PRS. We used whole-genome sequencing (WGS) data of 52 individuals, consisting of patients having ulcerative colitis (UC/IBD), their family members and control samples. To obtain final lists of genomic positions (GP-lists) for PRS, potentially relevant genomic positions associated with UC/IBD were extracted from the EMBL-EBI GWAS Catalog and further filtered using an in-house application according to different parameters. We generated several GP-lists, divisible into three main types: 1) Based on individual GWAS studies, extracting positions from single studies included in the GWAS Catalog; 2) Aggregated results of several GWAS studies generated by merging all the studies meeting the keywords of the disease/trait of interest (while these were further filtered using different variables to reach higher diversity of alternative lists); 3) Disease specific and previously validated PRS lists, which were obtained directly from relevant publications. PRS calculations were performed using five different equations (non-weighted sum, linear weighted model with OR, linear weighted model with logarithmic OR, PLINK and PRSice-2) for all of these lists. Results of our work proved that PRSs represent highly potent applications in genomics. On the other hand, their implementation poses certain challenges, which require not only further studies, but also careful standardization and increased awareness when evaluating the reliability of PRS outputs. Since many variables affect the PRS results, and many different data sources and filtering steps are necessary to find the best results when standardizing PRS calculations, automated but flexible algorithms may be very helpful, at least until standardized and validated PRS for each relevant disease/trait will be developed. Financial support: Slovak Research and Development Agency (APVV-18-0319), Scientific Grant Agency (VEGA_2/0167/20, VEGA_2/0040/20) and by the OP Integrated Infrastructure co-financed by the European Regional Development Fund (ITMS: 313011V578, 313011W428).

Complex Traits Posters - Thursday

PB1583. Polygenic risk scores for quantitative estimation of complex vitamin intake needs to lower homocysteine in the adult population: from population based to personalized dietary intake recommendations.

Authors:

J. Hager¹, N. Zubair², E. Scuccimarra¹, F. Salesses¹, K. Heim³, P. Kamani³, N. Morris²; ¹Nestle Inst. of Hlth.Sci., Lausanne, Switzerland, ²Nestle Hlth.Sci. - Atrium Innovations, Bridgewater, NJ, ³Nestle Hlth.Sci. - Atrium Innovations, Bridgewater, NJ

Abstract Body:

Cardiovascular disease (CVD) is one of the most frequent non-communicable diseases in the world. In the UK more than 20% of the adult population have at least one CVD related morbidity. Biochemical risk factors for CVD include high LDL, triglyceride and homocysteine levels. A large meta-analysis showed that a decrease in homocysteine under 11 mmol/l decreases the risk for ischemic heart disease and stroke by 16% and 24% respectively. Homocysteine is a metabolite of the 1-carbon cycle. Vitamins B9, B12, B6 and choline all play a vital role in 1-carbon metabolism and homocysteine homeostasis and supplementation with these vitamins has been shown to decrease homocysteine levels. However, supplementation trials have reported variable results in respect to both the degree of homocysteine lowering as well as CVD outcomes. One reason may be the poor correlation between vitamin intake and vitamin blood levels. Recent GWA studies have shown that genetic factors have a significant effect on the blood concentrations of all four of these vitamins as well as homocysteine. To better understand the impact of genetic factors on the variability of homocysteine levels in relation to vitamin intake, we created a polygenic risk score for homocysteine from GWAS catalog summary statistics that included 31 independent SNPs after LD pruning. The effect size for the PRS was estimated in an independent data set of 2000 subjects with homocysteine measurements and was significantly associated with homocysteine levels ($p=6.65E-11$). The observed PRS in the sample ranged from 17 to 44. Each additional risk allele resulted in an average increase of homocysteine plasma levels of 0.12 mmol/l, meaning that individuals in the high-risk group (34-44 risk alleles), show on average 4.7 mmol/l higher homocysteine levels than the general population (7.84 mmol/l). We then established an algorithmic dose-response model for the intake of vitamins B9, B12, B6 and choline on homocysteine levels conditioned on the PRS. A linear dose response regression model showed that 41% of the homocysteine plasma concentration variance is explained by age, sex, choline, folate, vitamin B6 and B12 intakes. The rest being attributable to the genetic terms and other non-identified co-factors. In summary, a polygenic risk score explains a significant part of the variability of homocysteine levels. The dose - response model shows that individuals with a high PRS have significantly increased intake needs to achieve homocysteine levels compared to low PRS individuals.

Complex Traits Posters - Wednesday

PB1584. Polygenic risk scores identify heterogeneity in disease diagnosis and treatment in asthma and COPD.

Authors:

M. Moll¹, J. Sordillo², A. Ghosh³, L. HAYDEN^{4,5}, G. McDermott¹, M. McGeachie¹, A. Dahlin¹, A. Tiwari⁴, A. Saferali⁶, S. Begum¹, J. Ziniti⁴, A. Gulsvik⁷, P. Bakke⁷, C. Irribarren⁸, C. Hersh⁴, J. Sparks¹, B. Hobbs⁴, J. Lasky-Su⁹, E. Silverman¹⁰, S. Weiss⁹, A. Wu², M. Cho¹¹; ¹Brigham & Women's Hosp., Boston, MA, ²Harvard Med. Sch., Boston, MA, ³SUNY Upstate Med. Ctr., Syracuse, NY, ⁴Brigham and Women's Hosp., Boston, MA, ⁵Boston Children's Hosp., Boston, MA, ⁶Brigham & Women's Hosp., Boston, MA, ⁷Univ. of Bergen, Bergen, Norway, ⁸Kaiser Permanente Northern California, Oakland, CA, ⁹Brigham & Women's Hosp, Boston, MA, ¹⁰Brigham & Women's Hosp, Boston, MA, ¹¹Brigham and Women's Hosp., Boston, MA

Abstract Body:

Rationale: Asthma and chronic obstructive pulmonary disease (COPD) have distinct and overlapping genetic and clinical features. Clinically, the diagnoses may co-occur as asthma-COPD overlap (ACO). We hypothesized that separate polygenic risk scores (PRSs) for asthma and COPD, constructed from recent large genome-wide association studies, would demonstrate variable associations with asthma and COPD diagnoses and treatments across research and clinical cohorts. **Methods:** We developed a novel asthma PRS (derived from clinical definitions) and used a previously-published COPD PRS (derived from spirometry), and we applied these PRSs to two research (COPDGene and CAMP, with available spirometry) and clinical (MGB Biobank and GERA; diagnosis using billing codes) studies. We assessed the association of each PRS with COPD and asthma diagnosis, ACO (defined as COPD subjects with a history of asthma diagnosed by a physician before 40 years of age), lung function trajectories, asthma exacerbations, and medication use. Models were adjusted for age, sex, smoking variables (when available), and genetic ancestry. **Results:** In 102,260 participants, the PRS_{Asthma} was weakly correlated with the PRS_{COPD} (mean $r = 0.0043$, $p=0.01$). In meta-analyses, the PRS_{Asthma} was associated with asthma (OR 1.16 [95% CI: 1.13 to 1.18]) and the PRS_{COPD} was associated with COPD (OR 1.25 [95% CI: 1.21 to 1.30]). However, these results were inconsistent across cohorts, with no association of the PRS_{COPD} with billing-code-defined COPD in clinical cohorts (GERA and MGB). In COPDGene non-Hispanic whites, the PRS_{COPD} was associated with ACO (OR 2.2 [95% CI: 2.0 to 2.5]) and asthmatics (diagnosed at age <40 years old) who developed moderate-to-severe airflow obstruction and emphysema (OR 1.59 [95% CI: 1.15 to 2.2]). In GERA asthmatics, both PRS_{COPD} and PRS_{Asthma} were associated with asthma exacerbations (Caucasians: OR 1.18 and OR 1.1, respectively). In GERA White asthmatics, the PRS_{Asthma} was associated with greater inhaled corticosteroid use (+4 days/year/SD) and the PRS_{COPD} with greater long-acting beta agonist use (+4.5 days/year/SD). Results in non-European populations were similar but with smaller effect sizes. **Conclusions:** Asthma and COPD polygenic risk scores predict disease status, asthma-COPD overlap, and medication use; however, disease associations differ in research versus clinical cohorts. These differences may be due to differences in disease pathophysiology or clinical / administrative practices.

Complex Traits Posters - Thursday

PB1585. Polygenic score for height influences adolescent growth spurt in Southwestern Indigenous Americans.

Authors:

M. Ramirez Luzuriaga, S. Kobes, L. Wedekind, W-C. Hsueh, L. Baier, W. Knowler, R. L. Hanson; NIDDK, Phoenix, AZ

Abstract Body:

A recent meta-analysis of genome-wide association studies (GWAS) from the Genetic Investigation of ANthropometric Traits (GIANT) consortium and UK Biobank ($N \sim 700,000$) identified 3290 single nucleotide polymorphisms (SNPs) independently associated with adult height. However, it is unclear how these variants influence linear growth during adolescence. Moreover, most of the meta-analysis data come from European-ancestry cohorts with little representation of Indigenous Peoples. The current study used anthropometric and genotypic data from a longitudinal study of diabetes conducted in an Indigenous community in Arizona between 1965-2007. Biological parameters of adolescent growth (i.e., height, velocity, and timing of growth spurt) were derived from the Preece-Baines growth model, a parametric growth curve fitted to longitudinal height data, in 787 participants with height measurements spanning the whole period of growth. We examined the association of growth spurt parameters with a polygenic score for height derived from 2612 SNPs identified in the GIANT/UK Biobank meta-analysis for which genotypic data were available for the Indigenous study population. The correlation between polygenic score and growth parameters was analyzed with adjustment for sex, birth year, maternal diabetes, and the first 5 genetic principal components (to account for population structure) in a mixed model that accounted for relationships among individuals. We found moderate correlations of the height polygenic score with adult height ($r=0.25$, $p=3.2 \times 10^{-13}$), height at take-off ($r=0.20$, $p=6.3 \times 10^{-8}$), and height at peak velocity ($r=0.25$, $p=5.5 \times 10^{-12}$). Correlations of the polygenic score with velocity at take-off ($r=0.12$, $p=8.9 \times 10^{-4}$) and peak height velocity ($r=0.13$, $p=3.2 \times 10^{-4}$) were significant ($p < 0.05$) but weaker. We found no correlation of the polygenic score with age at take-off ($r=0.00$, $p=0.97$), age at peak height velocity ($r=-0.02$, $p=0.44$), or age at maturation ($r=-0.03$, $p=0.41$). To assess potential influences on fetal growth, we also analyzed the association of the height polygenic score with birth weight in 3700 individuals from the Indigenous population with available data; the correlation with birthweight was modest but significant ($r=0.05$, $p=2.9 \times 10^{-3}$). Our findings suggest that genetic variants associated with adult height influence the magnitude and velocity of height growth that occurs before and during the adolescent growth spurt, with little influence on the timing of growth spurt. These findings provide additional information on how polygenic determinants of height influence linear growth.

Complex Traits Posters - Wednesday

PB1586. Polygenic score prediction of atrial fibrillation following cardiac surgery: a retrospective hospital biobank study.

Authors:

A. Jeuken^{1,2}, É. Goulet¹, J. Cadrin-Tourigny^{1,2}, A. Petzl^{1,2}, D. Busseuil¹, L. Rivard¹, J-C. Tardif¹, R. Tadros^{1,2}; ¹Montreal Heart Inst., Montreal, QC, Canada, ²Université de Montréal, Montreal, QC, Canada

Abstract Body:

Background: Atrial fibrillation (AF) is the most common complication following heart surgery. Post-operative AF (POAF) prolongs hospital stay and is associated with increased risk of stroke and death. Identifying individuals at risk of POAF and AF-related stroke is an unmet need, considering the existence of simple interventions mitigating such risk. **Objective:** To assess the predictive ability of a validated AF polygenic score (PGS-AF) for POAF following cardiac surgery and remote recurrence of AF in individuals with POAF. **Methods:** We included participants in the Montreal Heart Institute biobank that underwent cardiac surgery from 2005 to 2018 and had no AF prior to surgery. POAF was defined as AF occurring in up to 30 days post surgery. Array genotyping was performed using the Global Screening Array followed by imputation using the TOPMed reference panel. PGS-AF was computed using weights from a published PGS (catalog ID: PGS000016) and adjusted for genotypic principal components. The association of PGS-AF with POAF was assessed using logistic regression with and without correction for known AF clinical predictors including the CHARGE-AF clinical score. The association of PGS-AF and CHARGE-AF with remote AF recurrence (>30 days after surgery) within the patient subgroup with POAF was assessed using a Cox Proportional Hazards Model. **Results:** We included 1984 patients (82% males; aged 64±10 at cardiac surgery), of which 721 (36%) developed POAF. POAF was more common following valve surgery than other cardiac surgeries (43% vs 32%, P<0.0001) and with increase in the CHARGE-AF score (P<0.0001). PGS-AF was significantly associated with POAF (P<0.0001; 41% increased risk of POAF per SD increase in PGS-AF; C statistic 0.60). In a model combining PGS-AF with clinical predictors (CHARGE-AF and valve surgery), the association of PGS-AF with POAF remained significant (P<0.0001). Compared to clinical prediction only, adding PGS-AF results in significant improvement of the model (C statistic 0.67 vs 0.64; likelihood ratio test P<0.0001). Among the 721 patients with POAF, 138 (19%) developed AF during a median 4.5-year follow-up. PGS-AF and CHARGE-AF were both significantly and independently associated with remote AF recurrence (P<0.05), where each SD increase in PGS-AF increases risk of remote AF by 23%. **Conclusion:** PGS-AF is significantly associated with AF occurring as a complication of cardiac surgery independently from clinical risk predictors with modest discriminative ability. PGS-AF is also associated with remote recurrence of AF. Broadly, these data implicate polygenic risk in complex clinical scenarios such as surgical complications.

Complex Traits Posters - Wednesday

PB1587. Population level genetic screening for coronary artery disease with polygenic risk scores and rare variants in a diverse electronic health record linked biobank

Authors:

K. Boulier, Y. Ding, V. Venkateswaran, R. Johnson, B. Pasaniuc; Univ. of California, Los Angeles, Los Angeles, CA

Abstract Body:

Large-scale genetic association studies have enabled polygenic risk scores (PGS) that estimate genetic risk for disease by aggregating small effects of common variants into a single genetic risk prediction. While promising, few, if any, have been validated for clinical use. Here we evaluate methods of genetic screening for coronary artery disease (CAD) risk assessment in ATLAS, an electronic health record (EHR) linked biobank at UCLA. We focus on CAD, a common, preventable disorder with strong genetic contribution across both monogenic and polygenic architectures. Familial hypercholesterolemia (FH) is a monogenic form of CAD, which affects roughly 1 in 200 individuals. The ATLAS study genotypes selected patients from the UCLA Health system (N=60,000). Focusing on a subset of adult European-ancestry ATLAS patients (N=13,900) with abundant clinical data, we find that a PGS for CAD is correlated with presence or absence of cardiovascular disease. Similarly PGS for LDL is correlated with measured LDL levels (Pearson correlation of 0.14 $p < 0.001$). PGSs performance was reduced in individuals of non European ancestries. Next, we investigate the predictiveness of existing PGS for CAD and relevant biomarkers LDL and lipoprotein A. We find that no PGS improved prediction in the presence of standard clinical variables. CAD PGS z score was statistically significant in Cox modeling but higher PGS scores protected against rather than predicted CAD (HR 0.88, 0.83-0.93). Harrell's C statistic for prediction of CAD was 0.79 (0.77 - 0.81) using clinical variables of age at first visit, gender, prior smoking, hypertension and diabetes by medication, whether or not CAD PGS was included. Next, we evaluate the penetrance of FH in the 250 patients found to have likely pathogenic or pathogenic variants in associated genes (233 in *LDLR*, 22 in *APOE*). Average age was 48.0 years (std 15.1). 60 (24%) had an ICD code for coronary disease compared to 9% of the overall population. Mean maximum LDL in the FH group was 142 (46.5) compared to 119 (41.2) in the general population ($p < 0.001$). Pooled cohort equation categorized 150 or the 178 patients between the ages of 40 and 79 as low or intermediate risk. We find that PGS for CAD is higher in these 250 patients as compared to random controls and investigate the role of PGS in risk stratifying patients with FH. Polygenic and monogenic variant effect sizes are lower in EHR data than reported in registry studies. While rare variants affect a small percentage of the population, their significantly larger effect sizes are more easily measurable with imperfect phenotyping. Combining polygenic and monogenic screening for CAD may capture more patients at elevated risk.

Complex Traits Posters - Thursday

PB1588*. Population-scale proteo-genomic analysis reveals novel loci associated with Parkinson's disease

Authors:

A. Doostparast Torshizi, D. Truong, B. Mautz, S. Li, M. Black; Population Analytics, Data Sci. Analytics & Insights (DSAI) ML/AI Analytics & Insights, Janssen R&D, Spring House, PA

Abstract Body:

Background: Parkinson's disease (PD) has strong genetic components. However, its etiological basis remains largely unclear, posing a challenge in identifying and developing novel, safe, and effective therapies for treatment.

Methods: In UK Biobank (UKB) participants with whole-exome sequencing (WES) and imputed array (IMP) data, we conducted genome-wide variant-/gene-level (collapsing pathogenic, putative loss of function and deleterious variants) analyses for PD using REGENIE v3.0. PD was defined using a combination of ICD9/10 codes, clinician-curated clinical codes and self-report data. Individuals diagnosed with recurrent depression, schizophrenia, or bipolar disorder were excluded from case and controls; controls diagnosed with any nervous system disease were also excluded. Genetic analysis was performed on curated PD cases (n=2859) and controls (n=158876). Significant variants were additionally tested for association with ~1500 proteins assayed from blood plasma samples collected at UKB enrollment as part of the UKB Proteomics Project. Finally, pathway enrichment analysis using KEGG was conducted on the identified associated proteins to specify enriched biological pathways.

Results: Three novel loci were identified by variant-level analysis in IMP including rs56214516 (*CRHRI*), rs7155501 (*GCHI*), and rs429358 (*APOE*) as well as two novel variants in WES including chr4_89936328_G_A (*MMRNI*) and chr17_46031540_A_G (*STH*). Several known variants were replicated including *LRRK2*, *SNCA*, *STK39*, and *KRTCAP2*. Gene-level tests also replicated *LRRK2*, *SNCA*, and *GBA*. Three variants were associated with several proteins: rs7155501 (1 protein), rs429358 (36 proteins) and rs56214516 (26 proteins); the majority of these protein quantitative trait loci (pQTLs) were in *trans*. Pathway enrichment analysis of the associated proteins shows significant enrichment in three pathways including glycosaminoglycan degradation, sphingolipid metabolism, and autophagy-lysosome. These pathways have been previously linked to PD/Parkinsonism pathogenesis through affecting α -synuclein aggregation, loss of synaptic plasticity, and neurodegeneration.

Conclusion: We identified several novel variant associations with PD and found these variants to be pQTLs for proteins enriched in multiple PD-relevant pathways. Our study provides novel insights into the genetic drivers of PD as well as their functional implications at the protein level.

Complex Traits Posters - Wednesday

PB1589. Portability of a multiethnic polygenic risk score for low-density lipoprotein cholesterol in a Samoan population

Authors:

J. Carlson¹, M. Krishnan², S. Liu¹, S. Rosenthal³, N. Hawley⁴, H. Cheng⁵, T. Naseri⁶, M. Reupena⁷, S. Viali⁸, R. Deka⁹, S. McGarvey¹⁰, R. Minster¹, D. Weeks¹¹; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²Univ. of North Carolina, Chapel Hill, NC, ³Univ. of Pittsburgh Sch. of Dental Med., Pittsburgh, PA, ⁴Yale Sch. of Publ. Hlth., New Haven, CT, ⁵Univ. of Cincinnati Coll. of Med., Cincinnati, OH, ⁶Ministry of Hlth., Apia, Samoa, ⁷Lutia i Puava ae Mapu i Fagalele, Apia, Samoa, ⁸Natl. Univ. of Samoa, Apia, Samoa, ⁹Univ Cincinnati Med Ctr, Cincinnati, OH, ¹⁰Brown Univ., Providence, RI, ¹¹Univ of Pittsburgh, Pittsburgh, PA

Abstract Body:

Polygenic risk scores (PRS) are a promising tool for improving health outcomes through personalized medicine. However, there is a well-documented need to improve diversity in the populations through which PRS are derived, to not exacerbate health disparities among underrepresented populations. Specifically for low-density lipoprotein cholesterol (LDL-C), a known risk factor for cardiovascular disease, recent efforts have been made to diversify PRS to represent non-Europeans, including African, Asian, and Hispanic populations. However, the transferability of these ‘multiethnic’ PRS has not been examined in several other populations, including Pacific Islanders, who have a disproportionate burden of cardiovascular disease and are underrepresented in health research.

Thus, we sought to assess the performance of a published multiethnic PRS for LDL-C, constructed using over 1 million individuals of African, East Asian, European, Hispanic, and South Asian ancestry (Graham et. al 2021), in a cohort of n=2,816 Samoan adults.

Variant information and corresponding weights for the multiethnic LDL-C PRS were downloaded from the Polygenic Score Catalog, and genome-wide data from 2,816 Samoans were aligned to the scoring file to assign individual-level risk scores. Of the 9,009 variants in the PRS, 8,653 (96%) were available for the Samoan participants, although 20% (1,747/8,653) of variants were monomorphic. The distribution of the PRS in Samoans ranged from 43.6 to 49.3 (mean 46.8). PRS utility was evaluated with a linear regression model adjusted for age, sex, and principal components of ancestry. Higher PRS was associated with higher LDL-C ($\beta = 13.7$ mg/dL, 95% CI 12.2 - 15.3 mg/dL, $p = 4.3e-66$). The partial r^2 for the PRS was 10.17% (95% bootstrap CI 8.23 - 12.42%), which is similar to the published PRS performance in other non-European groups (r^2 10 - 16%), despite there being no Pacific Islanders represented in the multiethnic PRS construction.

While there is still much work to do to improve the representation of non-European populations in health research, these results show that, at least for LDL-C, a PRS derived from multiple diverse ancestries does not perform more poorly in Samoans than in other non-European populations. Further work is needed to compare the LDL-C PRS performance to one derived in a Pacific Islander population, and to assess PRS performance for other traits in this population to see if they are equally transferable. However, this work highlights the importance of including several diverse populations in PRS construction as a first step to improving accuracy and transferability to underrepresented populations.

Complex Traits Posters - Thursday

PB1590. Potential effects of SORBS1 variants on serum immunoglobulin E levels specific to milk intake in relation to prevalence of atopic dermatitis in Korean children

Authors:

H-J. Lee¹, E-A. Choi¹, J-H. Do¹, S. Hong², S. Lee², Y-Y. Kim¹; ¹NIH Korea, Cheongju, Korea, Republic of, ²Asan Med. Ctr., Seoul, Korea, Republic of

Abstract Body:

Environmental factors are reported to alter the genetic effects on the allergic diseases through IgE-mediated type-1 hypersensitivity. We aimed to investigate that the association of atopic dermatitis with genetic polymorphisms within *SORBS1* based on milk exposure. The mother-infant pairs were recruited from the Cohort for Childhood Origins of Asthma and Allergic Diseases(COAOA). To identify the variants associated with milk-specific IgE, we performed linear regression using Korean chip genotype data and have found genome-wide significant 5 loci. Among them, we examined the effect of milk-specific IgE levels on SORBS1 polymorphisms, rs143705721 and rs75786636, in 1,281 children. As a result, SORBS1 minor alleles led to a significant increase of total IgE in the higher group of milk-specific IgE level. Also, SORBS1 rs143705721 variant was significantly associated with atopic dermatitis in 1-year children. In addition, we examined the expression of the *SORBS1* gene in a murine model using house dust mite, the most pervasive allergens, observed alterations of the lung. The lung tissue having chronic airway inflammation induced by chronic exposure to house dust mite (*Dermatophagoides pteronissunus*; HDM) extract through the intranasal for 8 weeks in Balb/c mice presented the lower expression of SORBS1. Our finding provides the possibility of the level of SORBS1 expression decreased is related to the risk of atopic dermatitis through food allergen exposure.

Complex Traits Posters - Wednesday

PB1591. Predicted loss-of-function variants for blood lipids in over a million individuals

Authors:

S. Koyama^{1,2,3}, Z. Yu^{1,2,3}, S. Choi¹, D. Klarin⁴, J. E. Huffman³, S. M. Damrauer^{5,6}, K-M. Chang^{5,6}, T. L. Assimes^{4,7}, P. S. TSAO^{4,7}, D. J. Rader⁵, G. M. Peloso^{8,3}, P. T. Ellinor^{1,2}, K. Cho^{3,9}, Y. V. Sun^{10,11}, P. W. F. Wilson^{10,11}, Million Veteran Program, P. Natarajan^{1,2,3}, ¹The Broad Inst. of MIT and Harvard, Cambridge, MA, ²Massachusetts Gen. Hosp., Boston, MA, ³VA Boston Hlth.care System, Boston, MA, ⁴Stanford Univ. Sch. of Med., Palo Alto, CA, ⁵Univ. of Pennsylvania, Philadelphia, PA, ⁶Corporal Michael J. Crescenz Dept. of VA Med. Ctr., Philadelphia, PA, ⁷VA Palo Alto Hlth. care, Palo Alto, CA, ⁸Boston Univ., Boston, MA, ⁹Massachusetts Gen. Brigham and Harvard Med. Sch., Boston, MA, ¹⁰Emory Univ., Atlanta, GA, ¹¹VA Atlanta Hlth.care System, Decatur, GA

Abstract Body:

Background: Predicted loss-of-function (pLOF) variants yield lifelong genetic inactivation permitting powerful human disease inference *in vivo*. Prior studies relating pLOF with clinical traits have had limited power due to low variant frequencies of this highly constrained variant class requiring the aggregation of molecularly heterogeneous variants inherently obscuring individual variant interpretation. However, recent expansion in the sample sizes of sequenced individuals enables us to genotype ultrarare pLOF variants by direct sequencing or high-quality genotype imputation with individual variant phenotype association.

Methods: We associated and meta-analyzed rare pLOF variants with blood lipid traits [LDL cholesterol (LDLC), HDL cholesterol (HDLC), and Triglycerides (TG)] in the densely genotyped population from Million Veterans Program (n = 617,279, imputed to the TOPMed reference panel), and UK Biobank (n = 410,482, whole exome sequenced). **Results:** We tested 42,087 pLOF variants across 12,023 genes. We identified statistically significant association ($P < 1 \times 10^{-6} = 0.05/42,087$ variants tested) between 18 genes (58 pLOF variants) and LDLC, 14 genes (30 pLOF variants) and TG, and 28 genes (59 pLOF variants) and HDLC respectively. pLOF variants were significantly more frequently associated with blood lipid traits than other classes of exonic variants (i.e., missense, synonymous variants). We also identified multiple associated pLOFs in 21 genes (6 genes for LDLC, 9 for HDLC, 6 for TG) with concordant effect directions, which includes 23 pLOF variants in *APOB* associated with LDLC, 12 in *PCSK9* and LDLC, and 12 in *CETP* and HDLC. Each of these variants showed large but varying effect sizes now yielding precise individual variant-level penetrance estimates. Each pLOF variant explained a small portion of phenotypic variance but collectively explained a considerable unexplained phenotypic variance of blood lipid traits. **Conclusion:** Through the pLOF analysis for blood lipids in over a million individuals, we obtained robust estimates for individual pLOF variant effect sizes at the population scale and insights into the genetic architecture of very rare functional variants.

Complex Traits Posters - Thursday

PB1592*. Predicted polygenic transcriptional risk score supports the inference of canalization of polygenic risk of common diseases and traits in the UKBiobank.

Authors:

S. Nagpal, Z. Xu, G. Gibson; Georgia Inst. of Technology, Atlanta, GA

Abstract Body:

Since organisms develop and thrive in the face of constant perturbations due to environmental and genetic variation, species may evolve resilient genetic architectures. We previously sought evidence for this process, known as canalization, through a comparison of the prevalence of phenotypes as a function of the polygenic score (PGS) across environments in the UK Biobank cohort study. Contrasting seven diseases and three categorical phenotypes with respect to 151 exposures in 408,925 people, the deviation between the prevalence-risk curves was observed to increase monotonically with the PGS percentile in one-fifth of the comparisons, suggesting extensive PGS-by-Environment (PGS×E) interaction. After adjustment for the dependency of allelic effect sizes on increased prevalence in the perturbing environment, cases where polygenic influences are greater or lesser than expected are seen to be particularly pervasive for educational attainment, obesity, and metabolic condition type-2 diabetes. Notably, body mass index showed more evidence for decanalization (increased genetic influence at the extremes of polygenic risk), whereas the waist-to-hip ratio showed canalization, reflecting different evolutionary pressures on the architectures of these weight-related traits. An additional 10 % of comparisons showed evidence for an additive shift of prevalence independent of PGS between exposures. Here we show that predicted polygenic transcriptional risk scores (PPTRS) based on the summation of transcriptome-wide association study (TWAS) effects performs comparably and supports the evidence of canalization. For Ulcerative colitis, smoking and some aspects of diet such as bread type were confirmed to exacerbate risk using both static genotype-based polygenic score and rectum-based PPTRS capturing cis-eQTL effects. Across all exposures, the correlation of the observed deviations of risk at the extremes of PGS vs PPTRS is 72%. These results provide the first widespread evidence for canalization using polygenic as well as predicted transcriptional risk score, protecting against disease in humans and have implications for precision medicine as well as understanding the evolution of complex traits.

Complex Traits Posters - Wednesday

PB1593. Prediction of Alzheimer's disease conversion using polygenic risk scores grouped by transcriptome profiles from blood and brain

Authors:

D. Goldstein¹, J. Hu¹, R. Panitch¹, N. Sahelijo¹, Y. Wang¹, R. Au^{2,3,4}, L. A. Farrer^{1,3,5,6}, T. D. Stein^{3,7,8}, G. R. Jun^{1,5,6}; ¹Dept. of Med. (BioMed. Genetics), Boston Univ. Sch. of Med., Boston, MA, ²Dept. of Anatomy & Neurobiology, Boston Univ. Sch. of Med., Boston, MA, ³Dept. of Neurology, Boston Univ. Sch. of Med., Boston, MA, ⁴Dept. of Epidemiology, Boston Univ. Sch. of Med., Boston, MA, ⁵Dept. of Ophthalmology, Boston Univ. Sch. of Med., Boston, MA, ⁶Dept. of Biostatistics, Boston Univ. Sch. of Med., Boston, MA, ⁷Dept. of Pathology & Lab. Med., Boston Univ. Sch. of Med., Boston, MA, ⁸Dept. of VA Med. Ctr., Bedford, MA

Abstract Body:

Introduction: Polygenic risk scores grouped by biologically connected gene networks preserved in brain and blood brain may predict the conversion of clinically normal individuals to Alzheimer's disease (AD). **Methods:** We selected AD associated variants ($p < 10^{-3}$) identified in a large AD GWAS (*Kunkle, 2019*) in genes contained in four previously identified biologically connected gene networks (modules) preserved in blood and brain transcriptome data from AD cases and controls (*Panitch, 2021*). We derived a module based polygenic risk score (mbPRS) for each module using AD associated variants for subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI, $n=1,605$) and Framingham Heart Study (FHS, $n = 8,481$). Individuals were stratified based on the proportion of their mbPRS to all mbPRSs within each study; individuals with mbPRSs less than the first quartile were classified as low risk, while those with mbPRS greater than the third quartile were assigned as high risk, and remaining subjects as undetermined (reference) subjects. We evaluated associations of cognitive and imaging traits using risk status as a binary outcome in ADNI. Conversion rate from normal or mild cognitive impairment to AD was estimated using the Cox proportional hazards model by comparing hazard ratios (HR) between high and reference groups in ADNI and FHS. **Results:** In each of the four modules (M1-M4) in the ADNI dataset, executive function and memory were significantly reduced in the high risk compared to low risk subjects (best $p=3.8 \times 10^{-13}$ with memory in M3 group). Hippocampal volume was significantly reduced in high-risk subjects in the M2-M4 groups (best $p=5.3 \times 10^{-9}$ in M3 group), whereas amyloid pathology measured by PET imaging was significantly increased in high risk subjects in the M2 and M4 groups. The conversion rate to AD was significantly greater in high risk compared to reference subjects from the M3 group in both ADNI (HR=1.74, $p=6.0 \times 10^{-3}$) and FHS (HR=1.63, $p=5.4 \times 10^{-7}$) datasets. In FHS, the conversion rate was also significantly greater among high risk compared to reference subjects in the M2-M4 groups ($p < 10^{-3}$). **Conclusion:** Our results demonstrate for the first time that polygenic risk scores derived from gene co-expression networks preserved in blood and brain can differentiate subjects with high risk of conversion to AD according to biological pathways or mechanisms. These findings also provide insight for precision medicine guided by genomic and endophenotypic profiles.

Complex Traits Posters - Thursday

PB1594. Pre-infection antiviral innate immunity attenuate SARS-CoV-2 infection and viral load in iPSC derived alveolar epithelial cells type-2.

Authors:

S. Kumar¹, J. Curran², J. Granados¹, E. De Leon¹, J. Thomas³, S. Williams-Blangero², J. Blangero²; ¹Univ. of Texas Rio Grande Valley Sch. of Med., McAllen, TX, ²Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX, ³Univ. of Texas Rio Grande Valley Sch. of Med., Edinburg, TX

Abstract Body:

The coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), manifests through a broad spectrum of symptoms and symptom severity; many who were infected remained asymptomatic and in those who developed symptoms, the disease ranged from mild cold like symptoms to severe illness and death. Age, sex, and presence of comorbidities such as cardiovascular disease, diabetes, hypertension, and obesity have been the major risk factors for severe COVID-19. However, a large portion of this heterogeneity in disease susceptibility and severity of illness remains poorly understood. SARS-CoV-2 infects humans through the angiotensin converting enzyme (ACE2) receptor, enriched in nasal epithelial cells, and then spread to the ACE2-expressing cells in the distal lung. A growing body of evidence suggests that alveolar epithelial cell type 2 (AT2), which constitute 10-15% of all alveolar cells and are critical for alveolar homeostasis, are the primary targets of SARS-CoV-2 infection. Damage to AT2s was reported to be positively correlated with poor prognosis in COVID-19 patients. Modeling SARS-CoV-2 infection in-vitro, in well characterized, highly uniform, (r^2 at 95% CI = 0.89 ± 0.03), induced pluripotent stem cell (iPSC) derived AT2s from ten participants of our Mexican American Family Study, showed interindividual variability in SARS-CoV-2 infection susceptibility and in cellular viral load. To understand the underlying mechanistic of the AT2's capacity to regulate SARS-CoV-2 infection and cellular viral load, we performed a systematic characterization of pre- and post- SARS-CoV-2 challenged AT2's transcriptomes. The expression of 760 genes in pre-infection AT2s showed statistically significant (FDR corrected p -value ≤ 0.05 ; r^2 absolute ≥ 0.64) correlation with infection susceptibility and cellular viral load. 307 of these genes, whose expression was negatively correlated, showed significant enrichment (Fisher's exact test p -value ≤ 0.001) in cytokine-mediated signaling pathways, positive regulation of interferon-beta production, positive regulation of the inflammatory response, type I interferon signaling pathway, and MDA-5 signaling pathway gene ontology (GO) terms and included innate immunity genes *IFIH1*, *OAS1*, *OAS3*, *DHX58*, *IRF7*, *ADAM8*, *NMI*, *ADAR*, and *GBP6*, whose expression also significantly upregulated (One way ANOVA, FDR corrected p -value range 5.69×10^{-21} to 8.04×10^{-11}) in post SARS-CoV-2 challenged AT2s. Our results strongly suggest that activated pre-infection innate immunity regulates susceptibility to SARS-CoV-2 infection and post infection cellular viral load in AT2s.

Complex Traits Posters - Wednesday

PB1595*. Prioritization of causal genes at Parkinson's disease associated loci using machine learning.

Authors:

E. Yu^{1,2}, R. Lariviere^{1,2}, R. Thomas^{1,2}, K. Senkevich^{1,2}, E. Fon^{1,2}, Z. Gan-Or^{1,2}; ¹McGill Univ., Montreal, QC, Canada, ²Montreal Neurological Inst.-Hosp. (The Neuro), Montreal, QC, Canada

Abstract Body:

In the most recent Parkinson's disease (PD) genome-wide association study (GWAS), 90 independent genetic variants within 78 loci were reported to be associated with the disease. Because most of these variants have unknown functional evidence, we faced a challenge to uncover the potential mechanism and identify causal genes. Previous studies trained machine learning models to predict causal genes from multiomic data. However, these methods lacked tissue and cell-type specific data that is relevant for each disease. In this study, we leveraged genomic datasets from dopaminergic neurons with other publicly available brain tissue quantitative trait loci (QTL) to train a model specific for Parkinson's disease. We trained a machine learning model to predict PD associated genes from GWAS loci. To train our model, we combined transcriptomic, epigenomic data including single-cell multiomic data by FOUNDIN-PD, dopaminergic neuron expression QTL (eQTL), GTEx brain tissues eQTL, enhance-promoter interactions. To select specific gene candidates, PD variants were fine mapped using echolocator to exclude passenger variants. Variant level information was transformed to a gene-feature matrix. We used loci which contained known PD genes (*GBA*, *GCH1*, *LRRK2*, *MAPT*, *SNCA*, *TMEM175*, *VPS13C*) for cross-validation by labeling the known PD genes as positive and other genes 1Mb upstream and downstream as negative. Across all 78 loci, we nominated the genes with high statistical evidence associated with PD such as *IP6K2*, *BIN3*, *IGSF9B*, *TMEM163* and others. Most of these genes are not extensively studied in PD. This study prioritizes multiple candidate genes at PD GWAS loci for functional studies and therapeutic targets for drug development. We used a comprehensive approach including PD relevant datasets from dopaminergic neurons to improve the accuracy of our model.

Complex Traits Posters - Thursday

PB1596*. Prioritizing genes and gene programs for disease by integrating genetic and perturbation data

Authors:

K. Dey¹, K. Jagadeesh², K. Geiger-Schuller³, B. Eraslan³, W. Zhou⁴, W. Bi⁵, Z. Zhao⁵, C. Frangieh⁶, E. Weeks², X. Jin⁷, R. Xavier², S. Lee⁸, H. Finucane⁴, A. Regev³, A. Price¹; ¹Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ²Broad Inst. of MIT and Harvard, Cambridge, MA, ³Genentech, South San Francisco, CA, ⁴Massachusetts Gen. Hosp., Boston, MA, ⁵Univ. of Michigan, Ann Arbor, MI, ⁶Dana-Farber Cancer Inst., Boston, MA, ⁷Scripps Res., San Diego, CA, ⁸Seoul Natl. Univ., Seoul, Korea, Democratic People's Republic of

Abstract Body:

GWAS have identified thousands of disease-associated variants, but identifying genes and gene programs that impact disease remains challenging. Here, we propose a new method, PolyGene, that assigns a disease score to each gene by integrating GWAS summary statistics, gene-level annotations, and gene-gene interactions to prioritize genes with functional relationships to disease GWAS genes. We applied PolyGene to 28 diseases/traits (mean $N=270K$) and evaluated gene score enrichments for gold-standard gene-disease pairs based on approved drug targets, Mendelian genes and CRISPR-implicated genes. PolyGene attained higher gene score enrichments than other methods (e.g. mean excess enrichment 5.01x higher than MAGMA, $P=8e-32$). For a subset of 7 well-powered UK Biobank blood cell traits, incorporating gene-based association statistics from WES data (PolyGene-S) produced a small but statistically significant further improvement (1.14x, $P=3e-04$ vs. PolyGene). We used PolyGene to evaluate the informativeness for 18 autoimmune diseases and blood cell traits of small gene programs constructed from Perturb-seq data in bone marrow dendritic cells, defined by the top 500 genes with downstream effects on gene expression of knocking out (or knocking down) a target gene. We analyzed 1,030 gene perturbation programs constructed from Perturb-seq of E3 ligase genes and their interacting partners (Geiger-Schuller*, Eraslan* et al. in preparation). We observed higher enrichment for immune-related blood cell traits than RBC/platelet-related traits (5.3x, $P<1e-50$), consistent with the immune function of dendritic cells; PolyGene attained higher gene score enrichments than other methods (e.g. 3.4x vs. MAGMA, $P=2e-12$). The *NFKB1* and *CUL3* gene programs were enriched across all autoimmune diseases (1.5x, $P=2e-06$ and 1.7x, $P=4e-12$), and the *CHFR* gene program was specifically enriched for type 1 diabetes (1.4x, $P=3e-06$). Notably, gene-level PolyGene scores of target genes were significantly correlated with gene score enrichments of their perturbation programs (mean $r=0.23$, $P=4e-05$). We also used PolyGene to evaluate the informativeness for autism of perturbation programs for 35 *de novo* autism risk variant genes in 5 brain cell types, derived from in-vivo Perturb-seq data (Jin et al. 2020 Science). The perturbation programs in astroglia and microglia attained the highest enrichments for autism (mean of 2.4x, $P=5e-07$ vs. mean of 1.2x, $P=3e-04$ in other cell types), consistent with the known role of these two cell types in autism. In conclusion, PolyGene is a powerful approach for prioritizing genes and gene programs for disease to produce biological insights.

Complex Traits Posters - Thursday

PB1598. Profiling the neurobiology underlying brain structure in living human subjects.

Authors:

A. Lund, The Living Brain Project Team; Icahn Sch. of Med., New York, NY

Abstract Body:

While large genome-wide association studies have shown that the structure of the brain is at least partially controlled by genetic variation, the relationship between molecular neurobiology and brain structure has yet to be thoroughly characterized. Many studies have investigated the relationship between molecular neurobiology and neuroimaging, but, due to the uncommon occurrence of profiling of living brains, few have been able to do so in a living population. The Living Brain Project (LBP) was designed to overcome this limitation of human brain research by obtaining samples from the dorsolateral prefrontal cortex (dlPFC) in a large cohort of living subjects undergoing deep brain stimulation. Leveraging clinical multimodal neuroimaging performed on the subjects at the time of surgery and molecular profiling generated from their dlPFC biopsies, we look at 171 living individuals for which we have both neuroimaging and RNAseq data. Using this resource, we provide what is, to our knowledge, the most extensive characterization of the molecular profiles of brain structure. Using differential expression, we identify genes whose levels of expression associate with imaging metrics of the brain in the dlPFC including cortical thickness, volume and area. This analysis significantly contributes to the body of work that attempt to generate a complete picture of the molecular properties that play a role in brain anatomy.

Complex Traits Posters - Wednesday

PB1599. PRS-PheWAS of psychiatric/ behavioral traits to disentangle clinical heterogeneity in Substance Use Disorders.

Authors:

L. Vilar-Ribo^{1,2,3}, J. Cabana-Domínguez^{1,2,3,4}, S. Alemany^{1,2}, L. Arribas^{1,2}, N. Llonga^{1,2}, C. Daigre⁵, L. Grau-Lopez⁵, J. Ramos-Quiroga^{1,2,3}, M. Soler Artigas^{1,2,3,4}, M. Ribasés^{1,2,3,4}; ¹Psychiatric Genetics Unit, Group of Psychiatry, Mental Hlth.and Addiction, Vall d'Hebron Res. Inst. (VHIR), Barcelona, Spain, ²Dept. of Mental Hlth., Hosp. Univ.ri Vall d'Hebron, Barcelona, Spain., Barcelona, Spain, ³BioMed. Network Res. Ctr. on Mental Hlth.(CIBERSAM), Madrid, Spain., Barcelona, Spain, ⁴Dept. of Genetics, Microbiol., and Statistics, Faculty of Biology, Univ.t de Barcelona, Barcelona, Spain., Barcelona, Spain, ⁵Addiction and Dual Diagnosis Unit, Dept. of Psychiatry, Hosp. Univ.ri Vall d'Hebron, Barcelona, Spain, Barcelona, Spain

Abstract Body:

Substance Use Disorders (SUDs) are known to be highly polygenic, and many genetic risk factors have been identified by genome-wide association studies (GWASs). Phenome-wide association studies (PheWAS) have been designed to explore how the genomic profile associates with a range of phenotypes. Individuals with SUD suffer from psychiatric comorbidity, resulting in higher clinical severity, and genetic effects are to a substantial extent pleiotropic across psychiatric diseases. Here, we perform an integrated polygenic risk score (PRS)-PheWAS to explore the association between the genetic liability for 10 psychiatric and behavioral traits with the clinical phenome of SUD, to explain the **clinical heterogeneity** seen in SUD patients. We genotyped 1427 SUD patients, whom underwent deep phenotyping, and selected 41 clinical phenotypes related to: (a) lifetime substance use, (b) psychiatric comorbidity, and (c) sociodemographic and health features. We selected 10 pre-existing GWASs from psychiatric and related traits and constructed PRSs in the SUD cohort, using PRSs, to test their association with the clinical phenotypes. We performed logarithmic transformations and used appropriate regression models depending on the nature of the data. We also used available data on lifetime personal abuse (physical, sexual or emotional) to test for a potential abuse by PRS interaction. We observed that PRS constructed from anxiety, depression, schizophrenia and bipolar disorder, were nominally associated ($P < 0.05$) with phenotypes including, psychiatric family history, number of ambulatory treatments, age at onset of consumption and multiple psychiatric conditions, as well as lower educational level and less employment. Similarly, PRSs constructed from educational attainment and well-being showed a negative association with criminal record, substance use family history and lifetime DSM-IV disorder, among others. Our results also showed that PRSs for depression, schizophrenia and suicide attempt seemed to modulate the effect of lifetime personal abuse on phenotypes related to number of therapeutic interventions, psychiatric family history and mental health problems. We observed that the clinical heterogeneity of SUD can be explained, at least in part, by genetic signatures related to other comorbid disorders and traits. Moreover, the presence of genetic risk variants underlying psychiatric conditions in SUD patients can be indicative of higher clinical severity and comorbidity rates, and unfavorable social status. Despite our limited sample size, we are able to show promising preliminary results that should be further investigated in bigger studies.

Complex Traits Posters - Wednesday

PB1600*. Quantification of race, ethnicity, and genetic ancestry disparities in anti-hypertensive drug efficacy in the All of Us Research Program

Authors:

S. Goleva¹, J. Keaton², D. Schlueter³, T. Cassini⁴, H. Mo⁵, A. Williams⁶, T. Ferrara⁶, O. Stubblefield³, C. Zeng⁶, J. Dai⁶, J. Denny⁷; ¹NIH, Bethesda, MD, ²Med. Genomics and Med. Genetics, Clarksburg, MD, ³Natl. Human Genome Res. Inst., Bethesda, MD, ⁴NIH, Chevy Chase, MD, ⁵NHGRI, Bethesda, MD, ⁶NIH, Bethesda, MD, ⁷NIH, Kensington, MD

Abstract Body:

Hispanic ethnicities and non-White races have been, and continue to be, underrepresented in clinical trials. Pharmacokinetic and pharmacogenetic differences in race, ethnicity, and genetic ancestry remain understudied despite evidence that they affect drug efficacy. Essential hypertension affects 47% of Americans, and several mechanisms can contribute to its development, each of which drugs can target. We developed a generalizable method to determine systemic race, ethnicity, and genetic ancestry disparities in common anti-hypertensive drug efficacy. The NIH All of Us Research Program (AoU) is a longitudinal cohort of 194,420 participants with electronic healthcare record (EHR) data, drug exposures, and self-reported demographics, 98,640 of which have whole genomes sequenced. We used a self-controlled case study design to determine the efficacy of common anti-hypertensive medication using the reduction in systolic blood pressure (SBP) values after medication (delta SBP). We analyzed 16 antihypertensive medications present in AoU, each of which over 500 participants used for >28 days. We predicted African, Admixed American, European, East Asian, South Asian, and Middle Eastern genetic ancestry using genetic principal components analysis. Linear regressions were used to determine whether delta SBP for each medication was significantly different in each race, ethnicity, or genetic ancestry compared to non-Hispanic White participants after adjusting for covariates. We were able to determine the most effective hypertensive drugs within each demographic (e.g. chlorthalidone in Black participants). We were also able to determine drugs which have the most variable effectiveness based on race/ethnicity (e.g. labetalol delta SBP = -3.73 ± 0.76 mmHg in White, -7.87 ± 4.56 mmHg in Asian, -9.01 ± 0.96 mmHg in Black, and -12.07 ± 1.64 mmHg in Hispanic participants). 14 out of 16 anti-hypertensive drugs were less effective in Black and 8 out of 16 were less effective in Hispanic, compared to non-Hispanic White, participants after Bonferroni correction. This provides support for the further need to increase racial- and ethnic-minority representation in clinical research. Results using genetic ancestry data closely mirrored those using race/ethnicity. The diuretic chlorthalidone was most effective at reducing SBP in African ancestry participants (delta SBP = -10.32 mmHg, SE = 1.2). This warrants further examination of underlying genetic variants that may contribute to the pharmacogenomic predictors of drug outcomes in future studies. Together, this data demonstrates the potential of real-world evidence to guide medication therapy selection.

Complex Traits Posters - Wednesday

PB1601. Quantifying differential multiple sclerosis risk via polygenic risk scores in African American and Hispanic cohorts.

Authors:

A. Hernandez¹, M. Davis²; ¹Brigham Young Univ., Provo, UT, ²Brigham Young Univ, Provo, UT

Abstract Body:

Polygenic risk scores (PRSs) have been used as a method to learn about disease risk due to genetics. However, the majority of PRSs have been developed using Genome-Wide Association Studies (GWASs) that overrepresent individuals of European ancestry and underrepresent diverse populations such as individuals of African or Latin ancestry. This suggests that PRSs have limited generalizability to diverse populations, which poses medical concerns and could contribute to health disparities. Previous research has found that autoimmune diseases collectively exhibit ethnic disparities, such as Multiple Sclerosis (MS) which shows that non-Hispanic black patients with MS died at an earlier age and have an increasing mortality trend in comparison to white patients with MS. Calculating accurate PRSs for MS can help accommodate ethnic minorities and contribute to closing existing gaps in health care. We acquired de-identified electronic health records (EHRs) from Vanderbilt University Medical Center BioVU that were genotyped on the Illumina MEGAex platform. The software program *Tractor* was used to facilitate the inclusion of admixed individuals and generate ancestry-specific effect size estimates and boost GWAS power. Using *Tractor*, we replicated the 200 previously identified variants for increased MS risk among Hispanic and African Americans by Beecham, A H et al. as well as the variants rs12373588, rs1335532, and rs4821544 which we previously identified to have statistically significant associations with later age of onset in African Americans. We then calculated PRSs using the formula $PRS_j = \sum^N S_i * G_{ij} / P * M_j$ for African American and Hispanic populations using the software program PLINK. We will analyze the efficiency of the PRSs calculated by conducting a linear regression.

Complex Traits Posters - Thursday

PB1602. Quantifying mitochondrial dysfunction at biobank scale using insights from rare disease

Authors:

R. Gupta^{1,2,3,4}, R. Sharma^{1,2}, P. Surendran^{5,6}, C. Gijavanekar⁷, A. Bloemendal^{3,4}, F. Scaglia^{7,8,9}, S. H. Elsea⁷, A. Butterworth^{5,10,11}, B. M. Neale^{3,4}, V. K. Mootha^{1,2,3}; ¹Dept. of Molecular Biology, Massachusetts Gen. Hosp.; Dept. of Systems Biology, Harvard Med. Sch., Boston, MA, ²Howard Hughes Med. Inst., Chevy Chase, MD, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Analytic and Translational Genetics Unit, Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁵British Heart Fndn. Cardiovascular Epidemiology Unit, Dept. of Publ. Hlth. and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, ⁶Dept. of Haematology, Univ. of Cambridge, Cambridge BioMed. Campus, Puddicombe Way, Cambridge, United Kingdom, ⁷Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, ⁸Texas Children's Hosp., Houston, TX, ⁹BCM-CUHK Ctr. of Med. Genetics, Hong Kong, Hong Kong, ¹⁰British Heart Fndn. Ctr. of Res. Excellence, Univ. of Cambridge; Natl. Inst. for Hlth. Res. Blood and Transplant Res. Unit in Donor Hlth. and Genomics, Univ. of Cambridge, Cambridge, United Kingdom, ¹¹Hlth. Data Res. UK Cambridge, Wellcome Genome Campus and Univ. of Cambridge, Cambridge, United Kingdom

Abstract Body:

Mitochondrial dysfunction, defined as a decline in energy production via oxidative phosphorylation (OXPHOS), has been linked to many common diseases of aging, including type 2 diabetes (T2D) and cardiovascular disease. Though often called a “hallmark” of aging, the etiology of age-related OXPHOS dysfunction is poorly defined. Human genetics provides a natural approach to study OXPHOS dysfunction and could contextualize genetic associations with age-related disease. However, classical measures of OXPHOS dysfunction, such as OXPHOS activity assays and histochemical stains of biopsy material, are low-throughput and impractical to perform at biobank scale.

Here, we created a new measure of OXPHOS dysfunction we call “MitoScore,” defined using plasma metabolite profiles from patients with mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS), a monogenic OXPHOS disease. Using a cohort of 134 patients and healthy controls, we trained an elastic net model for disease status. The resultant linear model, with non-zero contributions from 8 metabolites, produced scores well-correlated to disease severity measures in held-out samples (e.g., $r=0.42$ for urine heteroplasmy) with better performance than any single included metabolite.

We then assessed if MitoScore generalizes to other OXPHOS diseases. In a Leigh syndrome cohort ($N=18$), MitoScore produced perfect case-control separation ($AUC=1$). In a distinct cohort with diverse inborn errors of metabolism ($N\sim 230$), MitoScore highlighted patients with lesions in OXPHOS genes over those with other severe genetic diseases ($AUC=0.8$).

After validation in 3 independent cohorts, we performed a GWAS of MitoScore computed in UK Biobank (UKB; $N\sim 80,000$). We identified 15 genome-wide significant associated loci, 86% of which showed concordant effect sizes in INTERVAL ($N\sim 12,000$). Several nominated genes are biochemically linked to the NADH/NAD⁺ ratio, which often rises with OXPHOS impairment due to a reduced respiratory chain capacity to oxidize NADH. Genetic correlations with MitoScore highlighted T2D-related traits, including T2D ($r_g=0.34$) and sex hormone-binding globulin ($r_g=-0.26$), which is notable as elevated NADH/NAD⁺ has recently been linked to insulin resistance.

We expect that MitoScore will be useful in many contexts. For rare disease, MitoScore may be useful in aiding diagnosis (e.g., interpreting variants of uncertain significance) and monitoring disease progression and treatment response. In the general population, we anticipate using MitoScore-associated loci as a genetic instrument to test the causal link between OXPHOS dysfunction and age-related disease.

Complex Traits Posters - Thursday

PB1604. Rare copy number variants in *CACNA1H* implicated in essential tremor.

Authors:

M. Medeiros^{1,2}, C. Liao^{1,2}, D. Spiegelman^{1,2}, P. A. Dion^{1,2}, G. A. Rouleau^{1,2}; ¹McGill, Montreal, QC, Canada, ²Montreal Neurological Inst., Montreal, QC, Canada

Abstract Body:

Background: Essential tremor (ET) is a neurological condition and the most common movement disorder in the world, affecting approximately 5% of the population by the age of 65¹. ET leads to significant impairment and diminished quality of life due to tremors occurring in the hands during voluntary motion (action tremors)¹. Despite ET having an overall heritability of between 45-90%, only a limited number of associated genomic loci have been identified². In the past, linkage disequilibrium, sequencing, and genome-wide association studies were done to identify candidate ET genes; however, copy number variants (CNVs) have been studied less extensively in ET³. Here we leveraged a large ET cohort of 2,133 cases and 10,336 controls to identify genomic deletions or duplications of genes (CNVs) associated with ET.

Methods: The sample cohort was limited to those with European ancestry inferred through PCA. CNVs were called from SNP array data using the program PennCNV and annotated for functional consequence and population frequency using gnomAD. Only rare CNVs (gnomAD frequency less than 1%) intersecting protein coding regions of the genome were tested. The enrichment of CNVs per gene between cases and controls was investigated using Fisher's exact test, and only the candidate genes that passed Bonferroni correction were considered.

Results: Following quality control, 3,516 CNVs across 1,638 genes were tested across 1,365 cases and 9,585 controls. Of the genes tested, 17 were significantly different between cases and controls. One of these genes, the Calcium Voltage-Gated Channel Subunit Alpha 1H (*CACNA1H*), was overlapped by 18 unique CNVs. Genomic duplications overlapping *CACNA1H* were exclusively observed in ET cases and were not observed in any controls ($p = 7.1 \times 10^{-9}$). Approximately 1.3% of the cases in our cohort carried a rare CNV spanning the protein-coding regions of *CACNA1H*.

Discussion: Previously, different calcium channel subunits (*CACNA1G*, *CACNA1C*, *CACNA1A*) have been implicated in ET³. The observation of duplications of the *CACNA1H* gene implicates this gene as a risk factor for ET, supplements the involvement of calcium channels in ET and adds to the involvement of *CACNA* genes in ET pathology.

References: 1. 2001. Louis, E. D., *N. Engl. J. Med.*; 2. 2016. Muller, S. H. et al., *Brain*; 3. 2021. Agúndez, J.A.G et al., *Pharmaceuticals*

Complex Traits Posters - Wednesday

PB1605. Rare copy number variation in the GR@ACE/DEGESCO dementia dataset of Spanish population.

Authors:

I. de Rojas^{1,2}, GR@ACE, DEGESCO, G. Garcia-Ribas³, P. Sánchez Juan^{4,2}, J. Pérez-Tur^{5,2}, P. Pastor⁶, M. Bullido^{7,2}, V. Álvarez^{8,9}, L. Real¹⁰, P. Mir^{11,2}, G. Piñol-Ripoll^{12,13}, J. García-Alberca^{14,2}, M. Medina^{15,4}, M. Sáez¹⁶, Á. Carracedo^{17,18}, L. Tàrraga^{1,2}, M. Boada^{1,2}, S. van der Lee^{19,20,21}, A. Ruiz^{1,2}; ¹Res. Ctr. and Memory Clinic. Ace Alzheimer Ctr. Barcelona – Univ. Internacional de Catalunya, Spain., Barcelona, Spain, ²CIBERNED, Network Ctr. for BioMed. Res. in Neurodegenerative Diseases, NIH Carlos III, Madrid, Spain, Madrid, Spain, ³Hosp. Univ.rio Ramon y Cajal, IRYCIS, Madrid, Madrid, Spain, ⁴CIEN Fndn./Queen Sofia Fndn. Alzheimer Ctr., Madrid, Spain, ⁵Unitat de Genètica Molecular, Inst. de Biomedicina de València-CSIC, Valencia, Spain, ⁶Unit of Neurodegenerative diseases, Dept. of Neurology, Univ. Hosp. Germans Trias i Pujol, Badalona, Barcelona, Spain, ⁷Centro de Biología Molecular Severo Ochoa (UAM-CSIC), Madrid, Spain, ⁸Laboratorio de Genética. Hosp. Univ.rio Central de Asturias, Oviedo, Spain, ⁹Inst. de Investigación Sanitaria del Principado de Asturias (ISPA), Asturias, Spain, ¹⁰Unidad Clínica de Enfermedades Infecciosas y Microbiología. Hosp. Univ.rio de Valme, Sevilla, Spain, ¹¹Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología. Inst. de Biomedicina de Sevilla (IBiS), Hosp. Univ.rio Virgen del Rocío/CSIC/Univ. de Sevilla, Sevilla, Spain, ¹²Unitat Trastorns Cognitius, Hosp. Univ.ri Santa Maria de Lleida, Lleida, Spain, ¹³Inst. de Recerca Biomedica de Lleida (IRBLLeida), Lleida, Spain, ¹⁴Alzheimer Res. Ctr. & Memory Clinic, Andalusian Inst. for NeuroSci., Málaga, Spain, ¹⁵CIBERNED, Network Ctr. for BioMed. Res. in Neurodegenerative Diseases, NIH Carlos III, Madrid, Spain, ¹⁶CAEBI, Centro Andaluz de Estudios Bioinformáticos, Sevilla, Spain, ¹⁷Grupo de Med. Xenómica, Centro Natl. de Genotipado (CEGEN-PRB3-ISCIH). Univ.e de Santiago de Compostela, Santiago de Compostela, Spain, ¹⁸Fundación Pública Galega de Med. Xenómica- CIBERER-IDIS, Santiago de Compostela, Spain, ¹⁹Alzheimer Ctr. Amsterdam, Neurology, Vrije Univ.it Amsterdam, Amsterdam UMC location VUmc, Amsterdam, Netherlands, ²⁰Amsterdam NeuroSci., Neurodegeneration, Amsterdam, Netherlands, ²¹Section Genomics of neurodegenerative Diseases and Aging, Dept. of Clinical Genetics, VU Univ. Med. Ctr., Amsterdam, Netherlands

Abstract Body:

Recent studies have found that duplications or deletions of DNA fragments, known as copy number variants (CNVs), may play a role in missing heritability for complex human diseases. In that sense, we conducted a scan for CNVs in the GR@ACE/DEGESCO dementia dataset of Spanish population (n=20,080 individuals using Axiom 815K Spanish biobank array (Thermo Fisher)). We ran CNV calling algorithms from PennCNV software to obtain high-confidence calls for CNVs. After extensive quality control (QC) for individuals (sex discrepancies, excess of heterozygosity, high missingness, familiar relations and population outliers were excluded), CNVs (>50kb and nSNPs>10 were included), removing spurious CNVs in telomeric/centromeric regions, gene QC (ANOVA comparisons between DNA origin source evaluated in controls), 8,275 controls and 7,818 dementia cases were selected for following analysis. Global burden analyses revealed highly significant differences between dementia cases and controls (dem/ctrl) in CNV rate ($p < 2.2e-16$, $mean_{ctrl-dem} = 6.31-5.23$), distribution of deletions or duplications (both $p < 2.2e-16$; duplications $mean_{ctrl-dem} = 2.22-2.42$; deletions $mean_{ctrl-dem} = 4.10-2.81$), but not in number of genes affected by CNVs ($p = 0.1085$). We also observed a nominal deletion-gene-affected association by genes related to nervous system development pathway or intellectual disability such as *VPS13B* (nine cases affected), *PKP3* ($Freq_{ctrl-dem} = 0.30-0.45\%$) and *FBRSL1* ($Freq_{ctrl-dem} = 0.27-0.47\%$). The current study did not detect significant differences in CNVs that affect known Alzheimer's disease (AD) loci identified by recent genome-wide association studies ($p = 0.128$). Nevertheless, we found three

ASHG 2022 Annual Meeting Poster Abstracts

important AD genes with CNVs potentially associated to dementia. Specifically, we detected *MAPT* protective deletions ($\text{Freq}_{\text{ctrl-dem}} 0.16\text{-}0.01\%$), *ABCA7* risk deletions ($\text{Freq}_{\text{ctrl-dem}} 0.05\text{-}0.14\%$) and *APP* CNVs detected in five dementia cases (0.064%). Because the technology used in our study has limitations in detecting small CNVs, future studies must carefully assess the presence of smaller CNVs and their relationship with dementia.

Complex Traits Posters - Thursday

PB1606. Rare *Plasmodium falciparum* coronin gene mutations following ACT treatment of malaria in South Western Nigeria

Authors:

O. Ajibaye¹, I. C. Nwankwo¹, Y. A. Olukosi¹, O. V. Adaramoye², C. Nnam¹, E. Oriero³, M. Oboh³, B. A. Iwalokun¹, A. Amambua-Ngwa³; ¹Nigerian Inst. of Med. Res., Lagos, Nigeria, ²Coll. of Med., Univ. of Lagos, Lagos, Nigeria, ³Med. Res. Council, The Gambia, Fajara, Banjul, Gambia

Abstract Body:

Plasmodium falciparum Non-*Pfkelch13* protein variants have been implicated in Artemisinin (ART) resistance, but not in Africa. The specific genetic markers driving in vivo reduced ART efficacy among African *P. falciparum* populations are currently unclear. Here we investigated SNPs in *Plasmodium falciparum* actin-binding protein (Pfcoronin) associated with in vivo reduced sensitivity to ART in Nigeria.

Seven isolates showing parasitaemia after Day 3 in a 28-day therapeutic efficacy study of artemether-lumefantrine among 51 volunteers in Lagos, Nigeria were investigated. Molecular diagnosis was done by conventional and real-time PCR amplification of Pf18S rRNA gene, var acidic terminal sequence, telomere-associated repetitive elements-2 and coupled conventional and real-time Pf18S rRNA PCR. High resolution melting (HRM) and 12 neutral *P. falciparum* microsatellite loci genotyping were analyzed to confirm recrudescence in comparison with *msp2* genotyping. We genotyped drug resistance targets (DHFR_51, DHFR_59, DHFR_108, DHFR_164, MDR1_86, MDR1_184, DHPS_581 and DHPS_613) and sequenced *PfCoronin* and *PfKelch13* bi-directionally to investigate presence and association of SNPs with ART insensitivity.

The Real-Time PCR methods equally detected *P. falciparum* infections. A total of 7 (26.92%) cases were either early treatment failure (ETF), late parasitological failure (LPF) or late clinical failure (LCF), and 18 (73.08%) showed adequate clinical and parasitological response (ACPR). Only one sample was confirmed as recrudescence by microsatellites and HRM analysis out of the four identified as recrudescence infections by *msp2* genotyping. Presence of the drug resistance-associated haplotypes, pfdhfr/pfdhps/pfmdr1 (108T/N/51I/164L/59R/581G/86Y/184F) was observed in two samples. Expected heterozygosity (*He*), allelic diversity, for each of the microsatellite loci from pre- and post-drug administration, revealed no significant difference in *He*, ($P = 0.19$, Mann-Whitney test). Analysis of the allele sizes and frequency per locus implicated one isolate as recrudescence. LD measured as a standardized association index *IAS* between multiple *P. falciparum* loci revealed significant LD ($IAS = 0.2865$, $P = 0.02$, Monte Carlo simulation). Genetic analysis identified 7 new Pfcoronin single nucleotide variants (V55L, V67E, I68G, K69G, L77I, D154Y and E200Q) and the P76S earlier reported. None of the ART resistance-associated mutations earlier reported in *pfkelch-13* and Pfcoronin was observed. Pfcoronin mutations here reported are important for investigations on mechanisms of non-*keclh13* markers of emerging African ART resistance phenotypes.

Complex Traits Posters - Wednesday

PB1607. Rare protein-truncating DNA variants in APOB or PCSK9, low-density lipoprotein cholesterol, and risk of coronary artery disease.

Authors:

J. Dron^{1,2}, TOPMed Lipids Working Group; ¹Massachusetts Gen. Hosp., Boston, MA, ²Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract Body:

Introduction: Protein-truncating variants (PTVs) in either apolipoprotein B (*APOB*) or proprotein convertase subtilisin/kexin type 9 (*PCSK9*) are associated with significantly lower low-density lipoprotein (LDL) cholesterol concentrations. Using data from prospective cohort studies, we quantified the relationship between PTVs in *APOB* and *PCSK9*, LDL cholesterol concentrations and protection against coronary heart disease (CHD).

Methods: We considered participants in five prospective cohorts from the NHLBI Trans-Omics for Precision Medicine (TOPMed) program and the UK Biobank. PTVs were defined as nonsense, frameshift, and splice-site variants disrupting *APOB* or *PCSK9*. The impact of PTVs on LDL cholesterol levels was assessed using linear regression, and the hazard ratio (HR) for CHD between PTV carriers and non-carriers was estimated using a Cox proportional hazard model. Models were adjusted for age, sex and the first five principal components of ancestry. Age-dependent probabilities for cumulative incidence of CHD in carriers and non-carriers were determined by summing all events by age at most recent follow-up. The cumulative incidence of CHD by age 75 years between carriers and non-carriers was assessed using a standardized Cox proportional hazard model, adjusted for sex and the first five principal components of ancestry.

Results: From the NHLBI cohorts (N=19,073; 44.4% male; mean [SD] age of 52 [17] years; 67.0% white, 23.9% Black), PTVs were identified in 139 (0.7%) participants and were associated with a 49 mg/dL (95% CI 43-56) reduction in LDL cholesterol. Over a median follow-up of 21.5 years, incident CHD was observed in 12 carriers (8.6%) versus 3,029 non-carriers (16.0%), corresponding to an HR of 0.51 (95% CI 0.28-0.89). From the UK Biobank (N=190,464; 45.0% male; mean [SD] age of 58 [8] years; 93.9% white, 1.6% Black), a PTV was identified in 662 (0.4%) participants and were associated with a 45 mg/dL (95% CI 42-47) reduction in LDL cholesterol. By age 75, estimated cumulative exposure to LDL cholesterol was 31.6% lower in carriers, and the estimated CHD risk was 3.7% (95% CI 2.0%-5.3%) in carriers compared to 7.0% (95% CI 6.9%-7.2%) in non-carriers, corresponding to an HR of 0.51 (95% CI 0.32-0.81).

Conclusion: Results of this large-scale genetic association study confirm and extend prior cross-sectional analyses in identifying that a PTV in either *APOB* or *PCSK9*—owing to significantly decreased exposure to LDL cholesterol—is associated with a 49% reduction in risk of CHD.

Complex Traits Posters - Thursday

PB1608*. Rare protein-truncating variants in *ZNF518A* are associated with shorter female reproductive lifespan

Authors:

S. Shekari^{1,2}, S. Stankovic³, E. J. Gardner³, N. D. L. Owens¹, G. Hawkes¹, K. A. Kentistou³, A. Azad⁴, R. N. Beaumont¹, F. R. Day³, Y. Zhao³, A. R. Wood¹, K. Ong³, C. N. Wright¹, E. R. Hoffmann⁴, K. S. Ruth¹, J. R. B. Perry³, A. Murray¹; ¹Univ. of Exeter Med. Sch., Exeter, United Kingdom, ²Univ. of Queensland Publ. Hlth.Sch., Brisbane, Australia, ³MRC Epidemiology Unit, Wellcome–MRC Inst. of Metabolic Sci., Univ. of Cambridge, Cambridge, United Kingdom, ⁴DNRF Ctr. for Chromosome Stability, Dept. of Cellular and Molecular Med., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark

Abstract Body:

Female reproductive lifespan is characterised by declining ovarian reserve, but the underlying mechanism is poorly understood. At 6 months gestation, there are around 6 to 7 million non-renewable oocytes in the fetal ovary, which are lost by apoptosis throughout life until menopause, when ovarian reserve reaches around 1 thousand oocytes. Reproductive longevity varies due to environmental and genetic factors that affect the size of the initial oocyte pool and the rate of follicle loss.

We assessed the role of rare (MAF<0.1%) protein-coding variants on ovarian function in data from 104,733 females of European ancestry in the UK Biobank using gene burden association tests. We identified protein-truncating variants in Zinc Finger Protein 518A (*ZNF518A*) associated with both earlier age at natural menopause (effect size 5.61 years, $P=1.2 \times 10^{-10}$) and later puberty timing in girls (effect on menarche timing, 0.42 years, $P=9.8 \times 10^{-4}$). The effect on menopause timing is larger than for any previously reported gene or variant. *ZNF518A* has not been previously implicated in ovarian ageing and its function is unknown. *ZNF518A* was highly expressed in human fetal primordial germ cells as well as primary and secondary oocytes during folliculogenesis. By integrating ChIP-Seq data from ENCODE, we determined that *ZNF518A* has >18,000 binding sites in the genome, many occurring in regulatory regions for genes expressed in the human fetal ovary. Common GWAS variants associated with age at natural menopause were also enriched in the binding sites of *ZNF518A*.

Our results suggest that *ZNF518A* is involved in establishing initial ovarian reserve during fetal life and may regulate many genes that are important for ovarian development. Heterozygous protein-truncating variants in *ZNF518A* are a risk factor for primary ovarian insufficiency (POI; menopause before 40 years), with 12% of carriers having POI.

Complex Traits Posters - Wednesday

PB1609*. Rare variant genetic architecture of mitochondrial DNA copy number from 415,422 exomes.

Authors:

V. Pillalamarri^{1,2,3}, W. Shi¹, C. Say⁴, S. Yang¹, J. Lane⁵, E. Guallar⁶, N. Pankratz⁵, D. Arking¹; ¹McKusick-Nathans Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ²Predoctoral Program in Human Genetics, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ³Maryland Genetics Epidemiology and Med. Training Program, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ⁴Johns Hopkins Univ. Dept. of Biology, Baltimore, MD, ⁵Dept. of Lab. Med. and Pathology, Univ. of Minnesota Med. Sch., Minneapolis, MN, ⁶Dept.s of Epidemiology and Med., and Welch Ctr. for Prevention, Epidemiology and Clinical Res., Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD

Abstract Body:

A portion of interindividual variation in the number of mitochondrial genome copies (mtDNA-CN) reflects mitochondrial function and has recently been shown to have a heritable component. We hypothesized that nuclear loci would harbor rare, functional variants that could influence variation in mtDNA-CN. We examined 415,422 exomes of self-reported White ancestry individuals from the UK Biobank with a median age of 58, and tested the impact of rare variants, both at the level of single variants and through aggregate variant-set mixed model association tests, on mtDNA-CN adjusted for cell counts and covariates including genome-wide independent common variants. A survey across nine variant sets chosen to cover coding sequence and loss-of-function mutations at multiple minor allele frequency thresholds identified 14 genes at experiment-wide significance and three genes at marginal significance. These included associations at known mitochondrial DNA depletion syndrome genes (mtDNA helicase *TWNK*, $p=5.7 \times 10^{-29}$; mitochondrial transcription factor *TFAM*, $p=4.3 \times 10^{-13}$; mtDNA maintenance exonuclease *MGME1*, $p=1.3 \times 10^{-6}$). Novel gene associations included the tyrosine kinase *JAK2* ($p=7.1 \times 10^{-17}$) mutated in myeloproliferative disease, the ATP-dependent protease *CLPX* ($p=9.9 \times 10^{-9}$) involved with mitochondrial proteome quality, and the mitochondrial adenylate kinase *AK2* ($p=5.3 \times 10^{-8}$) involved with hematopoiesis. The most significant association was a missense variant in *SAMHD1* ($p=4.2 \times 10^{-28}$), found on a rare, 1.2 Mb shared ancestral haplotype on chromosome 20. *SAMHD1* is a cytoplasmic host restriction factor involved with viral defense response and the mitochondrial nucleotide salvage pathway, and the missense mutation falls within a critical region of Aicardi-Goutieres syndrome 5 associated with childhood encephalopathy and a chronic inflammatory response. Using leave-one-out analysis, we identified an additional three independent variants driving the *SAMHD1* signal and observed a linear dose-response relationship between the genetic association with mtDNA-CN and genetic association with breast cancer risk across the four conditionally independent, uncorrelated variants at the *SAMHD1* locus (inverse-variance weighted Mendelian randomization causal estimate = 1.6, SE = 0.3, $p=5.5 \times 10^{-8}$), supporting a mechanistic link. Rare variants were enriched in Mendelian mtDNA depletion syndrome loci, and these variants further implicated core processes in mtDNA replication, nucleoid structure formation, and maintenance. Together, these data indicate strong-effect mutations from the nuclear genome contribute to the genetic architecture of mtDNA-CN.

Complex Traits Posters - Thursday

PB1610. Rare variants in ADAMTS13 lead to COVID-19 autosomal dominant disorder, conditioned by SARS CoV-2 infection, sex, and age.

Authors:

C. Fallerini^{1,2}, K. Zguro¹, M. Baldassarri^{1,2}, F. Fava^{1,2,3}, G. Beligni^{1,2}, S. Daga^{4,2}, R. Leoncini⁵, L. Galasso⁵, M. Cirianni⁵, D. Tacconi⁶, C. Spertilli Raffaelli⁶, P. Katsikis⁷, M. Lorubbio⁸, P. Calzoni⁵, A. Ognibene⁸, M. Tozzi⁹, A. Bucalossi⁹, G. Marotta⁹, S. Furini¹, GEN-COVID multicenter study, A. Renieri^{1,2,3}; ¹Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Italy, ²Med. Genetics, Univ. of Siena, Siena, Italy, ³Genetica Medica, Azienda Ospedaliero- Univ.ria Senese, Siena, Italy, ⁴Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Iceland, ⁵UOC Laboratorio Patologia Clinica, Azienda Ospedaliero- Univ.ria Senese, Siena, Italy, ⁶Dept. of Specialized and Internal Med., Infectious Diseases Unit, San Donato Hosp. Arezzo, Arezzo, Italy, ⁷Dept. of Immunology, Erasmus Med. Ctr., 3015GD, Rotterdam, Netherlands, ⁸UOC Laboratorio Analisi Chimico Cliniche, Arezzo, Italy, ⁹Stem Cell Transplant and Cellular Therapy Unit, Univ. Hosp., Siena, Italy

Abstract Body:

We have modeled COVID-19 using a machine learning approach and Whole Exome Sequencing (WES). As an extreme end of this polygenic model, we have identified a Mendelian form of X-linked recessive COVID-19 affecting males, due to rare variants in TLR7. We report here another extreme Mendelian form due to ADAMTS13, affecting mainly females (in the non-reproductive phase of life) and young males. WES analysis of 2,988 SARS-CoV-2 infected subjects was performed and the post-Mendelian model was applied. The ADAMTS13 activity was measured by TECHNOZYM® ADAMTS-13 Assay. Segregation analysis within families was performed. We found an association between ADAMTS13 ultra-rare variants (Minor Allele Frequency < 0.001) and severity in female patients (OR=3.32; 95% CI 1.34 to 8.18; p-value= 4.9x10⁻³). One of several heterozygous ADAMTS13 ultra-rare variants was identified in 124 SARS-CoV-2 infected patients (4.2%), including 49 females and 75 males. Most subjects (106 - 85.5%) had severe COVID-19 disease requiring hospitalization with signs of hyperinflammation, higher D-dimer, and a tendency to platelet consumption. The remaining not-hospitalized 18 subjects (14.5%) were either females between 18 and 50 years (14 subjects) or males over 50 years of age, likely due to the protective role of estrogens, which increase ADAMTS13 transcript. Among the hospitalized patients, there were also 3 female children younger than 10 months. In all cases in which fresh blood was available a reduction of ADAMTS13 activity was demonstrated (p-value = 0.017 at Wilcoxon test). Segregation analysis was available in 2 families: i) a 66-year-old female who required oxygen support transmitted the mutation to the 34-year-old son who required CPAP treatment; ii) a 73-year-old female treated by oxygen support transmitted the mutation to the 40-year-old daughter who was oligosymptomatic likely due to the young age; her sister 76-year-old without the mutation was asymptomatic during infection. Here, we show that carriers of Thrombotic Thrombocytopenic Purpura, (a well-known severe autosomal recessive disorder in which carriers are reported asymptomatic), have a micro-thrombotic form of COVID-19, which segregates in families as an autosomal dominant disorder, conditioned by SARS-CoV-2 infection, sex, and age.

Complex Traits Posters - Wednesday

PB1611. Rare variation in ADHD and related diagnoses in the Mass General Brigham Biobank

Authors:

F. Satterstrom¹, M. J. Daly²; ¹Broad Inst., Cambridge, MA, ²Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

Attention-deficit/hyperactivity disorder (ADHD) is estimated to affect more than 5% of children, making it a potentially important context for sequencing studies. In a Danish cohort, individuals with an ADHD diagnosis had a greater burden of constrained rare protein-truncating variants than controls, similar to individuals with an autism diagnosis (Satterstrom et al., 2019, Nature Neuroscience 22:1961-1965). We seek to follow up on this finding by examining ADHD and related psychiatric diagnoses in the biobank of the Mass General Brigham hospital system in Boston, MA, which contains the exome sequences of 49,593 unrelated individuals after sample-level quality control.

The biobank contains 2,743 individuals with an ADHD diagnosis, and we have compared their rates of constrained (i.e. the lowest decile of LOEUF) ultra-rare (i.e. gnomad non-neuro frequency $<1e-5$ and occurring only once in the biobank) protein-truncating variants to the ~30,000 individuals in the biobank without a psychiatric diagnosis. In a preliminary analysis stratified by predicted population (based on PCA), the case enrichment of these variants is greater than 10%. We plan to identify the genes harboring these variants and identify those occurring more often in cases than controls.

In addition, the biobank contains 2,004 samples with a diagnosis of schizophrenia or schizoaffective/psychotic disorder and 2,218 samples with a diagnosis of bipolar disorder (in addition to smaller numbers of autism and intellectual disability cases), enabling within-biobank comparison of these diagnoses to ADHD. We also plan to leverage the data from recent large-scale sequencing efforts from the Autism Sequencing Consortium and the Schizophrenia Exome Sequencing Meta-Analysis consortium to compare risk genes across these disorders and ADHD. From these data we will seek to identify genetic factors contributing to risk for psychiatric disorders in general as well as factors contributing to risk for specific diagnoses.

Complex Traits Posters - Thursday

PB1612. Rare-Inherited variations in Sensory genes increased burden of Autism Spectrum Disorders: A Whole Exome Sequencing study of Indian families

Authors:

A. Siddappa Niranjana Murthy, S. Valiya Parambath, S. Durbagula, A. Korlimarla, N. Kumar Gowda, A. Mysore Vishweshwaraiah; St John's Res. Inst., Bengaluru, India

Abstract Body:

De novo Mutations (DNMs) and rare inherited variations have long been implicated as the strongest contributors to risk for Autism Spectrum Disorders (ASD). Since DNMs do not account for all incidences of ASDs, it is suspected that rare inherited variations add formidable pressure toward disease manifestation. Considering that inherited variations might provide important clues on the subtle yet common behavioral dysfunctions that accumulate over generations to produce phenotypic effects in offspring, we performed the first trio Whole Exome Sequencing (WES) of 14 Indian ASD simplex families with male probands on the Illumina platform at 100X coverage. Data was processed in BWA, GATK, VarScan, and VEP. Genes were prioritized based on their evolutionary conservation, haploinsufficiency, and mutation intolerance scores and further subjected to gene-set enrichment analysis, brain-tissue expression, and protein-protein interaction analysis using the tool IPA. Of the total 539 ASD behaviors genes, 149 genes (27.6%) were enriched for core ASD behaviors- social-interaction and restricted repetitive behaviors (RRB) and 200 genes (37%) for associated ASD behaviors including language, speech, communication, learning, cognition, intelligence, emotions, and moods. Unexpectedly, inherited genes were significantly enriched for sensory perception of mechanical stimulus ($p \leq 1.27E-07$), and hereditary hearing loss ($p \leq 9.78E-06$). A total of 116 sensory genes were found carrying inherited variations. Each child carried mutations in an average of 9 sensory genes that were equally transmitted maternally and paternally; hearing and tactile sensation genes were most prevalent. Notably, 74% of these sensory genes displayed neuronal functions and 70% were directly linked to ASD core and/or associated behaviors. 62% of RRB genes had sensory functions indicating a strong interplay between them. The majority of the prenatally upregulated sensory genes produced core ASD behaviors ($p \leq 0.048$). Our cohort carried inherited variations in well-established ASD genes *CNTNAP2*, *GABRB3*, *SPAST*, *EP300*, *CHD7*, *HIVEP2*, *DLG2*, *CNTN5* as well as novel genes *APC*, *WFS1*, *KIF1B*, *ESR1*, *HGS*, *JAK2* producing sensory dysfunctions. Supporting evidence from a previous study on individuals with sensory dysfunctions without ASD found increased mutations in ASD genes. Thereby, we suspect that underlying, undiagnosed sensory dysfunctions in parents might collectively increase the risk for ASD in offspring. Sensory functions of ASD genes must be studied for a better understanding of the role of sensory dysfunctions in ASD manifestations.

Complex Traits Posters - Wednesday

PB1613. Relationship between the gut microbiome and depression in individuals with post traumatic stress disorder

Authors:

C. Finnicum¹, J. L. Scholl², C. Davis¹, S. Viet¹, P. L. Jason¹, L. A. Baugh², E. A. Ehli³; ¹Avera McKennan Hosp. & Univ. Hlth.Ctr., Sioux Falls, SD, ²Univ. of South Dakota, Vermillion, SD, ³Avera Inst. for Human Genetics, Sioux Falls, SD

Abstract Body:

The gut microbiome is a large collection of microbial inhabitants, living within the human GI system. The gut microbiome is known to be involved in various human traits and diseases, including links to mental health disorders such as post-traumatic stress disorder (PTSD) and depression. We sought to further understand how the gut microbiome is related to depression, particularly in individuals with PTSD. DNA was extracted from stool samples collected from PTSD-associated participants (N = 16) and was used for 16S rRNA profiling to survey the bacterial composition of the gut microbiome. Our preliminary analyses, comparing the gut microbiomes of individuals with PTSD and varying degrees of depression (minimal, mild, moderate, severe), identified a single significant *genus*, *Dialister* ($p < 1 \times 10^{-5}$) after controlling for multiple testing. This *genus* was particularly enriched in the PTSD-associated individuals with moderate depression, while markedly depleted in the PTSD-associated individuals with severe depression. No significant enrichment of taxa was observed at the species or phylum levels. Furthermore, no significant associations were observed with the alpha diversity, although a trend towards a lower alpha diversity with higher severity of depression was observed. Active enrollment is planned and underway, which will aid in further increasing the statistical power of the study.

Complex Traits Posters - Thursday

PB1614. Results from largest GWAS in Gastroparesis point to macrophage polarization etiology

Authors:

S. Smieszek; Vanda Pharmaceuticals, Washington, DC

Abstract Body:

Gastroparesis is a serious medical condition characterized by delayed gastric emptying and symptoms of nausea, vomiting, bloating, fullness after meals, and abdominal pain. An innate immune dysregulation and injury to the interstitial cells of Cajal and other components of the enteric nervous system is likely central to the pathogenesis of gastroparesis. Thus far, little is known about the underlying genetic risk factors for gastroparesis. We have done a genome-wide association study comparing idiopathic and diabetic cases with controls as well as contrasting idiopathic versus diabetic cases.

The samples were obtained from patients with idiopathic and diabetic gastroparesis enrolled in 2 clinical trials of gastroparesis (VP-VLY-686-2301 VP-VLY-686-3301). The dataset consisted of 1215 WGS samples from screened subjects. Subjects included male and female adults age 18-70 with a diagnosis of diabetic or idiopathic gastroparesis.

We report a novel genetic association of SLC15A4 locus with idiopathic gastroparesis. This signal is driven by multiple variants with top missense variant SLC15A4:NM145648:exon2:c.T716C:p.V239A, rs33990080. Altogether, among EUR (PCA defined) cohort we report 69 carriers out of 214 versus controls 165 out of 896 (MAF: cases 0.18; MAF: controls 0.09) and OR: 1.9 (p-value logistic model 10^{-9}). The significant effect persists when idiopathic cases are compared with diabetic cases as the MAF of diabetic cases (0.10) is comparable to that of population controls (0.09). The variant is a strong SLC15A4 eQTL across multiple tissues reported in GTEx.

Solute carrier family 15 (SLC15) A4 is a lysosome-resident amino acid/oligopeptide transporter that is preferentially expressed in immune cells. It is required for TLR7/9-dependent type I interferon production and plays a critical role in the pathogenesis of lupus and colitis. SLC15A4 was shown to mediate M1-prone metabolic shifts in macrophages and guards immune cells from metabolic stress. Previous reports have shown that when the anti-inflammatory M2 macrophages are switched to pro-inflammatory M1 macrophages, delayed GE was evident in animal models. The variant is associated with severe nausea. We also differentiate idiopathic from diabetic gastroparesis with an established diabetes genetic risk score (p-value<0.0004).

Current genetic findings suggest that a mechanism directly connected to macrophage polarization may be implicated in a subset of idiopathic gastroparesis patients. The GWAS picture that is emerging implicates novel loci implicated in the pathophysiology of gastroparesis differentiating those of diabetic versus idiopathic etiology.

Complex Traits Posters - Wednesday

PB1615. *RGS3* and *ILIRAPL1* missense variants implicate defective neurotransmission in early-onset inherited schizophrenias

Authors:

A. Kanwal¹, J. Pardo², S. Naz³; ¹Sch. of Biological Sci., Univ. of the Punjab, Lahore, Pakistan, ²Minneapolis VA Hlth.Care System, minneapolis, MN, ³Sch. of Biological Sci., Univ. of the Punjab, lahore, Pakistan

Abstract Body:

Schizophrenia is a clinically and genetically heterogeneous complex disorder characterized by hallucinations, disorganized behavior, delusions, and negative symptoms. Despite identification of hundreds of susceptibility loci associated with schizophrenia, monogenic causes remain largely undiscovered. We designed a family study to identify inherited rare gene variants for schizophrenia. Two consanguineous families each with two patients affected by severe treatment-resistant schizophrenia were recruited. Structured clinical evaluations supported the diagnosis in the four patients. Whole-exome sequencing was performed for all participants. Data analyses identified few variants segregating with schizophrenia of which only two were likely pathogenic. The homozygous c.649C>T, p.(Arg217Cys) variant in *RGS3* and a hemizygous c.700A>G, p.(Thr234Ala) variant in *ILIRAPL1*, were the most likely causes of phenotype in patients of families PSYAK2 and PSYAK3, respectively. The two variants were rare in all publicly available databases (gnomAD allele frequency = 0.00004692 and 0.000005491, respectively) and absent from the DNA of at least 400 ethnically matched controls. Amino acids affected by these variants were conserved among different orthologs. *RGS3* is involved in the G-protein signaling pathway and was previously implicated in modulating sensory behavior in *C. elegans*. *ILIRAPL1* plays a known role in synapse formation and modulation. Variants of *ILIRAPL1* have been described to cause non-syndromic X-linked intellectual disability with or without behavioral defects in humans. Our work suggests that these missense variants are involved in causing schizophrenias. Cellular assays and variant-specific functional studies will elucidate the pathophysiology relevant to schizophrenias and will motivate translation to novel, personalized therapeutics. Our work is funded by NIMH, NIH USA, 1R21MH120692-01A1

Complex Traits Posters - Thursday

PB1616. Risk factors involved in Congenital Diaphragmatic Hernia, a case-control study in Bogotá and Cali, Colombia.

Authors:

C. Acevedo Castaño¹, G. Villamil Patiño¹, M. Uribe Gaviria¹, S. A. Suarez Gomez¹, L. A. Quintero Riaño¹, L. J. T. Rincon Abreo¹, M. S. Parra Artunduaga¹, S. C. Bernal Bejarano¹, K. Sarmiento¹, I. M. Zarante¹, P. Hurtado², J. A. Holguín³; ¹Pontificia Univ. Javeriana, Bogotá, Colombia, ²Pontificia Univ. Javeriana, Cali, Colombia, ³Secretaria de Salud Pública de Cali, Cali, Colombia

Abstract Body:

The aim of the study was to determine, classify, quantify and qualify the possible risk factors involved in the development of congenital diaphragmatic hernia (CDH), through the analysis and comparison of sociodemographic, clinical and anthropometric variables; related to the mother, the newborn and the pregnancy that could possibly be involved, as well as to determine the degree of association between these variables. In order to expand the knowledge of the practitioner and thus facilitate the identification of the disease and generate a prompt diagnosis. **Materials and Methods** A descriptive retrospective study of cases and controls was carried out, with patients registered in the Surveillance and Follow-up Program for Congenital Defects of Bogotá and Santiago de Cali (PVSACB-C) with 51 and 3 Hospitals, respectively, between 2001-2018. With 546,123 births reported between these periods. A sample of 83 cases and 332 controls was selected with a ratio between cases and controls of 1:4, respectively. In our data analysis results CDH is significantly associated with maternal age (MA) >35 years, low birth weight, male sex, primigravid mother, socioeconomic status, cesarean delivery, weight for gestational age (WGA) (OR 1.95, 4.6, 1.75, 2.90, 6.17, 3.59, 8.28), (CI 1.04-3.67, 2.64-8.01, 1.08-2.86, 2.59-14.7, 1.91-6.73, 4.78-14.35) respectively; negatively with maternal BMI greater than 25 (OR 0.39 (0.16-0.96)) and was not associated with gestational age < 35 weeks, alcohol consumption of illicit substances during pregnancy, parental consanguinity (OR 0.99, 1.34, 9.04, 1.53), (CI 0.42-2.35, 0.49-3.68, 0.8-101.66, 0.94-2.48) respectively, among others. It was identified that of those born with HD (n = 83), 91.5% were born at term, 59% had a birth weight greater than 2,500 g, 71.3% were the mother's first birth, 27.7% of the children were classified as polymalformed since they had one or more additional malformations to HD and 7.2% were included within some syndrome. In conclusion, despite the lack of etiological knowledge and given the high prevalence of this disease, the data obtained in this study provide useful information to identify and learn more about CDH, as well as its associated anomalies, in order to achieve a prompt diagnosis.

Complex Traits Posters - Wednesday

PB1617*. Risk factors that affect performance of polygenic risk scores across diverse cohorts

Authors:

D. Hui, M. D. Ritchie; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Polygenic risk score (PRS) performance varies across datasets, which prevailing evidence suggests is mostly due to ancestry differences. However, differences in personal or environmental covariates may also affect PRS performance. Little is known why select covariates do so and assessing myriad environmental exposures is infeasible; covariate-specific effects are also notorious for poor replicability in human genetics studies. To explore risk factors affecting PRS, we analyzed the effects of covariate stratification and interaction on PRS for body mass index (BMI) across four cohorts of European (EUR) and African (AFR) ancestry - UK Biobank (UKBB), Penn Medicine Biobank (PMBB), eMERGE, and GERA. We assessed 62 covariates for R^2 differences, stratifying on binary covariates and quintiles for continuous covariates. With UKBB EUR as discovery, 18 covariates had significant differences in R^2 among groups, including age, sex, alcohol, physical activity, Townsend index, diet, lipids, blood pressure, and A1c. Relative R^2 s between worst and best performing physical activity, alcohol intake, and HDL quintiles were 71%, 61%, and 56%, respectively, comparable to differences due to ancestry. PRS R^2 of bottom age quintile PMBB AFR individuals ($R^2 = .058$, mean = 28) was higher than top age quintile PMBB ($R^2 = .051$, mean = 78) and eMERGE ($R^2 = .047$, mean = 82) EUR individuals. Main covariate effects on BMI and max R^2 differences across quintiles had .56 correlation, suggesting PRS R^2 increases for individuals with higher BMI-associated covariates (i.e., most at-risk). This result may also reduce search for PRS-interacting covariates as outcome-associated covariates are often known. Next, we modeled PRS-covariate interaction terms and observed replicable interactions with age, sex, lipids, and blood pressure. Magnitude of main and interaction effects had .43 correlation, suggesting BMI-associated covariates excessively alter PRS effects. Unlike published reports, smoking and education interactions were insignificant, supported by their small main effects on BMI. R^2 increases when including interaction terms were small (max increase .0024). Given PRS-covariate dependence evidence, we assessed R^2 gains using machine learning models (neural networks) versus L1-regularized linear regression. Relative cross-validated R^2 increased up to 67% (mean 25%) across cohorts with age and sex as covariates, and up to 16% (mean 7%) with all available covariates. These results suggest covariate-specific effects can affect PRS comparably to ancestry, covariates affect PRS proportionately to their outcome association, and machine learning models improve PRS prediction.

Complex Traits Posters - Thursday

PB1618. RNA editing regulates interferon induction and host immune response in SARS-CoV-2 infection

Authors:

M. Huang¹, A. Mark², K. Vera¹, C. Meydan³, J. Foox³, D. Butler³, C. Mozsary³, A. Saravia-Butler⁴, A. Beheshti^{4,5}, C. Mason³, Q. Jiang⁶, K. Fisch^{1,2}; ¹Dept. of Obstetrics, Gynecology & Reproductive Sci., Univ. of California, San Diego, La Jolla, CA, ²Ctr. for Computational Biology & Bioinformatics, Univ. of California, San Diego, La Jolla, CA, ³Dept. of Physiology and Biophysics, Weill Cornell Med., New York, NY, ⁴KBR, NASA Ames Res. Ctr., Moffett Field, CA, ⁵Broad Inst. of MIT and Harvard, Cambridge, MA, ⁶Div. of Regenerative Med., Univ. of California, San Diego, La Jolla, CA

Abstract Body:

Viruses enhance the RNA editing function of *ADARI* in the host and emerging data points to a role of *ADARI*-mediated A-to-I editing in the inflammatory response associated with severe COVID-19 disease. Understanding the relationships between the SARS-CoV-2 induced cytokine storm, disease severity, and clinical outcomes is key to developing effective treatments and prognostic biomarkers for COVID-19 infection.

Human whole transcriptome data from nasopharyngeal swabs was obtained from 208 SARS-CoV-2 positive, 82 positive for other viral illness (OVI) and 411 SARS-CoV-2 negative patients. We performed RNA editing analysis to obtain A-to-I editing events present in at least 5% of samples, differential editing (DE) at each site using a log likelihood ratio test, enrichment analysis using clusterProfiler, cellular deconvolution with XCell, and interrogated *ADARI* expression within immune cell types in SARS-CoV-2 infected nasopharyngeal swabs with the BROAD Single Cell Portal. Validation was performed by activating CD8 and CD4 naive T cells with CD3/28 beads. *ADARI* expression and IL2 level was quantified by flow cytometry and RT-qPCR, and knockdown of *ADARI* was performed using lentiviral vectors in CD8 and CD4 T-cells followed by analysis of apoptosis (Annexin V), cell cycle (Ki67), and IL2 production using flow cytometry.

Overall editing levels ($p = 0.0026$) and variant allele frequencies ($p = 6.2E-6$) were higher in SARS-CoV-2 positive patients than negative patients. RNA editing levels increased with decreasing viral load ($p = 2.4E-5$). We identified DE sites ($\text{padj} < 0.05$) in SARS-CoV-2 vs None: 1,466; OVI vs None: 1,031; SARS-CoV-2 vs OVI: 1,139; SARS-CoV-2 High vs Low Viral Load: 734. DE in SARS-CoV-2 vs None were enriched in 23 pathways including interferon signaling ($\text{padj} = 6.6E-8$) and antiviral mechanisms by interferon stimulated genes ($\text{padj} = 0.00021$) in High vs Low Viral Loads. Interferon responsive cytotoxic CD8 T-cells highly express *ADARI* at single cell resolution and cellular deconvolution revealed significantly higher enrichment for CD8+ naive T-cells in OVI and None relative to high viral load ($p = 0.015$ and 0.00016 , respectively). We observed a striking increase in *ADARI* levels within 48 hrs (>2.5 fold) by quantifying the *ADARI* protein and mRNA levels in CD3/38-activated CD4 and CD8 T cells. Finally, lentiviral directed *ADARI* knockdown in naive CD4 cells followed by CD3/28 activation revealed that *ADARI* knockdown leads to accelerated T-cell proliferation coupled with necroptosis and overproduction of IL2. This study highlights the importance of *ADARI* activation in T-cell homeostasis with relevance to SARS-CoV-2 infection.

Complex Traits Posters - Wednesday

PB1619. RNASeq in nasal epithelium from African ancestry subjects in the CAAPA consortium reveal *IL4* and *TGFBI* as upstream regulators of differentially expressed networks for asthma

Authors:

B. Szczesny¹, K. Kammers², M. P. Boorgula³, M. Taub⁴, M. Campbell³, R. Johnson⁵, I. Ruczinski⁶, S. Chavan³, C. Figueiredo⁷, C. N. Rotimi⁸, R. C. Landis⁹, H. Watson⁹, N. N. Hansel¹⁰, I. V. Yang¹¹, C. O. Olopade¹², C. Ober¹³, A. H. Liu¹⁴, E. Kenny¹⁵, K. C. Barnes¹⁶, R. A. Mathias², CAAPA Consortium; ¹Johns Hopkins Univ., Baltimore, MD, ²Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ³Univ. of Colorado, Denver, CO, ⁴Johns Hopkins, Baltimore, MD, ⁵Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ⁶Johns Hopkins Univ., Dept. of Biostatistics, Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ⁷Univ. of Bahia, Salvador, Brazil, ⁸Natl. Human Genome Ctr., Howard Univ. Coll. of Med., Washington, DC, ⁹The Univ. of the West Indies, Queen Elizabeth Hosp., Bridgetown, Barbados, ¹⁰Johns Hopkins Sch. of Med., Baltimore, MD, ¹¹Univ. of Colorado, Aurora, CO, ¹²Univ. of Chicago Med., Chicago, IL, ¹³Dept. of Human Genetics, Univ. of Chicago, Chicago, IL, ¹⁴Children's Hosp. Colorado, Aurora, CO, ¹⁵Icahn Sch. of Med. at Mt Sinai, New York, NY, ¹⁶Univ. of Colorado, Anschutz Sch. of Med., Aurora, CO

Abstract Body:

Asthma, a complex chronic lung disease affecting the airways, has striking disparities across racial and ethnic groups. The second phase of the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) aims to understand genomic signatures of asthma and asthma severity in populations of African descent.

RNASeq data was generated from nasal epithelium in subjects recruited from 7 sites (Baltimore, Washington DC, Chicago, Denver, Brazil, Barbados, and Nigeria). Analysis was limited to active asthmatics (N=253, 8-80 years of age) determined using a respiratory health questionnaire and the Composite Asthma Severity Index, and never-asthmatic controls (N=283, 8-75 years of age). Pre-alignment quality control, adapter trimming, and alignment of reads to GRCh38 were performed using FastQC and Picard tools, BBDUK, and HISAT2, respectively. Raw counts were generated by CoCo, mean normalized counts were generated by DESeq2, and 21,831 advanced to differential analysis in LIMMA and edgeR. Linear models were fit to the normalized expression data adjusting for relevant covariates.

At a q-value<0.05, we found 389 differentially expressed genes (DEGs). The top 10 DEGs included *FNI*, *POR*, *VSIG4*, *SUSD4*, *PTCHD4*, *HS3ST4*, *PPP1R9A*, *SPTBNI*, *SNTG2*, and *RHEX*. Several of these have been previously implicated and/or have biological relevance for asthma: *SNTG2* (q-value= 1e-4) is the target of multiple miRNAs related to asthma. *SUSD4* (q-value=1e-6) plays a role in cell-to-cell adhesion and has been shown to harbor complement inhibitory activity; it has similar homology to *CD46* which has a known regulatory role in T cell activation. *PPP1R9A* expression (q-value=7e-5) was previously determined to be influenced by IL-13 in mouse lung, and IL-13 is documented to contribute to asthma pathogenesis. *SPTBNI* (q-value=1e-4) plays a key role in mediating TGF-beta signaling. The most significant differentially expressed gene, *FNI* (q-value=3e-9), was downregulated in asthmatics. *FNI* encodes fibronectin, which plays a key role in airway remodeling. Ingenuity Pathway Analysis reveals several networks with upstream regulators highly relevant for asthma; the top 5 of these include two genes: *IL4* (p=7.4e-10) and *TGFBI* (p=6.8e-8), and two drugs: dexamethasone (p=4.3e-10) and fluticasone propionate (p=6.8e-8).

Our analyses reveal that many genes with known and plausible mechanistic roles in asthma are differentially expressed in nasal epithelium of asthmatics of African ancestry in CAAPA. Ongoing work

ASHG 2022 Annual Meeting Poster Abstracts

includes testing for networks related to asthma and asthma severity, and integration with eQTLs and methylation data in the nasal epithelium of these same subjects.

Complex Traits Posters - Thursday

PB1620. Role of amino acids in COVID-19 virulence and pathogenicity : an observational study

Authors:

H. Singh; CHC, Pilkhuwa, Hapur, Hapur, India

Abstract Body:

Role of amino acids in Covid 19 virulence & pathogenicity Virus machinery is dependent on host body for their activation & survival. The multiplication within the host body requires a suitable environment rich in specific Amino acids (AA) like arginine to be more specific. The alteration in host nitrogen balance, protein & lipid metabolism, and catabolic effects make the host body vulnerable to viral infection due to altering AA concentration. Currently, AA restriction diet is been advised in cancer patients to sustain cancer progression & proved to be successful. AAs are mediators of metabolic cross-talk between host & pathogen. Methodology The study comprises 28 COVID 19 patients including 12 new variants (UK strain). All the patients were informed & taken consent before taking a small drop of blood from finger pulp, on filter papers. After that, liquid base chromatography was performed to titer the AA profile of blood spots i.e. alanine, arginine, citrulline, glycine, leucine, valine, methionine, ornithine, phenylalanine, tyrosine & proline. Symptoms of the patients were noted & later, the correlation was attempted between AAs & symptoms to understand & the virulence & pathogenicity of COVID-19. All the patients were included under the act of ethics. Result We observed arginine being the main highlight of the study, which was raised to 100 $\mu\text{MOL/L}$ in one patient J6 (Jaipur, India patient 6), whose symptoms were aggressive in comparison to other patients. Also, leucine & valine values were also significantly increased to 450 & 352 $\mu\text{MOL/L}$ respectively. Patient J5 showed increased arginine value up to 64 $\mu\text{MOL/L}$, which was near to maximum limit. That patient J5 showed symptoms like loss of smell & taste for 7 days including 2/3 days severe body ache. All the United Kingdom (UK) variants of COVID-19 patients were asymptomatic & their arginine value was noted to be very low in comparison to other patients. Conclusion- we hypothesized that arginine plays a major role in the pathogenicity & virulence of Coronavirus. The body microenvironment of AAs in body may decide the outcome of infections. So, the diet should be considered important along with focusing on hygiene, wearing masks, & medication

Complex Traits Posters - Wednesday

PB1621. SARS-CoV-2 sequencing: A comparison of high-throughput methods.

Authors:

R. Olasso¹, Z. Gerber¹, C. Daviaud¹, D. Delafoy¹, F. sandron¹, E. Alidjinou², J. Mercier¹, S. Gerber³, V. Meyer¹, A. Boland¹, L. Bocket², J. Deleuze¹; ¹CEA - CNRGRH, Evry, France, ²CHU de Lille, Lille, France, ³MNHN - CNRS, Paris, France

Abstract Body:

Introduction: The COVID-19 pandemic continues to threaten public health and burden healthcare systems worldwide. Whole SARS-CoV-2 genome sequencing has become essential for epidemiological monitoring and identification of new variants, which could represent a risk of increased transmissibility, virulence, or resistance to vaccines or treatment. In this study, we assess the performance of various target enrichment methods for whole SARS-CoV-2 sequencing.

Methods: We applied three target enrichment methods - two multiplex amplification methods and one hybridization capture - to the same set of nasopharyngeal patient samples (N = 93) in high-throughput mode. SARS-CoV-2 genome was obtained using short-read next-generation sequencing.

Results: All three methods provided excellent breadth of coverage of SARS-CoV-2 genome (above 99%), albeit with vastly different sequencing depth (5-fold difference) and uniformity of coverage (20% difference in coefficient of variation). Poor local coverage has negative impact on variant calling in the concerned region, leading to an occasional allele drop-out (1.2% SNPs affected for one method).

Conclusion: We discuss the performance of each target enrichment method and their potential for scaling up, in order to promote prospective programs of large-scale genomic surveillance of SARS-CoV-2 worldwide. Genomic surveillance will be crucial to overcoming the ongoing pandemic of COVID-19, despite its successive waves and continually emerging variants.

Funding: LabEx GENMED (grant number ANR-10-LABX-0013).

Complex Traits Posters - Thursday

PB1622. Screening of variant rs11190870 nearby *LBXI* gene for association with Adolescent Idiopathic Scoliosis in the population of North India

Authors:

H. Singh¹, S. Kowra¹, M. Gupta², P. Bhau¹, T. Chalotra¹, R. Manotra¹, N. Gupta³, G. Gupta⁴, A. Pandita⁵, M. Bhat⁶, R. Sharma⁶, S. Pandita⁷, V. Singh¹, E. Rai¹, B. Garg², S. Sharma¹; ¹Shri Mata Vaishno Devi Univ., Katra, India, Jammu, India, ²All India Inst. of Med. Sci., New Delhi, India, ³District Hosp. Poonch, Jammu and Kashmir, Poonch, India, ⁴Dept. of Radiology, Shri Mata Vaishno Devi Narayana Superspeciality Hosp., Jammu, India, ⁵Accidental Hosp., Chowki Choura, Jammu, India, ⁶Government Med. Coll., Jammu, India, ⁷Radiology section, Chest Disease Hosp., Bakshi Nagar, Jammu, India

Abstract Body:

Adolescent idiopathic scoliosis (AIS) is the most common spinal deformity of adolescent individuals accompanied with the lateral curvature and twisting of spine with Cobb angle of $\geq 10^\circ$. Our recently published epidemiological study, where 9500 post adolescent children were screened, has shown uniquely low incidence (0.61%) of AIS in northwest India. Globally, genetic predisposition to AIS has been studied extensively. However, any such study in Indian populations is yet lacking. The present study is the first of the kind that has been carried out to evaluate variant rs11190870 nearby *LBXI* gene, reported to be associated with AIS in literature, in North Indian population. In addition to samples collected during population screening, samples from special care hospitals and clinics have been collected to increase the patient sample size from the region, in this SMVDU Institutional Ethical Review Board approved study. A total of 95 AIS cases and 282 healthy controls were screened in this pilot case control association study for AIS from North India. The genotyping was carried out using TaqMan Allele discrimination assay on Mx3005p Real-time PCR System. The genotypes distributions, in cases (CC - 0.141, CT - 0.353 and TT - 0.506) and controls (CC - 0.136, CT - 0.398, TT - 0.466) separately as well as in total samples, were found to be following Hardy Weinberg Equilibrium. The allele frequencies distribution in cases were T = 0.682 and C = 0.318 and in controls T = 0.665 and C = 0.335, respectively and the variant was not found associated with the AIS in the population. Interestingly, similar allele frequencies were noted in healthy public datasets for other Indian populations. However, in contrast to other global populations in literature, where T allele has been noted in lower frequencies in controls than AIS cases. Similar distribution of alleles in cases and controls resulted in non-association of the variant in screened population. This is an interesting finding and worth reporting timely. *LBXI* is a candidate gene for AIS and rs11190870 is nearby variant reported as part of haplotype providing susceptibility to AIS. Present study highlights need for screening of other SNPs across the gene to elucidate if haplotypic differences exist w.r.t Indian populations. Also, study in an increased sample size would be ideal to increase the confidence in the findings. This study provides an insight in the genetics of AIS in an Indian population group. It also highlights the possibility of the potential haplotypic differences or genetic heterogeneity and lays a foundation for carrying out a genome wide association study in Indian population for better understanding of AIS, a task we are pursuing.

Complex Traits Posters - Wednesday

PB1623*. Secretoglobin family 1D member 2 (SCGB1D2) protein inhibits growth of *Borrelia burgdorferi* and affects susceptibility to Lyme disease

Authors:

S. Strausz^{1,2}, G. Blacker², S. Galloway², H. Paige^{2,3}, S. Jones¹, E. C. Sanders^{2,3}, N. Sinnott-Armstrong², FinnGen, I. L. Weissman², M. Daly¹, T. Aivelo¹, M. C. Tal^{2,3}, H. M. Ollila^{1,4,5,6}; ¹Univ. of Helsinki, Helsinki, Finland, ²Stanford Univ., Stanford, CA, ³Dept. of Biological Engineering, Massachusetts Inst. of Technology, Cambridge, MA, ⁴Broad Inst. of MIT and Harvard, Cambridge, MA, ⁵Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁶Anesthesia, Critical Care, and Pain Med., Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA

Abstract Body:

Lyme disease is a tick-borne disease caused by bacteria of the genus *Borrelia*. The disease can initially manifest as an erythema migrans rash and, if able to evade the host immune defenses, can progress into a secondary stage chronic disease with debilitating physical or neurological manifestations. The host factors that modulate susceptibility for Lyme disease have remained mostly unknown. Here we show a novel host defense mechanism against Lyme disease in humans. Using epidemiological and genetic data from FinnGen, we identify a common missense variant at the gene encoding for Secretoglobin family 1D member 2 (SCGB1D2) protein that increases the susceptibility for Lyme disease. The genetic variant changes proline at position 53 to leucine and is predicted as deleterious. Consequently, we validate the dysfunction of this protein variant using live *Borrelia burgdorferi* (Bb). Recombinant reference SCGB1D2 protein inhibits the growth of Bb twice as effectively as the recombinant SCGB1D2 P53L deleterious missense variant. Together, these data suggest that SCGB1D2 is a host defense factor present in the skin, sweat, and other secretions which protects against Bb infection. This finding provides a novel therapeutic avenue for drug development to prevent and treat Lyme disease.

Complex Traits Posters - Thursday

PB1624. Self-reported and device-based physical activity and risk of Alzheimer's disease and related dementias: a Mendelian randomization study.

Authors:

G. Arani¹, T. J. Thompson¹, M. S. Newell², K. W. Choi^{3,4,5,6}, C-Y. Chen^{7,8,4}, D. A. Raichlen⁹, G. E. Alexander¹⁰, Y. C. Klimentidis¹¹; ¹Dept. of Epidemiology and Biostatistics, Mel and Enid Zuckerman Coll. of Publ. Hlth., Univ. of Arizona, Tucson, AZ, ²Veteran's Hlth. Admin., United States Dept. of VA, Phoenix, AZ, ³Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ⁴Psychiatric and Neurodevelopmental Genetics Unit, Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁵Dept. of Psychiatry, Massachusetts Gen. Hosp., Boston, MA, ⁶Stanley Ctr. for Psychiatric Res., Broad Inst., Cambridge, MA, ⁷Biogen, Cambridge, MA, ⁸Analytic and Translational Genetics Unit, Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁹Dept.s of Psychology and Psychiatry, NeuroSci. and Physiological Sci. Interdisciplinary Programs, BIO5 Inst., and Evelyn F. McKnight Brain Inst., Tucson, AZ, ¹⁰Dept.s of Psychology and Psychiatry, NeuroSci. and Physiological Sci. Interdisciplinary Programs, BIO5 Inst., and Evelyn F. McKnight Brain Inst., Univ. of Arizona, Tucson, AZ, ¹¹Dept. of Epidemiology and Biostatistics, Mel and Enid Zuckerman Coll. of Publ. Hlth., Univ. of Arizona, Tucson, AZ

Abstract Body:

Background: Current observational studies suggest that physical activity may have a protective role in Alzheimer's disease and related dementias (ADD). However, these may be prone to residual confounding, bias, and reverse causation. Mendelian randomization (MR) is an alternative approach that may avoid these limitations in order to better assess the causal role of physical activity in ADD. **Aim:** We conducted Mendelian randomization analyses to examine whether genetically-predicted levels of physical activity (PA) are associated with ADD. **Methods:** We used GWAS of three self-reported measures ($n_{\max}=377,234$): moderate-to-vigorous physical activity (MVPA), vigorous physical activity (VPA), and strenuous sports or exercises (SSOE), along with two accelerometry-based measures ($n_{\max}=91,084$): acceleration average (AAv) and fraction of accelerations >425 mg (AF). We used the most recent GWAS of ADD that included 111,326 cases and 677,633 controls from the European Alzheimer's & Dementia Biobank. (EADB) We used a p-value threshold of $p < 5 \times 10^{-7}$ to identify genetic instruments for each PA phenotype. We primarily assessed MR associations using the inverse-variance weighted (IVW) method and conducted multiple sensitivity analyses. **Results:** We found positive trends in the estimates for MVPA (0.12, 95% CI: -0.10, 0.34), and AF (0.13, 95%CI: -0.64,0.89), with ADD outcome. We found negative trends in the estimates for VPA (-0.17, 95% CI: -0.99, 0.65), SSOE (-0.29, 95% CI: -0.77, 0.20), and AAv (-0.02, 95% CI: -0.05, 0.02), with ADD outcome. However, none of these IVW estimates were statistically significant. Our sensitivity analyses also did not reveal any evidence of associations between genetically-predicted PA phenotypes and the risk of ADD. **Conclusion:** We leveraged both self-reported and objective measures of PA, and the largest ADD GWAS to date in our Mendelian randomization analyses. However, we did not find any evidence to support that the genetic propensity for different types of physical activity is associated with the risk of Alzheimer's disease and related dementias. The lack of evidence for a causal effect may be explained by the absence of a true causal effect, or by the limitations of the MR approach.

Complex Traits Posters - Wednesday

PB1625. Sequencing-based genome-wide association study of triglycerides in East Africans.

Authors:

K. Meeks, A. Bentley, M. Gouveia, A. Doumatey, D. Shriner, G. Chen, L. Lei, J. Zhou, A. Adeyemo, C. Rotimi; NIH, Bethesda, MD

Abstract Body:

Despite dramatically different environmental exposures, studies consistently show that West African ancestry populations in Africa and in the diaspora have substantially lower triglyceride (TG) levels compared with European ancestry populations. In contrast, East Africans (EA) experience disproportionately high TG levels. The reasons for the high TG levels in EA are unclear, although there is evidence that suggests a population-specific genetic architecture for TG. We performed whole-genome sequencing (WGS) on a subset of 199 unrelated Kenyans from the Africa America Diabetes Mellitus (AADM) study using an extreme phenotype sampling strategy: 99 samples with the highest TG levels (mean = 363 mg/dl) and 100 with the lowest TG levels (mean = 70 mg/dl). Participants were aged 37-80 years and were not on lipid-lowering medication. The groups were matched on age, sex (52-54% women), and ethnic group (60-64% Kalenjin). WGS was performed on the Illumina NovaSeq 6000 platform and variant calling was done using GATK best practices guidelines. After filtering based on variant quality score recalibration, missingness (> 50%), minor allele count (< 3), and Hardy Weinberg Equilibrium (< 1×10^{-6}), 15.9 million variants remained for analysis. Common variants (MAF > 5%) were analyzed using a mixed linear model in GENESIS with adjustment for body mass index. We identified 10 loci with P-value < 5×10^{-6} . Half of the variants best representing these loci had a higher frequency in the high-TG group (*NEDD9*, *SERPINA13P*, *LOC107985969*, *PPP1R3A*, *AC020551.1*) and the other half had a higher frequency in the low-TG group (*REL*, *VPS13C*, *DOK6*, *FRMD4A*, *TMEM170A*). Gene-based burden testing of variants with MAF \leq 5% revealed *TMEM170A* as the gene set with the lowest P-value (1.5×10^{-5}). Next, we identified rs1291962476 (*LOC105377067*) and rs369751109 (*TUBGCP5*) as having a frequency of > 8% in the low and high TG group, respectively, and being monomorphic in the other. Annotation showed that one of the common loci (*REL*) was found only in people of African descent. The *REL* gene encodes a protein involved in the regulation of apoptosis, inflammation, immune response, and oncogenic processes. Two common variants were eQTLs for their annotated gene (*VPS13C*, *TMEM170A*). *TMEM170A* is involved in endoplasmic reticulum and nuclear envelope organization. In summary, these first genome-wide sequencing analyses in East Africans identified candidate loci that may play a role in the high TG levels in this population. The loci identified will be further investigated through association studies in a larger cohort of Africans and functional studies.

Complex Traits Posters - Thursday

PB1626. Serum osteopontin levels in Mexican patients with systemic lupus erythematosus with lupus nephritis and molecular interaction review

Authors:

A. Rivera-Cameras^{1,2}, M. P. Gallegos Arreola², M. C. Morán Moguel³, M. Salazar Páramo⁴, M. Alcaraz López⁵, J. F. Topete Reyes⁵, L. Bobadilla Morales⁶, I. Cuero Quezada^{1,6}, I. P. Dávalos Rodríguez^{1,2}; ¹Doctorado en Genética Humana, CUCS, UDG., Guadalajara, Mexico, ²División de Genética, CIBO, IMSS., Guadalajara, Mexico, ³Departamento de Disciplinas Filosófico-Methodológicas e Instrumentales, CUCS, UDG, Guadalajara, Mexico, ⁴Departamento de Fisiología, CUCS, UDG, Guadalajara, Mexico, ⁵Hosp. Gen. de Zona 46, Inst. Mexicano del Seguro Social, Guadalajara, Mexico, ⁶Centro de Registro e Investigación de Anomalías Congénitas (CRIAC), Servicio de Unidad de Genética y Citogenética, División Pediátrica, Hosp. Civil Dr. Juan I. Menchaca de Guadalajara, Guadalajara, Mexico

Abstract Body:

Introduction: Osteopontin (OPN) is encoded by the *SPP1* gene, it is a glycoposphoprotein, which acts as a cytokine. OPN overexpression has been associated with a predisposition to systemic lupus erythematosus (SLE) and a poor prognosis. OPN is known to mediate polyclonal B-cell activation, promote T follicular helper cells (TFH) and enhance antinuclear antibody production, where OPN's role in lupus nephritis (LN) is central to tissue damage. **Objectives:** Compare the serum levels of OPN in Mexican patients with SLE, without LN and with LN and analyze the protein-protein molecular interaction networks of OPN. **Materials and methods:** The study included forty-three patients (18-82 years) affected by SLE, 17 with LN and 26 without LN. OPN serum levels were quantified using the ELISA technique (Invitrogen Osteopontin/SPP1 Human ELISA Kit). Version 11.5 of the STRING database was used for the prediction of functional interactions of the OPN protein. **Results:** Serum OPN levels in patients with SLE without LN was 185.92 ± 259.86 pg/mL ($p=.0001$), and 311.08 ± 498.40 pg/mL in patients with SLE and LN ($p=0.121$), in the reference group was 72.59 ± 77.62 pg/mL ($p=0.002$). Additionally, were analyzed involved pathways that play a relevant role in renal tissue damage, finding interaction nodes and protein networks in relation to the OPN that could explain a final impact on LN. The pathways involved were selected based on relevance: fibronectin, integrin, protease, and collagen binding, as well as extracellular matrix binding. **Conclusions:** These results indicate that serum OPN levels were associated with patients with LES and LN, however we considered increase N sample. The molecular interaction of OPN was observed in the participation of fibronectin in glomerular injury is based on the internalization of IgG mediated by fibronectin, so this node of interaction is a fundamental piece, in turn the role of macrophages is mediated by the nodes of interaction of the binding to integrin and protease, correlating with the infiltration of diffuse macrophages in the kidneys developing LN.

Complex Traits Posters - Wednesday

PB1627. Serum proteomic signatures predicting unintentional weight loss in patients with chronic obstructive pulmonary disease

Authors:

J. Chiles¹, A. C. Wilson¹, A. Rocco¹, H. B. Rossiter², R. Casaburi², E. A. Regan³, R. Bowler³, C. P. Hersh⁴, M-L. N. McDonald¹; ¹Univ. of Alabama at Birmingham, Birmingham, AL, ²The Lundquist Inst., Los Angeles, CA, ³Natl. Jewish Hlth., Denver, CO, ⁴Brigham and Women's Hosp., Boston, MA

Abstract Body:

RATIONALE: Unintentional weight loss in patients with chronic obstructive pulmonary disease (COPD) is associated with increased mortality and morbidity. There are limited therapeutic options for patients once unintentional weight loss occurs. Populations with weight loss have been extensively characterized; however, there have been no studies of molecular risk factors that may precede the manifestation of weight loss in patients with COPD. Our goal was to identify proteomic predictors of unintended weight loss. **METHODS:** Data from 971 participants with COPD from the COPDGene study were analyzed: 842 controls who never reported unintentional weight loss and 129 cases who developed unintentional weight loss in the 5 years between visits 2 and 3. COPD was diagnosed using post-bronchodilator lung function. Serum was collected at visit 2 and used to generate proteomic data with SomaScan v4.0 (SomaLogic, Boulder, CO), which measures 4,776 human proteins. Linear regression models were built to test the association of each peptide with future unintentional weight loss while including age, sex, smoking status, and body mass index as covariates with false-discovery rate (FDR) correction. Gene-set enrichment analyses were conducted using peptides with an unadjusted p-value < 0.05. **RESULTS:** Age, sex, race, and baseline body-mass index were not different between groups, but participants reporting unintentional weight loss lost significantly more weight ($p < 0.001$). No single peptide met the FDR significance threshold, but 300 single peptides had an unadjusted p-value < 0.05. The top single peptides associated with future unintentional weight loss, each with FDR adjusted p-value of 0.15, were serum amyloid A1 (1.32 fold greater in cases), serum amyloid A2 (1.22 greater), complement factor 9 (1.08 greater), and C-reactive protein (1.34 greater); all are markers of inflammation. Gene set enrichment analysis of the up-regulated nominally significant peptides showed enrichment for eight pathways, including apoptosis (FDR $p = 0.018$). Down-regulated nominally significant peptides showed enrichment for 16 pathways, including adipogenesis (FDR $p < 0.001$). **DISCUSSION:** This study is the first to investigate peptide biomarkers that predict the development of unintentional weight loss in patients with COPD. Although no single peptide met significance criteria for association with future unintentional weight loss, there are patterns of increased inflammation and apoptosis with decreased adipogenesis. These are consistent with proposed mechanisms of unintentional weight loss in COPD. Further protein quantitative trait loci (pQTL) analyses are planned.

Complex Traits Posters - Thursday

PB1628. Sex dependent transcriptional changes in response to stress in patients with myalgic encephalomyelitis/chronic fatigue syndrome.

Authors:

J. Gamer¹, D. Van Booven², A. Joseph¹, M. Perez¹, O. Zarnowski¹, N. Klimas¹, E. Oltra³, L. Nathanson¹; ¹Dr. Kiran C. Patel Coll. of Osteopathic Med., Nova Southeastern Univ., Fort Lauderdale, FL, ²Univ Miami Miller Sch Med, Miami, FL, ³Catholic Univ. of Valencia, Valencia, Spain

Abstract Body:

Background The underlying mechanisms associated with the onset and progression of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS), a condition characterized by symptoms such as debilitating fatigue or easy fatigability, memory problems, muscle and joint pain, gastrointestinal issues, neurological problems, and hormonal imbalance remains unclear. **Objectives.** The main objectives of this research proposal are to evaluate 1) sex differences in circulating immune cells' gene expression in ME/CFS patients, 2) sex differences in ME/CFS patients in response to stress modeled by exercise challenge, and 3) the need to consider sex differences in diagnosis and treatment of ME/CFS. **Methods.** We evaluated differential gene (DG) expression (fold change > 1.5 in either direction, FDR < 0.1) by RNA-seq in circulating lymphocytes of 20 ME/CFS female patients, 14 ME/CFS male patients and 40 sex-, age- and BMI-matched healthy controls (HC) at T0 (baseline), T1 (exercise, maximal exertion) and T2 (4 hours of recovery after T1). After removing genes from sex chromosomes, analysis was done separately in men and women between ME/CFS patients and HC at each time point and in response to exercise (between time points for each group). **Results.** At baseline we found 20 DG between ME/CFS patients and HC in females and 160 DG in males. The largest difference in response to exercise in men and women with ME/CFS was between T2 and T1 time points. At these time points 395 DG were the same in men and women, 167 DG were only in men, and 174 only in women. Functional analysis revealed that at baseline in males with ME/CFS the most affected were regulation of leukocyte activation (-log₁₀(P)=6.99) and regulation of cell adhesion (-log₁₀(P)=5.26) compared to HC, while regulation of viral process (-log₁₀(P)=5.57) and cellular response to hypoxia (-log₁₀(P)=4.12) were the most affected in females with ME/CFS compared to HC. **Conclusion.** Identification of sex-specific biomarkers and therapeutic targets of ME/CFS can provide insight into sex-specific disease onset and progression, which will lead to optimal therapeutic intervention for such challenging disease.

Complex Traits Posters - Wednesday

PB1629*. Sex-specific analysis of rare variant associations with quantitative traits in the UK Biobank.

Authors:

R. Hoffing, A. Deaton, A. M. Holleman, L. Krohn, P. J. LoGerfo, P. Nioi, M. E. Plekan, C. Willis, L. Ward; Alnylam Pharmaceuticals, Cambridge, MA

Abstract Body:

Genetic variation can manifest differently between sexes and lead to substantial changes in health outcomes. Most genome-wide studies aggregate data by sex, which can mask meaningful and novel associations and exacerbate health disparities between the sexes. In this study, we perform a sex-stratified analysis of 100 quantitative traits in the UK Biobank across 363,661 exomes. We use burden tests to aggregate rare high-confidence predicted loss-of-function (pLOF) and damaging missense variants and test their association with these 100 traits. We elucidate several novel sex-specific associations, including *PREB* pLOF which associates with decreased apolipoprotein B in men ($P = 2.3 \times 10^{-6}$, -0.55 SD decrease). Additionally, we show several other novel associations including *WNT5B* pLOF + damaging missense with increased heel bone mineral density in women ($P = 8.9 \times 10^{-7}$, 0.28 SD increase), *MEN1* pLOF + damaging missense with increased calcium in women ($P = 1.9 \times 10^{-7}$, 0.40 SD increase), and *PDE3B* pLOF with decreased LDL cholesterol in women ($P = 5.2 \times 10^{-7}$, -0.21 SD decrease). These findings show instances of genetic loci with sexually dimorphic effects on complex traits and highlight potential new therapeutic strategies for diseases such as osteoporosis and cardiovascular disease with disparate effects across sexes.

Complex Traits Posters - Thursday

PB1630. Sex-specific and multiomic integration to enhance accuracy of peripheral blood biomarkers of major depressive disorder

Authors:

A. Delahaye-Duriez^{1,2}, A. MOKHTARI¹, E. Ibrahim³, A. Gloaguen⁴, D. P. Cohen¹, M. Derouin¹, I. Yalcin⁵, C. Marie-Claire⁶, B. Etain⁶, R. Belzeaux³, P-E. Lutz⁵; ¹NeuroDiderot UMR1141 INSERM Université Paris Cité, Paris, France, ²UFR SMBH, Université Sorbonne Paris Nord, Bobigny, France, ³Inst. of NeuroSci. of la Timone, Université Aix-Marseille, Marseilles, France, ⁴Inst. François Jacob, CNRGH - CEA, Université Paris-Saclay, Evry-Courcouronnes, France, ⁵Inst. des NeuroSci.s Cellulaires et Intégratives, UPR 3212, CNRS, Université de Strasbourg, Strasbourg, France, ⁶UMR-S1144 INSERM Université Paris Cité, Paris, France

Abstract Body:

Background/Objectives: Major depressive disorder (MDD also known as major depression or depression) is a psychiatric disorder that affects mood, behavior, and overall health, with a reduced life expectancy. Women are nearly twice as likely as men to be diagnosed with MDD. Clinical heterogeneity of MDD and the implication of multiple environmental risk factors contribute to the difficulty in identifying reliable biomarkers for diagnostic and therapeutic purposes. Here, we implemented a multi-omic integrative framework, designed to extract peripheral blood molecular signatures of MDD, and to evaluate the added-value of combining several types of omics data when classifying patients and controls. **Methods:** We analyzed transcriptomic (RNA-sequencing) and epigenomic processes (microRNA-sequencing and DNA methylation arrays) in blood samples from a well-characterized cohort of individuals with MDD (n=80), and healthy controls (n=89). We combined several strategies: first, we applied gene co-expression network and annotation enrichment analyses; second, we used supervised machine learning methods for matrix reduction; finally, we prioritized a subset of features (individual protein-coding genes, micro-RNAs, and methylation probes) that maximize the accuracy of the clustering of MDD patient and healthy controls. The performance of each combination of omics data and feature selection approach is evaluated through a 5 fold cross validation (CV) approach. **Results:** Gene co-expression network analyses identified gene modules that significantly associate with SNPs implicated in the risk for MDD in previous studies, and also show significant enrichment for differentially expressed genes, differential DNA methylation, and regulation by differentially expressed micro-RNAs. The joint dimension reduction approach allowed the identification of features that achieve good clustering of cases and controls (best accuracy=0.81). In both the pooled and male only cohort, the clustering based on the combination of mRNA and DNAm outperformed the single omic based approaches. While in the female mRNA based feature selection slightly outperformed the mRNA DNAm combination. **Conclusion:** Our analyses show a sex dependent biological signal of MDD that is important to consider when seeking to identify MDD biomarkers. These results also provide support for the hypothesis that, compared to single omics, integration of multi-omic datasets have the potential to significantly improve patient clustering, with implication for the development of MDD biomarkers.

Complex Traits Posters - Wednesday

PB1631. Sex-Specific associations of gene expression in brain with Alzheimer's disease pathology and cognitive performance.

Authors:

M. Clifton¹, M. Seto¹, R. F. Buckley^{2,3,4}, K. A. Gifford¹, A. L. Jefferson¹, P. L. De Jager^{5,6}, D. A. Bennett⁷, L. L. Barnes⁷, J. A. Schneider⁷, T. J. Hohman^{1,8}, L. Dumitrescu^{1,8}; ¹Vanderbilt Memory and Alzheimer's Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Dept. of Neurology, Massachusetts Gen. Hosp./Harvard Med. Sch., Boston, MA, ³Ctr. for Alzheimer's Res. and Treatment, Dept. of Neurology, Brigham and Women's Hosp./Harvard Med. Sch., Boston, MA, ⁴Melbourne Sch. of Psychological Sci., Univ. of Melbourne, Melbourne, Australia, ⁵Ctr. for Translational & Computational Neuroimmunology, Dept. of Neurology, Columbia Univ. Med. Ctr., New York, NY, ⁶Cell Circuits Program, Broad Inst., Cambridge, MA, ⁷Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, ⁸Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Alzheimer's disease (AD) disproportionately affects women, who make up two-thirds of all prevalent AD. While sex differences in AD neuropathology, the response to pathology, and the genetic predictors of clinical AD have been well described, sex differences in the brain transcriptomic signatures of AD phenotypes have not been fully characterized. We leveraged bulk RNA-sequence data from three brain regions (dorsolateral prefrontal cortex [DLPFC], posterior cingulate cortex [PCC], and caudate nucleus) from 2201 samples and 994 participants from the Religious Orders Study and Rush Memory and Aging Project (35% male; mean age at death 87.6 years). Sex-stratified and sex-interaction (int) regression models assessed sex-specific transcript associations with amyloid and tau burden, along with global cognition. Age at death, latency to death, and post-mortem interval were included as covariates. Sex-specific genes were defined as related to a trait in one sex (FDR-corrected $p < 0.05$) but not in the other sex and that showed evidence of a sex-modifying effect (int $p < 0.05$). Of the more than 20,000 significant autosomal gene expression associations with the three AD endophenotypes, 9% were sex-specific. Given the larger sample size, it was unsurprising that we observed more female-specific (8%) compared to male-specific (1%) effects. A number of genes showed particularly strong effects among females, including *BDNF-ASI* with tau tangles in the PCC ($P_{\text{men}}=0.93$, $P_{\text{women}}=3.03 \times 10^{-7}$, $P_{\text{int}}=7.25 \times 10^{-4}$), *LRIG3* with amyloid in the DLPFC ($P_{\text{men}}=0.11$, $P_{\text{women}}=6.99 \times 10^{-11}$, $P_{\text{int}}=0.012$), and *TMX4* in multiple tissues with cognition ($P_{\text{men}}=0.89$, $P_{\text{women}}=1.4 \times 10^{-3}$, $P_{\text{int}}=0.04$), tangles ($P_{\text{men}}=0.30$, $P_{\text{women}}=2.4 \times 10^{-3}$, $P_{\text{int}}=0.01$), and amyloid ($P_{\text{men}}=0.57$, $P_{\text{women}}=3.7 \times 10^{-3}$, $P_{\text{int}}=0.02$). Additionally, 17 genes showed male-specific associations with at least two traits including *HIP1R* ($P_{\text{men}}=5.7 \times 10^{-5}$, $P_{\text{women}}=0.07$, $P_{\text{int}}=0.03$), *CD83* ($P_{\text{men}}=8.4 \times 10^{-4}$, $P_{\text{women}}=0.45$, $P_{\text{int}}=0.04$), and *ZFHX4-ASI* ($P_{\text{men}}=1.2 \times 10^{-3}$, $P_{\text{women}}=0.45$, $P_{\text{int}}=1.4 \times 10^{-3}$). Similar sex-specific patterns with AD were observed in independent AMP-AD cohorts for *LRIG3*, *BDNF-ASI*, and *ZFHX4-ASI* (<https://agora.adknowledgeportal.org/>). A formal replication analysis is forthcoming, as is an examination of X chromosome genes. Our results highlight sex-specific transcriptomic associations with AD phenotypes, including genes along protein misfolding and neurotrophic signaling among females and macrophage markers among males. These findings highlight the exciting potential of precision medicine approaches that consider sex-specific biological pathways to select new targets for mechanistic evaluation.

Complex Traits Posters - Thursday

PB1632. Sex-specific genetic loci linked to early and late onset type 2 diabetes.

Authors:

J. Berumen¹, L. Orozco², H. Gallardo^{3,4}, R. E. Benuto⁵, E. Ramos-Martinez¹, F. Rivas¹, H. García-Ortiz², M. Marín-Medina⁵, E. Barrera¹, E. Juárez-Torres⁵, L. León-Angulo⁵, A. Alvarado-Silva⁵, E. Mendoza-Tenorio¹, L. Martínez-Juárez⁴, J. Lomelín-Gascón⁴, A. Montoya⁴, J. Ortega-Montiel⁴, R. Tapia-Conyer^{4,1}; ¹Univ. Natl. Autónoma de México, México, Mexico, ²Natl. Inst. of Genomic Med., México, Mexico, ³Univ. de Guadalajara, México, Mexico, ⁴Fundación Carlos Slim, México, Mexico, ⁵Laboratorio Huella Génica, México, Mexico

Abstract Body:

PurposeTo investigate the effect of sex and age on the timing of a type 2 diabetes (T2D) diagnosis and the influence T2D-related genes, parental history of T2D, and obesity have on T2D development.**Methods**In total, 1012 T2D cases and 1008 healthy subjects were selected from the Diabetes in Mexico Study database. Participants were stratified by sex and age at T2D diagnosis (early, ≤ 45 years; late, ≥ 46 years). The percentage contribution (R^2) of 71 SNPs located in T2D-related genes, parental history of diabetes, and obesity (body mass index [BMI] and waist-hip ratio [WHR]) on T2D development was calculated using univariate and multivariate logistic regression models.**Results** T2D-related genes influenced T2D development greatest in males who were diagnosed early ($R^2 = 23.5\%$; females diagnosed early, $R^2 = 13.5\%$; males and females diagnosed late, $R^2 = 11.9\%$ and $R^2 = 7.3\%$, respectively). With an early diagnosis, insulin production genes were more influential in males (76.0% of R^2) whilst peripheral insulin resistance genes were more influential in females (52.3% of R^2). With a late diagnosis, insulin production genes from the chromosome region 11p15.5 notably influenced males while peripheral insulin resistance and inflammation genes notably influenced females. Influence of parental history was higher among those diagnosed early (males, 19.9%; females, 17.5%) versus late ($\leq 6.4\%$). Unilateral maternal T2D history was more influential than paternal T2D history. BMI influenced T2D development for all, while WHR exclusively influenced males.**Conclusions** The influence of T2D-related genes, maternal T2D history, and fat distribution on T2D development was greater in males than females.

Complex Traits Posters - Wednesday

PB1633. Sex-specific genetic predictors of brain amyloidosis in Alzheimer's disease.

Authors:

T-C. Wang^{1,2}, D. B. Archer^{1,2}, M. Ali^{3,4}, A. J. Lee^{5,6,7}, Y. Gao^{5,6,7}, M. Bilgel⁸, E. Mormino⁹, D. Goldstein^{5,6,7}, T. Harrison¹⁰, C. D. Engelman¹¹, A. J. Saykin^{12,13}, C. D. Whelan¹⁴, Alzheimer's Disease Neuroimaging Initiative (ADNI), A4 Study Team, Alzheimer's Disease Genetics Consortium (ADGC), Alzheimer's Disease Sequencing Project (ADSP), J. Z. Liu¹⁴, W. Jagust¹⁰, M. Albert¹⁵, S. C. Johnson¹⁶, H. S. Yang^{17,18}, K. A. Johnson^{17,18}, P. Aisen¹⁹, S. M. Resnick⁸, R. Sperling^{17,18}, P. L. De Jager^{5,6,7,20,21}, J. Schneider²², D. A. Bennett²³, B. Vardarajan^{5,6,7,24}, R. Mayeux^{5,6,7,24}, C. Cruchaga^{3,4,25,26,27}, T. J. Hohman^{1,2}, L. Dumitrescu^{1,2}; ¹Vanderbilt Memory and Alzheimer's Ctr., Dept. of Neurology, Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Dept. of Psychiatry, Washington Univ., St. Louis, MO, ⁴NeuroGenomics and Informatics. Washington Univ., St. Louis, MO, ⁵Dept. of Neurology, Columbia Univ. Med. Ctr., New York, NY, ⁶Dept. of Neurology, The New York Presbyterian Hosp., New York, NY, ⁷Taub Inst. for Res. on Alzheimer's Disease and The Aging Brain, Columbia Univ. Med. Ctr., New York, NY, ⁸Lab. of Behavioral NeuroSci., Natl. Inst. on Aging, NIH, Baltimore, MD, ⁹Dept. of Neurology and Neurological Sci., Stanford Univ., Stanford, CA, ¹⁰Helen Wills NeuroSci. Inst., Univ. of California, Berkeley, CA, ¹¹Dept. of Population Hlth.Sci., Univ. of Wisconsin, Sch. of Med. and Publ. Hlth., Madison, WI, ¹²Dept. of Radiology and Imaging Sci., Ctr. for Neuroimaging, Sch. of Med., Indiana Univ., Indianapolis, IN, ¹³Dept. of Med. and Molecular Genetics, Sch. of Med., Indiana Univ., Indianapolis, IN, ¹⁴Res. and Early Dev., Biogen Inc., Cambridge, MA, ¹⁵Dept. of Neurology, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ¹⁶Alzheimer's Disease Res. Ctr., Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, ¹⁷Dept. of Neurology, Massachusetts Gen. Hosp., Boston, MA, ¹⁸Ctr. for Alzheimer Res. and Treatment, Dept. of Neurology, Brigham and Women's Hosp., Boston, MA, ¹⁹Alzheimer's Therapeutic Res. Inst., Keck Sch. of Med., Univ. of Southern California, San Diego, CA, ²⁰Ctr. for Translational and Computational Neuroimmunology, Dept. of Neurology, Columbia Univ. Med. Ctr., New York, NY, ²¹Cell Circuits Program, Broad Inst., Cambridge, MA, ²²ush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, ²³Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, ²⁴The Inst. for Genomic Med., Columbia Univ. Med. Ctr., New York, NY, ²⁵Knight Alzheimer's Disease Res. Ctr., Washington Univ., St. Louis, MO, ²⁶Hope Ctr. for Neurologic Diseases, Washington Univ., St. Louis, MO, ²⁷Dept. of Genetics, Washington Univ. Sch. of Med., St. Louis, MO

Abstract Body:

Nearly two-thirds of Americans with Alzheimer's disease (AD) are women, yet the role of sex-specific genetic drivers of AD or AD neuropathology has not been well-characterized. Amyloid positron emission tomographic (PET) imaging allows for *in vivo* detection of amyloid plaques in the brain, a neuropathological feature of AD. While amyloid PET levels have been leveraged as AD biomarkers in previous GWAS, sex-stratified models have not been incorporated. Hence, we analyzed sex-specific genetic associations with amyloid PET measured in 4,871 non-Hispanic white participants from seven cohorts of aging and AD ($N_{\text{men}}=2,137$, $N_{\text{women}}=2,734$, $\text{Age}_{\text{both sexes}}$: mean=72.37 years, SD=6.66). Amyloid PET was harmonized by applying a gaussian mixture model using modified nonlinear distributional mapping. We then performed sex-stratified and sex-interaction GWAS within each dataset and meta-analyzed. Covariates included age and the first five principal components for the sex-stratified GWAS and sex additionally for the sex-interaction GWAS. Outside of the expected association within the *APOE* locus, we identified a novel locus on chromosome 6 (index SNP rs544047, closest gene: *ROSL*, MAF=0.41) that was significantly associated with amyloid PET in men ($\beta=-0.17$, $p_{\text{men}}=3.86 \times 10^{-8}$) but not women ($\beta=-0.01$, $p_{\text{women}}=0.59$; $\text{pint}=2.45 \times 10^{-5}$). This top variant at the locus is an eQTL for *VGLL2* in

adipose tissue (GTEx, $p=2.2 \times 10^{-5}$) and *FAM26F* in blood (eQTLgen, $p=4.53 \times 10^{-10}$) and a methylation QTL in prefrontal cortex for *ROSI* (Brain xQTLServe, $\beta=-0.26$, $p=4.96 \times 10^{-10}$). *ROSI* encodes a receptor tyrosine kinase involved in chromosomal rearrangements, and some *ROSI* fusions were found in gliomas. Interestingly, *ROSI* rearrangement in cancer has shown evidence of a sex bias, and *ROSI* is downregulated in the AD brain, making it a fascinating sex-specific candidate gene. In our previous sex-agnostic GWAS we identified a genome-wide locus in *RBFox1*. Here, we observed nominal evidence that the association may be driven by females (rs56081887; $\beta_{\text{men}}=0.07$, $p_{\text{men}}=0.21$; $\beta_{\text{women}}=0.20$, $p_{\text{women}}=7.96 \times 10^{-6}$), though sex-interaction was not significant ($p_{\text{int}}=0.11$). In validation analyses, higher expression of both *ROSI* and *RBFox1* in prefrontal cortex was associated with lower amyloid plaque burden in women ($p=1.64 \times 10^{-4}$; $p=2.34 \times 10^{-4}$) but not men ($p=0.80$, $p_{\text{int}}=1.57 \times 10^{-2}$; $p=0.56$, $p_{\text{int}}=0.14$) at autopsy. Our results highlight the utility of sex-aware genome-wide analysis for identifying novel sex-specific genetic contributors to disease and the critical need for larger sample sizes to explore sex-specific genetic effects. Future work will seek replication of the novel male-specific *ROSI* locus.

Complex Traits Posters - Thursday

PB1634*. Sex-specific genetic predictors of memory performance in older adults.

Authors:

J. M. Eissman^{1,2}, D. B. Archer^{1,2}, S. Mukherjee³, M. L. Lee³, S-E. Choi³, P. Scollard³, E. H. Trittschuh^{4,5}, J. B. Mez⁶, W. S. Bush⁷, C. D. Engelman⁸, Q. Lu^{9,10}, D. W. Fardo^{11,12}, K. F. Widaman¹³, R. F. Buckley^{14,15,16}, E. C. Mormino¹⁷, B. W. Kunkle¹⁸, A. C. Naj^{19,20}, L. R. Clark⁸, K. A. Gifford¹, Alzheimer's Disease Neuroimaging Initiative (ADNI), A4 Study Team, Alzheimer's Disease Genetics Consortium (ADGC), Alzheimer's Disease Sequencing Project (ADSP), M. L. Cuccaro¹⁸, J. M. Vance¹⁸, C. Cruchaga²¹, M. A. Pericak-Vance¹⁸, L. A. Farrer^{6,22,23}, L-S. Wang²⁰, G. D. Schellenberg²⁰, J. L. Haines⁷, A. L. Jefferson¹, S. C. Johnson⁸, W. A. Kukull³, M. S. Albert²⁴, C. D. Keene²⁵, A. J. Saykin²⁶, E. B. Larson^{3,27}, R. A. Sperling¹⁴, R. Mayeux^{28,29,30}, P. M. Thompson³¹, E. R. Martin¹⁸, D. A. Bennett³², L. L. Barnes³², J. A. Schneider³², P. K. Crane³, T. J. Hohman^{1,2}, L. Dumitrescu^{1,2}; ¹Vanderbilt Memory & Alzheimer's Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Dept. of Med., Univ. of Washington, Seattle, WA, ⁴Dept. of Psychiatry and Behavioral Sci., Univ. of Washington Sch. of Med., Seattle, WA, ⁵VA Puget Sound Hlth.Care System, GRECC, Seattle, WA, ⁶Dept. of Neurology, Boston Univ. Sch. of Med., Boston, MA, ⁷Cleveland Inst. for Computational Biology, Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, ⁸Dept. of Population Hlth.Sci., Sch. of Med. and Publ. Hlth., Univ. of Wisconsin-Madison, Madison, WI, ⁹Dept. of Statistics, Univ. of Wisconsin-Madison, Madison, WI, ¹⁰Dept. of Biostatistics and Med. Informatics, Univ. of Wisconsin-Madison, Madison, WI, ¹¹Dept. of Biostatistics, Coll. of Publ. Hlth., Univ. of Kentucky, Lexington, KY, ¹²Sanders-Brown Ctr. on Aging, Univ. of Kentucky, Lexington, KY, ¹³Univ. of California at Riverside, Riverside, CA, ¹⁴Dept. of Neurology, Massachusetts Gen. Hosp./Harvard Med. Sch., Boston, MA, ¹⁵Ctr. for Alzheimer's Res. and Treatment, Dept. of Neurology, Brigham and Women's Hosp./Harvard Med. Sch., Boston, MA, ¹⁶Melbourne Sch. of Psychological Sci., Univ. of Melbourne, Melbourne, Australia, ¹⁷Dept. of Neurology and Neurological Sci., Stanford Univ., Stanford, CA, ¹⁸John P. Hussman Inst. for Human Genomics, Miller Sch. of Med., Miami, FL, ¹⁹Dept. of Biostatistics, Epidemiology, and Informatics, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, ²⁰Penn Neurodegeneration Genomics Ctr., Dept. of Pathology and Lab. Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, ²¹Dept. of Psychiatry, Washington Univ. Sch. of Med., St. Louis, MO, ²²Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, ²³Dept. of Med. (BioMed. Genetics), Boston Univ. Sch. of Med., Boston, MA, ²⁴Dept. of Neurology, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ²⁵Dept. of Lab. Med. and Pathology, Univ. of Washington, Seattle, WA, ²⁶Dept. of Radiology and Imaging Sci., Indiana Univ. Sch. of Med., Indianapolis, IN, ²⁷Kaiser Permanente Washington Hlth.Res. Inst., Seattle, WA, ²⁸Dept. of Neurology, Columbia Univ., New York, NY, ²⁹The Taub Inst. for Res. on Alzheimer's Disease and The Aging Brain, Columbia Univ., New York, NY, ³⁰The Inst. for Genomic Med., Columbia Univ. Med. Ctr. and The New York Presbyterian Hosp., New York, NY, ³¹Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ³²Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL

Abstract Body:

Two-thirds of clinical Alzheimer's disease (AD) cases are women compared to men, and women have more AD neuropathology at autopsy. Additionally, AD biomarker-positive women have steeper cognitive decline compared to AD biomarker-positive men. Sex-specific genetic factors may contribute to these observed sex differences, yet large-scale sex-aware genomic studies on late-life cognition are lacking. Thus, we sought to identify sex-specific genetic drivers of memory performance. Leveraging psychometric techniques, we harmonized item-level cognitive data into a continuous memory composite score that was scaled across four cohorts of cognitive aging ($N_{\text{men}}=9,618$, $N_{\text{women}}=13,199$, $\text{Age}_{\text{Both Sexes}}$:

mean=75 years). We performed sex-stratified and sex-interaction GWAS on baseline and longitudinal memory performance in non-Hispanic white (NHW; $N_{\text{total}}=20,205$) and in non-Hispanic black (NHB; $N_{\text{total}}=2,612$) participants. GWAS results were meta-analyzed both within and across each ancestry group. Among NHW men, we identified a locus on chromosome 18 associated with baseline memory performance (rs2590395; $P_{\text{men}}=2.15 \times 10^{-8}$, $P_{\text{women}}=0.65$, $P_{\text{sex-interaction}}=8.10 \times 10^{-7}$). The top variant in the locus, rs2590395, was an eQTL for *SERPINB2* and *SERPIN10* in blood (eQTLGen) and was a methylation QTL (xQTLServe) at two sites just downstream of the locus. Interestingly, our group previously identified a genome-wide significant *SERPIN* locus, *SERPINB1*, on chromosome 6, that was associated with CSF A β 42 burden among women. Among NHB participants, we did not identify any genome-wide significant loci, and no variants in linkage disequilibrium with the significant NHW male locus were significant in NHB men (rs2590395; $P_{\text{men}}=0.42$). In the Trans-Ancestry meta-analysis, among women, we identified two loci associated with longitudinal memory performance. First, we identified a chromosome 20 locus (rs2427384; $P_{\text{men}}=0.75$, $P_{\text{women}}=3.09 \times 10^{-8}$, $P_{\text{sex-interaction}}=4.33 \times 10^{-4}$), whereby the top variant, rs2427384, was an eQTL in the hippocampus for *LAMA5* (BrainSeq), a gene encoding extracellular matrix glycoproteins. Second, we observed an association at a known AD locus, *BINI*, that was genome-wide significant among women, but not among men (rs6733839; $P_{\text{men}}=7.60 \times 10^{-5}$, $P_{\text{women}}=3.29 \times 10^{-8}$) with no evidence for a true sex-specific effect ($P_{\text{sex-interaction}}=0.51$). Overall, our results highlight many promising sex-specific genetic predictors and candidate genes that may contribute to sex differences in late-life cognition. We are excited to follow-up on these results, including the opportunity to incorporate additional cognitive domains into our genetic analyses.

Complex Traits Posters - Thursday

PB1636. Sex-specific risk scores for coronary artery disease using machine learning models.

Authors:

Q. Ye¹, S. Gagliano Taliun²; ¹Montreal Heart Inst., Montreal, QC, Canada, ²Montréal Heart Inst./Université de Montréal, Montreal, QC, Canada

Abstract Body:

Coronary artery disease (CAD) affects millions of adults worldwide, and it affects women differently than men. Predicting CAD risk in a sex-specific manner using non-genetic and genetic factors is of great importance to moving towards personalized medicine and to improving human health. In recent years, polygenic risk scores for CAD built on linear models using genetic variants combined with phenotypic features as covariates have been proposed, and the area under the receiver operating characteristic curve (AUC) has reached 0.80. Here, we trained and compared two linear models (logistic regression (LR) and elastic net (EN)) and two non-linear models (support vector classifier (SVC) with non-linear kernels and multi-layer perceptron (MLP)) to predict CAD risk in the European-ancestry subset of the UK Biobank dataset (N cases = 14,740 with 22.30% female, and N controls= 299,580 with 55.25% female). We trained models in three sex categories (both sexes, female-only and male-only) considering various phenotypic feature sets and published polygenic scores. AUC was used to measure model performance. The highest AUCs for LR, EN, SVC, and MLP were 0.81, 0.79, 0.83 and 0.83, respectively with both sexes modeled together. The highest values were 0.76, 0.75, 0.79, and 0.78 on the female only dataset and 0.76, 0.75, 0.79, and 0.79 on the male only dataset. There was no significant decrease in AUC when a model was trained on the dataset with one sex and then tested on the opposite sex. We concluded that the non-linear models (SVC and MLP) achieved the highest AUCs, and both may out-perform linear models (LR and EN). Machine learning models can serve as an important tool for building risk scores for CAD.

Complex Traits Posters - Wednesday

PB1637. Sex-stratified meta-analysis of age-related cognitive decline across neurocognitive domains

Authors:

V. Acharya¹, K-H. Fan¹, B. Snitz², M. Ganguli³, S. Dekosky⁴, O. Lopez², E. Feingold¹, M. Kamboh¹; ¹Dept. of Human Genetics, Univ. of Pittsburgh Sch. of Publ. Hlth., Pittsburgh, PA, ²Dept. of Neurology, Sch. of Med., Univ. of Pittsburgh, Pittsburgh, PA, ³Dept. of Epidemiology, Univ. of Pittsburgh, Pittsburgh, PA, ⁴Dept. of Neurology, Coll. of Med., Univ. of Florida, Gainesville, FL

Abstract Body:

Cognitive aging is a huge public health concern due to increased elderly population worldwide. Age is the largest risk factor for cognitive decline in older adults. Increasing evidence suggests that cognitive decline across several cognitive domains exhibits sex-specific effects. Despite these sex-related differences, there is limited understanding of sex-specific genetic architectures of cognitive aging. To identify genetic variants involved in cognitive decline in males and females, we performed sex-stratified meta-analysis on 1499 males and 1569 females, ranging in age from 65 to 95 years and free of dementia at baseline, from three different prospective cohorts: Ginkgo Evaluation of Memory Study (GEMS), Monongahela-Youghiogheny Healthy Aging Team (MYHAT) and Monongahela Valley Independent Elders Survey (MoVIES). Longitudinal follow-up with cognitive evaluations ranged from 6 to 12 years. We calculated z scores across five cognitive domains: attention, memory, executive function, language, and visuospatial function, as well as global cognitive function score that was derived by averaging the performance across all tests within the five cognitive domains. Inter-individual domain-specific slopes over time were then generated by fitting a linear mixed effect model. We identified a genome-wide significant (GWS) locus for memory decline in males near *MIR1269A* on chromosome 4 ($p=2.94E-08$), a region which has previously been implicated with education attainment. We also observed a sub-threshold GWS locus for attentional decline in females near *GALR1* and *LINC01029* on chromosome 18 ($p=8.08E-08$). In addition, we identified multiple suggestive loci in males or females for the following domains: 4 loci for decline of attention in females (p from $p=3.41E-07$ to $p=9.71E-07$); 2 loci for decline of memory in females (p from $5.37E-07$ to $7E-07$); one locus for global cognitive function decline in females ($p=6.26E-07$); 2 loci for decline of memory in males (p from $4.22E-07$ to $8.21E-07$); 2 loci for decline of language in males (p from $1.22E-07$ to $8.96E-07$); 2 loci for global cognitive function decline in males (p from $9.46E-07$ to $9.89E-07$) and one locus for decline of visuospatial function in males ($p=2.35E-07$). The latter locus (*JAZF1*) has previously been implicated with Alzheimer's disease risk. These novel, exploratory data expand the landscape of the genetics of neurocognition and offer sex-specific findings of cognitive aging across the neurocognitive domains.

Complex Traits Posters - Thursday

PB1638. *SHANK3* domain specific functional characterization of a spectrum of neurodevelopmental disorders phenotypes.

Authors:

N. Kosaji¹, S. Al Shaibani¹, N. Nassir¹, B. Ashraf¹, A. Ahmed¹, M. Uddin^{1,2}; ¹MBRU (Mohammed bin Rashid Univ. of Med. and Hlth.Sci.s), Dubai, United Arab Emirates, ²GenomeArc (Cellular Intelligence (Ci) Lab, GenomeArc Inc.), Toronto, ON, Canada

Abstract Body:

Autism spectrum disorder (ASD) is a multifaceted neuropsychiatric condition characterized by impairments in social interaction, communication, and behavior. Previous experimental research and clinical studies established a correlation between *SHANK* protein impairment and the onset of neurodevelopmental disorders (NDD), particularly autism. *SHANK* proteins function as master organizers of the postsynaptic density, owing to their ability to form multimeric complexes with postsynaptic receptors, signaling molecules and cytoskeletal proteins. From this gene family, *SHANK3*, an excitatory synapse scaffolding protein, was found to be primarily involved in NDD pathogenesis. To explore those interactions, our research aimed to identify the most frequent *SHANK* mutations to mimic their pathophysiology using *in vitro* models with CRISPR/Cas9 technology. Furthermore, deducing distributional trends provides valuable insight into the mutations' characteristics, vital for understanding gene interactions and disease pathogenesis. In our study, we conducted a comprehensive meta-analysis of *SHANK* mutations in NDD cases by constructing a mutation database from published genomic studies on *SHANK* genes. For stratification analysis, we collected *SHANK* mutation clinical phenotypes to identify variable trends. The mutations were categorized into ASD, Asperger's syndrome, Schizophrenia, undiagnosed cases of intellectual disability, and others including ADHD, Phelan Mcdermid, and Rhett's Syndrome. Initially we congregated 1482 cases across 433 *SHANK* mutations. We then filtered the data for exonic and splice junction variants and investigated their pathogenicity following ACMG guidelines. After quality control, 297 proband cases across 211 distinct mutations were retained. Our analysis identified trends in mutation distribution across gender, type, location, pathogenicity, and disease. Our findings show that *SHANK* mutations predominantly affect males and are primarily involved in ASD amongst other NDDs. Additionally, the majority of *SHANK1/2* mutations are paternally inherited while *SHANK3* mutations are mostly *de novo*. Interestingly, the bulk of reported pathogenic *SHANK* mutations are in *SHANK3*. Since the results suggested the particular involvement of *SHANK3* into the pathogenesis of NDDs, we intend on exploring this phenotypic heterogeneity by CRISPR/Cas9 induced recurrent mutations within different *SHANK3* domains to model pluripotent stem cell differentiated knockout neurons. This model system will enable us to quantify morphological and molecular domain specific phenotypes that contribute into the pathophysiology of NDDs.

Complex Traits Posters - Wednesday

PB1639. Shared genetic basis informs the roles of polyunsaturated fatty acids in brain disorders

Authors:

H. Xu¹, Y. Sun¹, M. Francis², K. Ye^{1,2}; ¹Dept. of Genetics, Univ. of Georgia, Athens, GA, ²Inst. of Bioinformatics, Univ. of Georgia, Athens, GA

Abstract Body:

Polyunsaturated fatty acids (PUFAs) are associated with brain disorders, such as major depression (MDD) and Attention-Deficit Hyper-activity Disorder (ADHD). Supplementation of omega-3 PUFAs was also suggested to reduce psychotic symptoms, such as anxiety and anorexia. However, most previous findings were based on observational associations. Here, we systematically investigate the shared genetic basis between PUFAs and brain disorders, aiming to illuminate their interacting mechanisms. We will perform three major analyses using genome-wide association study summary statistics of 11 PUFA traits (n=114,999) and 20 brain disorders (n=9,725~762,917). First, we performed genetic correlation analysis with LDSC. We confirmed a previous observation that there are positive genetic correlations across most brain disorders. Among PUFA traits, there are strong correlations between PUFA, omega-6, linoleic acid (LA), omega-3, docosahexaenoic acid (DHA), and the percentage of omega-3 in total fatty acids (omega-3%), with the strongest correlation between LA and omega-6 ($r_g=0.95$, $P < 4.55 \times 10^{-3}$). Notably, for total PUFAs, omega-6, and LA, there are negative correlations between their absolute measurements and their relative percentages in total fatty acids. With pairwise analysis between PUFA traits and brain disorders, we identified a large number of significant correlations. Six PUFA phenotypes (PUFA%, omega-6%, LA%, omega-3%, DHA% and DHA) were negatively correlated with substance use disorders (opioid dependence: $r_g=-0.4\sim-0.23$; $P < 0.05$; alcohol dependence: $r_g=-0.3\sim-0.18$, $P < 0.05$; cannabis use disorder: $r_g=-0.27\sim-0.20$, $P < 0.001$), followed by ADHD ($r_g=-0.34\sim-0.23$, $P < 2.27 \times 10^{-4}$), Post Traumatic Stress Disorder ($r_g=-0.32\sim-0.16$, $P < 0.05$), anxiety ($r_g=-0.23\sim-0.22$, $P < 0.05$), insomnia ($r_g=-0.20\sim-0.11$, $P < 0.05$), MDD ($r_g=-0.19\sim-0.1$, $P < 0.05$) and neuroticism ($r_g=-0.14\sim-0.08$, $P < 0.05$). In addition, these six PUFA phenotypes were positively correlated with compulsive behaviors (obsessive-compulsive disorder: $r_g=0.14\sim0.3$, $P < 0.05$; anorexia nervosa: $r_g=0.17\sim0.28$, $P < 2.27 \times 10^{-4}$), followed by schizophrenia and alcohol consumption. The moderate to strong genetic correlations indicate a shared genetic basis between PUFAs and brain disorders. We are in the process of estimating the numbers of shared and unique causal variants using MiXeR. Moreover, we will apply colocalization analysis to pinpoint specific shared genes. This study will improve our understanding of the roles of PUFAs in brain disorders and inform their interacting biological pathways.

Complex Traits Posters - Wednesday

PB1641. Shriners Children's international program on the genetics of rare pediatric disorders

Authors:

K. Shazand¹, A. Gustafson¹, A. Quitadamo¹, X. Liu², V. Bijanki³, M. Lalande⁴; ¹Shriners Children's Genomics Inst., Tampa, FL, ²Univ. of South Florida, Tampa, FL, ³Shriners Children's St. Louis, St. Louis, MO, ⁴Shriners Hosp. for Children, Tampa, FL

Abstract Body:

The Shriners Children's recently initiated an ambitious large-scale research program, namely the Shriners Precision Medicine and Genomics (SPMG) program. The aim of SPMG is to build a comprehensive genetic database of over 5,000 trios (patients and their biologic parents) by Whole Genome Sequencing (WGS) by short and long read platforms. This work aims to first characterize the etiology of these disorders based on a statistically solid approach, and use this knowledge to improve patients' lives with highly specific diagnostic tools, more reliable treatment protocols and better therapeutic approaches. The pediatric disorders that we treat at our hospitals and clinics are neurologic, musculoskeletal and orthopedic, as well as burns and cleft lip and palate. In the last 2 years, we were able to get all 22 Shriners centers onboard for patient recruitment and accumulate over 4,000 samples of highly diverse ethnic descents, of which ~3,000 are sequenced. Due to its higher prevalence (3%), the adolescent idiopathic scoliosis (AIS) cohort reached a considerable number, allowing identification of some of the genes and pathways potentially involved in the onset and development of AIS during the embryonic and perinatal phases. We are developing animal models to functionally validate and isolate the most clinically relevant ones. The most recent results will be reported at the conference. In addition, the study of several other disorders such as cerebral palsy, arthrogryposis multiplex congenital, toe walking and pain management of burns have been ongoing in parallel and exciting developments are underway. Altogether, these projects have been successful in delivering some of the underlying genetic structure of the disorders under study and should contribute to the collective efforts in supporting patients across the world.

Complex Traits Posters - Thursday

PB1642. Significant heritability enrichment for Schizophrenia located in single neuron electrophysiology genes.

Authors:

X. Li, H. Hu, N. Johansen, L. Walker, G. Quon; Univ. of California, Davis, Davis, CA

Abstract Body:

The impact of psychiatric GWAS variants on cellular phenotypes such as single neuron electrophysiology (ephys) is unknown. We leveraged single neuron Patch-seq data (matched RNAseq and patch-clamping) to perform a genome-wide search for genes whose expression levels are correlated with different quantitative measurements of single neuron ephys in interneurons.

Globally, we identified 30 ephys features that can be robustly predicted (Spearman $\rho > 0.5$) from gene expression across mouse interneurons. Surprisingly, models trained on individual classes of interneurons (Sst, Pvalb, Lamp5, Vip) were highly cell type specific: models trained on one cell type but tested on another typically yielded a drop in 72% on average, suggesting distinct sets of ephys genes are responsible for control of the same ephys feature across different interneuron types. Unexpectedly, we found that of the genes predictive of ephys, ion channels only constitute a small fraction (average $< 4.3\%$) of them. Instead, we found significantly overrepresentation of adhesion molecules with known roles modulating ion channel activity, suggesting more indirect influence on ephys properties of neurons. We also found novel functional diversity among cell adhesion molecules such as Kirrel3 and Cxadr, in which individual adhesion genes established different ephys properties in different interneurons, despite similar expression levels across interneurons. This potential for cell adhesion function switching is previously unappreciated but was prevalent in our analysis.

To quantify the heritability of psychiatric diseases with respect to neuron ephys, we performed partitioned heritability analysis of different ephys gene sets and found the first action potential dV/dt genes explained significant heritability specifically for Schizophrenia (permuted $P < 0.001$). Co-localization analysis found four genes whose expression is likely modulated by SCZ GWAS variants, and are ephys genes: RGS6, STK4, TCTN1 and RBM26. Co-expression partners of these genes were significantly different between SCZ cases versus controls, suggesting changes in function between cases and controls. Further evidence gathered from epigenetic and chromatin accessibility data highly support two potential SCZ causal variants that may affect Sst neuron ephys. Overall, our work demonstrates a potential role of single neuron electrophysiology in mediating the genetic risk of psychiatric disorders.

Complex Traits Posters - Wednesday

PB1643. Single cell patch-seq in understanding the regenerative ability of cortical spinal motor neurons

Authors:

H. Kim, J. M. Saikia, A. Moore, E. Ha, K. Cervantes, D. Lusk, B. Zheng; Univ. OF CALIFORNIA, SAN DIEGO, San Diego, CA

Abstract Body:

Spinal cord injury (SCI) is a severe condition that results in loss of function in mobility. The corticospinal tract (CST) is a clinically important target for functional recovery after SCI. Multiple molecular pathways including the Pten/mTOR signaling pathway have been revealed to regulate axon regeneration and sprouting from the CST. However, among diverse populations of CST neurons, only a subset regenerates axons following molecular intervention and the number of regenerating neurons declines with age. Here, we have performed single cells RNA-seq using Patch-seq after retrograde tracer injection in animals with Pten/mTOR pathway modification. Through differential gene expression and pathway / network analyses, we identified known and new potential regulators of CST regeneration and found that regenerating transcriptomes differentially map to previously defined neuronal clusters based on single cell seq data.

Complex Traits Posters - Thursday

PB1644. Single nucleotide variants genotyping of IL-6, IL-1 β , TNF- α genes promoter region and its protein levels associated to inflammatory response in obstructive sleep apnea and periodontal disease

Authors:

F. Sir-Mendoza¹, M. Rey¹, F. Gonzalez², L. Otero³; ¹Univ. Natl. de Colombia, Bogotá, Colombia, ²Univ. de Cartagena, Cartagena, Colombia, ³Pontificia Univ. Javeriana, Bogota, Colombia

Abstract Body:

We aimed to determine single nucleotide variants (SNVs) in the promoter region of *IL-1 β* (c.-511 C>T, c.-31 T>C) and *TNF- α* (c.-308 G>A) gene, and the protein level of these proinflammatory cytokines in saliva of individuals with Obstructive Sleep Apnea (OSA) and Chronic periodontitis (CP). A case and control study that comprised a total 129 subjects from Bogotá, Colombia was conducted. The cases group was subdivided into a group with only CP, only with OSA, and presenting CP and OSA in concomitance. For genotyping, DNA extraction and PCR amplification were performed to carry out Sanger sequencing. Descriptive and inferential statistic was performed using R v3.6.2. The alleles *IL-1 β* c.-511T (OR 2.83 CI: 1.57-5.10), *IL-1 β* c.-31C (OR 2.81 CI: 1.51-5.20) and its homozygous states *IL-1 β* c.-511 TT (OR 7.92 CI: 2.09-29.8), *IL-1 β* c.-31CC (OR 7.2 CI: 1.8-27.9) had significant association with a risk effect. This effect was obtained when comparing individuals with OSA vs Controls and subjects with CP and OSA vs Controls, suggesting that individuals with the mentioned SNVs have two times more risk to develop these inflammatory phenotypes and the risk is even seven times more when having two copies of the allele. In both allelic and genotypes analyzes the risk effect of the SNVs was maintained in patients diagnosed with both diseases. In addition, this genotypic effect could be correlated to the protein level, homozygous individuals *IL-1 β* c.-511TT diagnosed simultaneously with OSA and CP had more IL-1 β protein level in saliva than controls *IL-1 β* c.-511CC (p-value 0.011), as well as when comparing controls *IL-1 β* c.-31TT with OSA and CP *IL-1 β* c.-31CT (p-value 0.001). In conclusion, the alleles *IL-1 β* c.-511T and *IL-1 β* c.-31C increase the risk of develop OSA and/or OSA with CP concomitantly, even more in a homozygous state.

Complex Traits Posters - Wednesday

PB1645. Single-Cell RNA-Seq Reveals the connections of amyloid-beta and ferroptosis tendency

Authors:

Y. Jin; Purdue Univ., West Lafayette, IN

Abstract Body:

Extracellular amyloid-beta plaques and intracellular neurofibrillary tangles have been regarded as the two primary pathological hallmarks of Alzheimer's disease (AD). The mechanism of how amyloid-beta plaques and intracellular neurofibrillary tangles trigger neuron death is still not fully understood. Evidence suggested that ferroptosis may be involved in neuron death in AD. We applied single-cell RNA sequencing method to investigate the relationship between amyloid-beta plaques, and ferroptosis tendency in AD. We used 48 individuals (24 cases and 24 controls) post-mortem brain tissue single cell data from the ROSMAP dataset. We found a positive correlation between the trends of amyloid-beta and ferroptosis. Then single-cell regulatory networks were built via SCENIC based on differentially activated transcription factors. Our findings showed there is a strong correlation between amyloid-beta and ferroptosis tendency.

Complex Traits Posters - Thursday

PB1646. SnRNA-seq probing of the differential vulnerability of motor neurons in ALS

Authors:

P. Alipour; McGill Univ., Montreal, QC, Canada

Abstract Body:

Background: selective vulnerability of some neuronal subpopulations is a common feature of neurodegenerative disorders. While upper and lower MNs degenerate in Amyotrophic Lateral sclerosis (ALS) ocular motor neurons including oculomotor nucleus (CNIII) seem to be largely spared. Diverse CNS cell types may contribute differently to neurodegeneration and are differentially vulnerable. Utilizing a single-nucleus RNA sequencing (snRNA-seq) method we aim to compare transcriptomic profile of differentially vulnerable regions to identify intrinsic properties of oculomotor nucleus that would be absent in medulla in ALS. **Methods:** Post-mortem CNS tissues from ALS cases and normal controls were obtained from the Douglas-Bell Canada Brain Bank, McGill university. Nuclei isolation was performed according to the Nagy et al. Following nuclei isolation, library preparations, sequencing, and Cellranger analyses were performed as per established protocol by 10X Genomics. Following the reception of the expression matrix, nuclei were analyzed in the single-cell analysis R package, Seurat. After quality control and clustering, the differentially gene expression analysis was performed to compare oculomotor nucleus vs medulla in ALS patients. **Results:** clustering of 92K nuclei revealed 23 cell types. Endothelial cell and astrocytes are the only major cell types under-represented in the ALS samples. The majority of DEGs in oculomotor nucleus attribute to astrocytes and oligodendrocytes precursor cells (OPCs), while in the medulla the majority of DEGs attribute to neurons, astrocytes and OPCs. Glial cells show divergent signature in medulla vs oculomotor; while in oculomotor OPCs are involved in pathway indicating activation of the proteasome and phagosome system in response to proteinopathy in medulla OPCs are mostly enriched in antigen processing and presentation including major histocompatibility complex class I (MHC-I); microglia subpopulation signature in medulla represents Disease Associated Microglia (DAM) which are associated with reactive microglia. **Conclusion:** Based on the observation we made thus far; it appears that glial cells are active in the two ALS regions examined here but top DEGs identified in oculomotor do not reveal neurotoxic pathways but are rather associated with the maintenance of cell homeostasis. Given the neuronal dysregulation observed in oculomotor nucleus from ALS cases is minimum, by comparison to what is seen in medulla region, it is presumed that glial cells might be more supportive in this region and better deal with cellular stress of the disease.

Complex Traits Posters - Wednesday

PB1647. SNX8 as a novel predisposing factor by an integrative study for the risk of childhood atopic dermatitis in COCOA

Authors:

E-A. Choi¹, S. Hong¹, S. Hong², S. Lee³, Y-Y. Kim¹, D. Yoon⁴, H-J. Lee⁵; ¹Korea NIH, Cheongju, Korea, Republic of, ²Korea NIH, Seoul, Korea, Republic of, ³Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, ⁴Korea NIH, Cheongju-si, Korea, Republic of, ⁵NIH Korea, Cheongju, Korea, Republic of

Abstract Body:

To discover the genetic determinants for underlying unclear mechanism of atopic dermatitis, we conducted genome-wide association study using Korean chip in Cohort for Childhood Origins of Asthma and Allergic Diseases (COCO A). A novel variant SNX8 rs6974490 showed the association with atopic dermatitis of 1- year children (n=1,236). Children having TT genotypes had a higher risk of atopic dermatitis than those having C allele. Also, to identify epigenetic biomarker, we performed in a cord blood-based DNA methylation analysis using Infinium HumanMethylation850 BeadChip (n=142). CC genotype carriers of SNX8 rs6974490 showed decreased CpG signal in cg09164228 in SNX8 gene when compared to CT or TT carriers. Furthermore, the interaction between childhood SNX8 genotype and their mother's specific IgE to Dermatophagoides farina during pregnancy influenced the risk of AD and the effect of eosinophil levels on SNX8 polymorphism in children. A SNX8 variant was significantly associated with childhood atopic dermatitis, blood eosinophil, and maternal IgE level following house dust mice exposure during pregnancy. An integrative multi-target analysis might reveal new functional implications related to epigenetic control according to genotype in atopic dermatitis.

Complex Traits Posters - Thursday

PB1648. Socioeconomic Status and Risk of Stroke: A Mendelian Randomization Study.

Authors:

E. Myserlis¹, M. K. Georgakis^{2,3,4,5,6}, E. Mayerhofer^{2,3,4,5}, L. Parodi^{2,3,4,5}, J. Rosand^{2,3,4,5}, C. D. Anderson^{2,3,4,7}; ¹Dept. of Neurology, Med. Univ. of South Carolina, Charleston, SC, ²Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ³Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Henry and Allison McCance Ctr. for Brain Hlth., Massachusetts Gen. Hosp., Boston, MA, ⁵Dept. of Neurology, Massachusetts Gen. Hosp., Boston, MA, ⁶Inst. for Stroke and Dementia Res. (ISD), Univ. Hosp., Ludwig-Maximilians-Univ. (LMU), Munich, Germany, ⁷Dept. of Neurology, Brigham and Women's Hosp., Boston, MA

Abstract Body:

Introduction: Low socioeconomic status (SES) has been associated with increased stroke risk in epidemiologic studies. The extent to which this association is explained by uncontrolled modifiable stroke risk factors or whether SES reflects an independent unaddressed burden on this patient population is unclear. We sought to disentangle the association between SES and stroke risk through mendelian randomization (MR) analysis.

Methods: We utilized summary statistics of a multi-trait genome-wide association study of household income (N=505,541) informed by its correlated trait, education attainment ($r_g=0.77$), to select 123 genome-wide, independent ($r^2<0.001$) SNPs to proxy SES. We performed univariable MR evaluating the effect of lower SES on stroke and subtypes (all stroke (N=40,585), ischemic stroke (N=34,217), large artery stroke - LAS (N=4,373), cardioembolic stroke - CES (N=7,193), small vessel stroke - SVS (N=5,386)) in MEGASTROKE (N controls = 406,111). We performed multivariable MR to control for known cardiometabolic risk factors (systolic blood pressure (SBP), type 2 diabetes mellitus (T2DM), low-density lipoprotein (LDL)). Lastly, we assessed to what extent the association between low SES and stroke risk is mediated by modifiable lifestyle risk factors (smoking, alcohol, waist-hip ratio). False discovery rate (FDR) was used to control for multiple hypothesis testing. Association estimates are expressed as OR per unit decrease in the logOR of genetically predicted income category.

Results: Lower socioeconomic status was associated with increased risk of all stroke and ischemic stroke subtypes, according to the univariate IVW analysis (all stroke: OR=1.47, 95% CI: 1.30-1.66, p-FDR=2.73x10⁻⁸; ischemic stroke: OR=1.46, 95% CI: 1.27-1.67, p-FDR=5.70x10⁻⁷; LAS: OR=1.63, 95% CI: 1.10-2.43, p-FDR=0.02; CES: OR=1.35, 95% CI: 1.04-1.73, p-FDR=0.022; SVS: OR=1.75, 95% CI: 1.28-2.38, p-FDR=0.00103). Effect estimates were largely directionally consistent across univariate MR methods. After controlling for SBP, T2DM, or LDL, the association between low SES and all stroke and ischemic stroke remained significant. While 33% (95% CI: 10-36%) of the association between SES and cardioembolic stroke risk was mediated by alcohol consumption, no other lifestyle risk factors were found to mediate associations between low SES and stroke or its subtypes.

Conclusion: We found that low SES is causally associated with increased stroke risk, independent of known cardiometabolic risk factors. Further studies are needed to address social and other determinants of health as an important way of decreasing stroke burden across populations.

Complex Traits Posters - Wednesday

PB1649*. Somatic mutations in chronic lung disease are associated with reduced lung function

Authors:

J. Yun¹, W. Khan², B. Hobbs³, P. Castaldi³, C. Hersh¹, P. Miller⁴, F. Sciruba¹, L. Barwick⁵, A. Limper⁶, K. Flaherty⁷, G. Criner⁸, K. K. Brown⁹, R. Wise¹⁰, F. Martinez¹¹, E. K. Silverman¹², M. H. Cho³, A. Bick¹³; ¹Brigham and Women's Hosp., Boston, MA, ²Vanderbilt Univ., Nashville, TN, ³Brigham and Women's Hosp., Boston, MA, ⁴Massachusetts Gen. Hosp., Boston, MA, ⁵The Emmes Company, Rockville, MD, ⁶Mayo Clinic, Rochester, MN, ⁷Univ. of Michigan, Ann Arbor, MI, ⁸Temple Univ., Philadelphia, PA, ⁹Natl. Jewish Hlth., Denver, CO, ¹⁰Johns Hopkins Univ., Baltimore, MD, ¹¹Weill Cornell Med., New York, NY, ¹²Brigham & Women's Hosp, Boston, MA, ¹³Vanderbilt, Nashville, TN

Abstract Body:

The lung carries one of the highest somatic mutational loads among tissues in healthy individuals, however whether somatic mutations in the lung are enriched among patients with chronic lung diseases is unknown. We hypothesized that somatic mutations would be increased in smoking-related chronic lung diseases. To identify somatic mutations, we identified somatic single nucleotide variants (SNV) in lung tissue from RNA sequencing paired with blood-derived whole genomes from the Lung Tissue Research Consortium (n=1,364). SNVs were called with GATK MuTect and filtered to a set of 285,000 high confidence mutations. To test for associations with phenotypes, we selected subjects with confirmed histopathology, smoking history, and lung function data which included 29 normal (based on histology and lung function), 352 with chronic obstructive pulmonary disease (COPD), 164 with idiopathic pulmonary fibrosis (IPF). Surprisingly, mutational burden (the number of SNVs) was not associated with age or smoking history, overall or within subgroups ($P > 0.2$). In addition, we did not find significant enrichment for somatic mutations in known lung cancer driver genes. However, we did find mutation burden significantly associated with reduced lung function (forced expiratory volume in 1 second (FEV1) % predicted) in COPD ($R = -0.16$, $p < 0.01$) and IPF ($R = -0.29$, $p < 0.001$), but not in controls ($R = -0.069$, $p = 0.7$). In addition, mutation burden was associated with an increased proportion of airway versus alveolar epithelial cells ($p < 0.001$ across all groups). There was no association of mutational burden with immune cell or stromal cell abundance. In multivariable regression analysis adjusted for age, sex, race and smoking history, both airway/alveolar epithelial ratio and lung function (FEV1% predicted) remained statistically significantly associated with mutational burden (Beta 16.7 and -7.27 respectively, $p < 10^{-4}$). Our study suggests that somatic mutations are associated with reduced lung function and specific cell types in chronic lung disease.

Complex Traits Posters - Thursday

PB1650. Spectrum of causal genetic variants in inherited cardiomyopathies and arrhythmias in a primarily French-Canadian Population: A 14-year retrospective study from a specialized Cardiovascular Genetic Center.

Authors:

S. Grondin^{1,2}, I. Soltani^{1,2}, A. Alaoui^{1,2}, A. Messina^{1,2}, L. Gaumond^{1,2}, A. Jeuken^{1,2}, C. Barahona-Dussault¹, G. Sylvain-Drolet¹, B. Neveu¹, L. Robb¹, J. Gagnon¹, V-A. Codina-Fauteux¹, D. Victoria Moron¹, S. Therrien-Laperrière¹, V. Hay¹, L. Rivard¹, G. Giraldeau¹, P. Lavoie-L'Allier¹, P. Garceau¹, M. Tremblay-Gravel¹, J. Cadrin-Tourigny^{1,2}, M. Talajic^{1,2}, J. Amyot^{1,2}, R. Tadros^{1,2}; ¹Montreal Heart Inst., Montreal, QC, Canada, ²Université de Montréal, Montreal, QC, Canada

Abstract Body:

Background: The genetic spectrum of inherited cardiomyopathies and arrhythmias in the French-Canadian population is unknown. Interpretation of genetic variants can be challenging in understudied populations where founder effects are common. **Objective:** To describe the genetic pool in probands with inherited cardiovascular disorders, where the majority are of French-Canadian ancestry. **Methods:** We included probands with suspected inherited cardiomyopathies and arrhythmias seen between 2006 and 2020 at the Montreal Heart Institute Cardiovascular Genetic Center, with clinical information and genetic results available. Medical charts and genetic results were retrospectively reviewed. Diagnoses and strengths of diagnoses (definite, probable, possible) were systematically adjudicated using predefined criteria. **Results:** A total of 1605 probands (mean age 48 ± 17 years) were included. The majority were males (64%) with at least one parent of French-Canadian ancestry (72%). Genetic testing was performed using diagnosis-specific gene panels (median 17 genes, range [1-133 genes]). A total of 609 distinct variants were identified, of which 213 (35%) were classified as likely pathogenic or pathogenic (LP/P). Overall, 369/1335 (28%) of probands with definite clinical diagnoses carried a LP/P variant, as opposed to 50/270 (19%) for those with non-definite diagnoses ($P=0.002$). The rate of LP/P variants varied across diseases: 32% for hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy (DCM), 48% for arrhythmogenic cardiomyopathy and 12% for Brugada syndrome. In DCM, the proportion of probands with LP/P variants was higher for those with a positive familial history (49/102; 48%) compared to sporadic cases (41/178; 23%; $P=3 \times 10^{-5}$). Reflecting the founder effect in the French-Canadian population, 57 unique LP/P variants were identified in at least 3 unrelated probands. In HCM, 3 recurrent LP/P variants together account for 16% of French-Canadian HCM probands overall, and 35% of gene-positive HCM. These 3 variants are: *TNNT2* (NM_001001432):p.Trp287*, *MYBPC3* (NM_000256):c.551dup and *MYH7* (NM_000257):p.Arg204Leu seen in 39, 23, and 10 apparently unrelated HCM probands, respectively. **Conclusion:** Genetic testing identified a disease-causing variant in 26% of probands with a definite diagnosis in a clinical context, highlighting the high yield of rare variant genetic testing in complex conditions. Rare disease-causing variants are found in multiple unrelated probands reflecting founder effects in the French Canadian population. In HCM, 3 variants account for a large portion of all gene-positive cases.

Complex Traits Posters - Wednesday

PB1651. Stratification of a PMS population based on their response to Human Growth Hormone and Insulin-like Growth Factor 1

Authors:

B. Moffitt¹, S. Sarasua¹, D. Ivankovic², L. Ward¹, K. Valentine¹, W. Bennett³, R. Rogers⁴, K. Phelan⁵, L. Boccuto¹; ¹Clemson Univ., Clemson, SC, ²Anderson Univ., CLEMSON, SC, ³Indiana Univ., Indianapolis, IN, ⁴Greenwood Gen Ctr, Greenville, SC, ⁵Florida Cancer Specialists, Fort Myers, FL

Abstract Body:

Background: Phelan-McDermid syndrome (PMS) is caused by pathogenic variants in the *SHANK3* gene or deletions of the distal region of chromosome 22. PMS is characterized by intellectual disability, autistic features, developmental delays, and neonatal hypotonia. Previous studies have found that insulin-like growth factor 1 (IGF-1) and human growth hormone (hGH) can reverse a range of deficits in PMS. This study aims to characterize the metabolic profile of a PMS cohort, outline potential biomarkers associated with abnormal responses to IGF-1 and/or hGH, and identify individuals with PMS who would benefit from IGF-1 or hGH treatments or who would be at higher risk for reduced or adverse responses to these compounds. Methods: Metabolic profiling of lymphoblastoid cell lines from 48 individuals with PMS and 50 controls was performed using Biolog Phenotype Mammalian Microarray plates (PM-Ms) encompassing 776 metabolites. Subpopulations were determined by taking the top and bottom 25% of responders to hGH and IGF-1. Correlation of the metabolic data versus the baseline glucose metabolic activity was conducted to determine population subgroups based on only the effect of the hormones on the metabolism of glucose. Results: A distinct metabolic profile for individuals with PMS showed a reduced ability to metabolize major energy sources and a higher metabolism of alternative energy sources. This is thought to be a compensatory mechanism. The analysis of the metabolic response to hGH or IGF-1 highlighted a major overlap between both high and low responders, validating the model and suggesting that the two growth factors share a large number of target pathways. The data allowed for stratification of the PMS population with remarkable differences between the high- and low-responder profiles. When we investigated the effect of hGH and IGF- on the metabolism of glucose, the correlation between the high responder subgroups showed less similarity, whereas the low responders were still relatively similar. Conclusions: The classification of individuals with PMS into subgroups based upon response to a hormone can be used to investigate different pathogenic mechanisms, identify molecular biomarkers, explore the response to candidate drugs in an in vitro model, and eventually select better candidates for clinical trials to optimize the response to treatment and minimize the side effects.

Complex Traits Posters - Thursday

PB1652. Stratified LD score regression with 3D genomic features in metabolically relevant cell types implicates a role for pancreatic alpha cells in the pathogenesis of childhood obesity.

Authors:

K. Trang¹, M. C. Pahl¹, J. A. Pippin¹, P. Seale², P. M. Titchenell², W. Yang³, K. Kaestner², A. D. Wells¹, S. F. A. Grant¹; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Inst. for Diabetes, Obesity and Metabolism, Univ. of Pennsylvania, Philadelphia, PA, ³Inst. for Translational Med. and Therapeutics (ITMAT), Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Obesity confers a considerable burden on society and quality of life. Of note, the prevalence of it in childhood is increasing worldwide, along with comorbidities that include type 2 diabetes and cardiovascular disease in later life. We previously led the GWAS efforts for childhood obesity on behalf of the Early Growth Genetics (EGG) Consortium, revealing loci associated with the trait. However, it is unclear in which cellular context these loci confer their effects; furthermore, it is clear that additional efforts are required to characterize the causal variant(s) and corresponding effector genes at these loci. Leveraging our high-resolution Hi-C, ATAC-seq and RNA-seq datasets derived from sorted human pancreatic beta, alpha, acinar cells, skeletal muscle myotubes differentiated from primary myoblasts, adipocytes differentiated in vitro from primary preadipocytes (pre- and post-differentiation) and primary hepatocyte cells, we carried out stratified LD regression score in metabolically relevant cell types in order to further our understanding of disease etiology. This approach calculates the proportion of genome-wide SNP heritability attributable to our derived cell type-specific features in order to estimate each cell-type polygenic contributions to the disease heritability. Interestingly, we observed a significant 8.9-fold enrichment in pancreatic alpha cells ($P=0.027$), along with trends in additional key cell types. Given this observation of significant enrichment of childhood obesity loci in the specific setting of pancreatic alpha cells with our genomic features of interest, we are now motivated to carry out a variant to gene mapping strategy in this cellular setting going forward.

Complex Traits Posters - Wednesday

PB1653. Study of gene-environment interactions suggests a personalised medicine approach for ALS

Authors:

J. Cooper-Knock¹, C. Harvey¹, S. Zhang², K. Kenna³, T. Julian⁴, S. Boddy¹, M. Islam¹, C. McDermott¹; ¹Univ. of Sheffield, Sheffield, United Kingdom, ²Stanford Univ., Palo Alto, CA, ³UMC Utrecht, Utrecht, Netherlands, ⁴Univ. of Manchester, Manchester, United Kingdom

Abstract Body:

Amyotrophic lateral sclerosis (ALS) is a common and incurable neurodegenerative disease. Although 10% of ALS is monogenic, the majority of disease is caused by a complex interaction of genetic and environmental risk factors. We have previously used Mendelian randomization (MR) to show that serum isoleucine is positively linked to risk of ALS and that this effect is mediated via depletion of vitamin B12 (Boddy et al., 2022). The principle of MR is that environmental exposures can be measured by genetic predisposition. We have taken advantage of this fact to perform a data-driven gene-environment screen. Serum isoleucine was quantified as a discrete continuous variable based on the number of isoleucine-associated alleles per individual. We then applied a version of rare variant burden testing (Lin et al., 2016) to test for a significant interaction between serum isoleucine and missense genetic variants genome-wide using whole genome sequencing data from 3,727 ALS patients and 1,687 controls (number of variants = 31). There was no evidence of statistical inflation ($\lambda=1.005$) and one gene was significant after Bonferroni multiple testing: PDIA3 (iSKAT, $p=9.01E-6$). My analysis suggests that missense mutations within PDIA3 protect against isoleucine-induced ALS. Elevated levels of PDIA3 have been found in spinal cord tissue from ALS patients and recently it has been shown that PDIA3 is involved in the triggering of apoptosis in response to misfolded proteins suggesting that PDIA3 function may exacerbate neurotoxicity. Depletion of vitamin B12 induces ER stress that would in-turn be expected to increase PDIA3 expression. We propose that missense mutations within PDIA3 avert isoleucine-induced neurotoxicity resulting from this pathway. The success of this method is the platform for a larger systematic study of gene-environment interactions using the total set of ALS-associated genetic and environmental risk factors.

Complex Traits Posters - Wednesday

PB1655. Telomere length associated rare variant candidate gene study in idiopathic pulmonary fibrosis

Authors:

J. Radder, J. F. McDyer, J. K. Alder, A. Morris; Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA

Abstract Body:

BACKGROUND: Idiopathic pulmonary fibrosis (IPF) is a chronic fibrosing lung disease with significant morbidity and mortality. Telomere length is increasingly recognized as a significant mediator of IPF pathogenesis. At least one third of individuals with familial IPF have telomere-related mutations and individuals with short telomeres even in the absence of known mutations have been demonstrated to have shorter lung transplant-free survival and more rapidly progressing disease. We sought to test for association between rare variants in genes previously reported to be associated with telomere length and IPF using a case-control approach.

METHODS: Whole genome sequencing was performed on 316 individuals with IPF (cases) and 429 individuals with no evidence of IPF (controls) using Illumina technology. Alignment to GrCh38 and variant calling was performed using a standard BWA/GATK4 pathway. Telomere length was calculated using TelSeq. Genes previously identified as significantly associated with telomere length were chosen as candidate genes and single nucleotide variants that were rare in a European reference population from gnomAD (MAF < 0.05). Association testing using a case-control approach was performed for gene-based tests using SKAT-O and single variant tests using the logistic score test with age, sex, and top eigenvectors from population stratification as covariates.

RESULTS: There was a significant difference in telomere length between cases and controls (cases 4.1, controls 4.5, $P=3.25 \times 10^{-10}$). Using a groupwise test across candidate genes we identified three significantly associated genes including *ABCB9* ($P=4.68 \times 10^{-6}$), *OGFOD2* (8.68×10^{-6}), and *SH2B3* (1.58×10^{-5}). We also tested for association between individual rare single nucleotide variants and the trait with top associations in the genes *ZMYM4*, *VIPR2*, *SH2B3*, and *ABCB9*.

CONCLUSIONS: Using gene-based and single rare variant-based association tests, we identified several candidate genes or rare variants in candidate genes that were associated with IPF. We chose to perform this analysis using all rare variants rather than filtering only for functional variants. Interestingly, while well-known telomere-associated genes like *TERT* reached suggestive study-wide significance using this approach, those reaching study-wide significance were less commonly recognized genes suggesting a need for further investigation of whether these genes play a non-causative but mediating role in the development of IPF.

Complex Traits Posters - Thursday

PB1656. Telomere length, and Parkinson's disease risk, age at onset, and presentation: Mendelian randomization analyses suggest no causal relationships

Authors:

A. Iyengar¹, E. Misicka¹, F. B. S. Briggs^{1,2}; ¹Case Western Reserve Univ., Cleveland, OH, ²Cleveland Inst. for Computational Biology, Cleveland, OH

Abstract Body:

Background: Parkinson's disease (PD) is the second-most common neurodegenerative disorder, affecting approximately 1 million Americans. PD is typically characterized by motor issues such as tremors and gait difficulty. Because PD is most prevalent in aging populations, shortened telomeres may be an underlying risk factor for development. This study explores a potential causal relationship between genetically-driven telomere length (TL) and various aspects of PD to determine whether TL affects PD risk, onset, and/or presentation.

Objective: To investigate the effects of TL on multiple PD phenotypes using Mendelian randomization (MR), a robust tool for causal analysis.

Methods: MR is a form of instrumental variable analysis utilizing genome-wide association (GWA) summary statistics to infer causality between phenotypes. Individual variants associated with an exposure or phenotype of interest are identified in both exposure and outcome GWA datasets, and the effect estimates for each are combined into a ratio that can then be meta-analyzed using inverse variance-weighted (IVW) meta-analysis to investigate the overall effect of the exposure on the outcome. IVW and other two-sample (2S) MR approaches were performed to investigate the potential causal effects of TL on PD risk, age at onset (AAO), tremor-dominant (TD) or postural instability/gait difficulty (PIGD) motor phenotype (MP), and a TD-PIGD score ratio (MPR). A genetic instrument for TL was created from 197 genetic variants associated with TL at genome-wide significance in a sample of 472,174 individuals from the UK Biobank. Summary statistics for PD phenotypes were extracted from the IPDGC data portal, with sample sizes of 3,212 for MP and MPR, 28,568 individuals for PD AAO, and 37,688 PD cases and 1.4 million controls for PD risk. Genetic variants in the TL instrument were clumped at an LD threshold of $r^2 < 0.05$ using the 1000 Genomes EUR Reference Panel.

Results: Using a 2SMR IVW approach, TL was not found to be associated with PD risk ($\beta_{IVW}=0.15$, $p=0.09$), AAO ($\beta_{IVW}=0.75$, $p=0.16$), MP ($\beta_{IVW}=0.36$, $p=0.17$), or MPR ($\beta_{IVW}=-0.11$, $p=0.26$). Null associations were also observed for other 2SMR approaches (i.e., MR-Egger, MR-PRESSO).

Conclusions: There is no evidence that genetically-driven TL causally influences risk, age at onset, or disease subtype of PD.

Complex Traits Posters - Wednesday

PB1657. THE ASSOCIATION OF *FSHR* POLYMORPHISMS WITH PCOS IN PUNJABI POPULATION

Authors:

A. Kaur; Guru Nanak Dev Univ., AMRITSAR, India

Abstract Body:

Background: Polycystic ovary syndrome (PCOS) is one of the most common endocrine-metabolic disorders affecting a significant number of women at their child bearing age around the world and characterized by hyperandrogenism, menstrual irregularities and polycystic ovarian morphology. Recent Genome-wide association studies (GWAS) have identified the 11 susceptibility loci for PCOS including *THADA*, *LHCGR*, *FSHR* etc. There is scarcity of data of North Indian population that signifies the underlying genetic mechanisms which lead to development of PCOS. In present study we investigated the association of *FSHR* genetic variants with PCOS in Punjabi population. **Methodology:** A case-control study, involving 200 women with PCOS and 200 age-matched healthy controls, was conducted. Relevant data of the participants was collected after taking informed consent. *FSHR* polymorphisms (rs6166, rs6165 and rs1394205) were genotyped using polymerase chain reaction-restriction fragment length polymorphism. Biochemical analysis was done in PCOS cases and controls. Statistical analysis was performed using SPSS (version 21, IBM SPSS, NY, USA). **Results:** Anthropometric measurements like BMI and WHR revealed a significant difference between PCOS cases and controls ($p < 0.000$ and $p = 0.0001$, respectively). Levels of cholesterol and triglycerides were found to be significantly higher and levels of HDL was lower in PCOS women than controls. In rs6166 polymorphism, it was observed that heterozygote genotypic frequency is more than both homozygote wild and mutant genotypic frequencies (Cases: AG-49%, AA-32% and GG-19%; Controls: AG-53%, AA-28.5% and GG-18.5%, respectively). Genotypic and allelic frequency is not significantly different with $p = 0.69$ and $p = 0.66$, respectively. Significant difference between genotypes of cases and controls was found for rs6165 polymorphism with $p = 0.03$. Heterozygote genotype (AG) of rs6165 was found to conferred 1.8 times risk of PCOS with $p = 0.01$. Dominant and co-dominant model of rs6165 was also revealed that both models provide significant risk towards the progression of PCOS (1.5 and 1.7 folds, respectively). Genetic distribution of rs1394205 did not show any significant difference for genotypic and allelic frequencies. **Conclusion:** Our study reveals the significant association of *FSHR* rs6165 polymorphism with PCOS in dominant and co-dominant model whereas no association was found for other two rs6166 and rs1394205 polymorphisms. It was concluded that polymorphism rs6165 could contribute to the impaired folliculogenesis in PCOS.

Complex Traits Posters - Thursday

PB1658. The bidirectional causal effects of brain morphology across the life course and risk of Alzheimer's disease: A cross-cohort comparison and Mendelian randomization meta-analysis.

Authors:

R. Korologou-Linden¹, B. Xu², E. Coulthard¹, E. Walton³, A. Wearn¹, G. Hemani⁴, T. White¹, C. Cecil⁵, T. Sharp¹, H. Tiemeier⁶, The IMAGEN consortium, L. Howe⁴, Y. Ben-Shlomo¹, N. Davies⁴, E. L. Anderson¹; ¹Univ. of Bristol, Bristol, United Kingdom, ²Erasmus MC Univ. Med. Ctr., Dept. of Child and Adolescent Psychiatry and Psychology Rotterdam, UK, Rotterdam, Netherlands, ³Univ. of Bath, Bath, United Kingdom, ⁴Univ. of Bristol, Bristol, United Kingdom, ⁵Erasmus MC Univ. Med. center, Rotterdam, Netherlands, ⁶Harvard Univ., Boston, MA

Abstract Body:

Neuropathological changes due to Alzheimer's disease (AD) can occur decades before clinical symptoms. We used bidirectional two-sample Mendelian randomization to estimate the effects of genetic liability to AD on global and regional cortical thickness, total intracranial volume, volume of subcortical structures and total white matter in 37,680 participants aged eight to 81 years. We also examined the effects of global and regional cortical thickness and subcortical volumes on AD risk in up to 37,741 participants. AD risk alleles have an age-dependent effect on a range of cortical and subcortical brain structures. Some of the identified structures are not typically implicated in AD, such as those in the striatum such as the thalamus, with consistent effects from childhood to late adulthood. We found little evidence to suggest brain morphology alters AD risk. Genetic liability to AD is likely to affect AD risk primarily through mechanisms affecting indicators of brain morphology in later life (e.g. potential neurodegeneration), rather than structural brain reserve.

Complex Traits Posters - Wednesday

PB1659. The causal effect of blood lipid metabolism related traits with depression phenotypes: the evidence from mendelian randomization study

Authors:

S. Tao¹, X. Ye², P. Huang², W. Duan³, S. Yang¹; ¹Dept. of Biostatistics, Ctr. for Global Hlth., Sch. of Publ. Hlth., Nanjing Med. Univ., Nanjing, China, Nanjing, China, ²Dept. of Epidemiology, Ctr. for Global Hlth., Sch. of Publ. Hlth., Nanjing Med. Univ., Nanjing, China, Nanjing, China, ³Dept. of Bioinformatics, Sch. of BioMed. Engineering and Informatics, Nanjing Med. Univ., Nanjing, China, Nanjing, China

Abstract Body:

Depression is one of the most universal psychiatric disorders with severe symptoms, such as loss of interest or pleasure, recurrent thoughts of death, suicide. Emerging genome wide association studies (GWAS) have identified significant loci of depression, but its etiology and mechanism are still unknown. Previous cohort studies and cross-sectional studies have demonstrated a positive association between blood lipid metabolism related traits, including triglyceride (TG), non-alcoholic fatty liver disease (NAFLD), and alanine transaminase (ALT) and depression phenotypes, including depressive symptom (DS) and major depressive disorder (MDD), but the pathogenesis of the association has not been elucidated. Here, we used two-sample mendelian randomization (MR) strategy to identify the potential causal effect of blood lipid metabolism related traits, including blood lipids, NAFLD, and transaminase on depression phenotypes in European population. Therefore, we collected four kinds of summary statistics: (i) depressive symptom (DS) and major depressive disorder (MDD); (ii) four of blood lipid, including total cholesterol (TC), TG, low-density lipoprotein cholesterol (LDL), and high-density lipoprotein cholesterol (HDL); (iii) NAFLD activity score; (iv) ALT and aspartate transaminase (AST); First, using the same clumping criteria ($P < 5E-8$ and $r^2 > 0.001$ within 10 Mb), we identified, on average, 30 (ranging from 11 to 54) instrumental variables (IVs) for the seven traits. Then, we estimated the association of non-alcoholic fatty liver, alanine transaminase, and blood lipids with DS or MDD using the inverse-variance-weighted (IVW) method. Furthermore, to validate the reliability and robustness of results from MR analyses, we performed three types of sensitivity analyses including heterogeneity test, pleiotropic test, and leave-one-out (LOO) test. We performed IVW with multiplicative random effects when instrument SNPs showed heterogeneity ($P < 0.05$). From the results of MR analysis supported by MR-IVW, TC, NAFLD activity score, and ALT were causally associated with DS, with beta of 0.0067 (95% CI: 0.0007-0.0126, $P = 0.028$), 0.0014 (95% CI: 0.0002-0.0027, $P = 0.026$), and 0.0601 (95% CI: 0.0018-0.1184, $P = 0.043$), respectively. TG was causally associated with MDD (OR = 1.111, 95% CI: 1.033-1.188, $P = 7.8E-3$). Our research provided new evidence to support the causal effect of blood lipid metabolism related traits on depression phenotypes, which indicates screening populations with abnormal indicators related to blood lipid metabolism would be a feasible strategy to detect depression and then reduce the psychiatric disorder burden.

Complex Traits Posters - Thursday

PB1660. The effect of population structure and ancestry-related assortative mating on the genetic architecture of autism spectrum disorder.

Authors:

J. Zhang¹, J. D. Weissenkampen¹, R. L. Kember¹, E. S. Brodtkin¹, L. Almasy², M. Bucan¹, R. Sebro³; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Children's Hosp. of Philadelphia, Philadelphia, PA, ³Mayo Clinic Florida, Jacksonville Beach, FL

Abstract Body:

Assortative mating in several neuropsychiatric disorders, including autism spectrum disorder (ASD) has been noted. It is unknown whether the phenotypic similarity between parents of children with ASD could be because spouse choice was based on phenotypes influenced by different ASD-related genes leading to correlations between parents ASD polygenic scores (PGS), or if spouse choice was based on ancestry and that there are ASD-related genes associated with underlying ancestry. To address these questions, we analyzed two family-based ASD collections: the Simons Foundation Powering Autism Research for Knowledge (SPARK) (1,505 families) and the Simons Simplex Collection (SSC) (2,285 families). We found phenotypic assortative mating for age ($r_{\text{SPARK}}=0.72$, $P<0.001$; $r_{\text{SSC}}=0.73$, $P<0.001$), education level ($r_{\text{SPARK}}=0.45$, $P<0.001$; $r_{\text{SSC}}=0.48$, $P<0.001$), and for quantitative autistic traits (in SSC) assessed by Social Responsiveness Scale (SRS) ($r_{\text{SSC_SRS_total}}=0.34$, $P<0.001$) and the Broad Autism Phenotype Questionnaire (BAPQ) ($r_{\text{SSC_BAPQ_total}}=0.13$, $P<0.001$) in families of European ancestry. We found ancestry-related assortative mating between spouses of European ancestry (measured by the spousal correlations of principal components (PC) from the principal component analysis with 1000 Genomes participants of European ancestry) in SPARK ($r_{\text{PC1}}=0.38$, $P<0.001$; $r_{\text{PC2}}=0.46$, $P<0.001$; $r_{\text{PC3}}=0.38$, $P<0.001$) and SSC ($r_{\text{PC1}}=0.45$, $P<0.001$; $r_{\text{PC2}}=0.54$, $P<0.001$; $r_{\text{PC3}}=0.38$, $P<0.001$). This ancestry-related assortative mating led to greater induced linkage disequilibrium (LD) measured by r^2 between highly ancestry-informative single nucleotide polymorphisms (SNPs) on different chromosomes compared to the LD between less ancestry-informative SNPs on different chromosomes (SPARK: $P_{\text{fathers}}=0.008$, $P_{\text{mother}}=0.008$, $P_{\text{unaffected_siblings}}=0.021$; SSC: $P_{\text{fathers}}=0.027$, $P_{\text{mother}}=0.055$, $P_{\text{unaffected_siblings}}=0.097$). There was no spousal correlation of ASD PGS after controlling for population substructure in SPARK ($r=0.04$, $P=0.139$). The presence of phenotypic assortative mating did not lead to a correlation between spouses' ASD PGS. This study shows that the phenotypic assortative mating for autistic traits is likely due to ancestry-based assortative mating in SPARK and SSC. The maximum induced LD between highly ancestry-informative SNPs on different chromosomes can be greater than 0.1 (measured by r^2), and although values of this magnitude were rare, is a phenomenon that is worth noting.

Complex Traits Posters - Wednesday

PB1661*. The first GWAS on intrahepatic cholestasis of pregnancy unveils novel bile acid metabolism related associations.

Authors:

J. Tyrmi^{1,2}, A. Havulinna^{3,4}, T. Laisk⁵, ICP Consortium, FinnGen, DBDS Genomic Consortium, J. Kettunen^{6,7}, M. Daly^{8,9,10}, D. Westergaard^{11,12,13}, S. S. Venkatesh^{14,15}, T. Tukiainen¹⁶, M. Nyegaard¹⁷, C. M. Lindgren^{18,15}, H. Laivuori^{19,20}; ¹Univ. of Oulu, Oulu, Finland, ²Tampere Univ., Tampere, Finland, ³Inst. for Molecular Med. Finland - FIMM, Helsinki, Finland, ⁴Univ. of Helsinki, Helsinki, Finland, ⁵Inst. of Genomics, Univ. of Tartu, Estonia, Estonia, ⁶Univ. of Oulu, Oulu, Oulu, Finland, ⁷Natl. Inst. for Hlth. and Welfare, Helsinki, Finland, Helsinki, Finland, ⁸Inst. for Molecular Med. Finland, Univ. of Helsinki, Helsinki, Finland, Helsinki, Finland, ⁹Broad Inst., Cambridge, MA, ¹⁰Analytical and Translational Genetics Unit, Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, ¹¹Dept. of Gynecology 4232, Rigshosp. et Univ. Hosp., Univ. of Copenhagen, Copenhagen, Denmark, ¹²Novo Nordisk Fndn. Ctr. for Protein Res., Univ. of Copenhagen, Copenhagen, Denmark, ¹³Methods and Analysis, Statistics Denmark, Copenhagen, Denmark, ¹⁴Big Data Inst., Li Ka Shing Ctr. for Hlth. Information and Discovery, Univ. of Oxford, Oxford, United Kingdom, ¹⁵Wellcome Ctr. for Human Genetics, Nuffield Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, ¹⁶Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, ¹⁷Aalborg Univ., Dept. of Hlth. Sci. and Technology, Aarhus, Denmark, ¹⁸Big Data Inst., Oxford Univ., Oxford, Oxfordshire, United Kingdom, ¹⁹Tampere Univ. Hosp. and Tampere Univ., Tampere, Finland, ²⁰Helsinki Univ. Hosp. and Univ. of Helsinki, Helsinki, Finland

Abstract Body:

Background: Intrahepatic cholestasis of pregnancy (ICP) is a common, complex disorder, usually diagnosed in the second half of pregnancy. It is characterized by pruritus and elevated serum bile acids. The incidence of ICP varies with geographic location and ancestry (between 0.2% and 2%). It is linked to increased risk of fetal complications including preterm labor, fetal asphyxia, meconium-stained amniotic fluid, stillbirth, and is associated with high maternal rates of gestational diabetes and preeclampsia. Genes coding for bile acid transporters have been implicated in ICP, but no genome-wide association study (GWAS) has been published so far. **Methods:** We performed multi-study ICP GWAS meta-analysis in >2,132 participants having a history of ICP and 415,682 female controls from the FinnGen, DeCODE, Danish blood donor study (DBDS) and Estonian genome center of the University of Tartu (EGCUT) studies. We identified putative candidate genes and assessed their potential causality within each associated locus by examining each variants eQTL, sQTL and pQTL effect in various databases. We also examined the relationship of ICP and other disorders via PheWAS, genetic correlation and comorbidity analysis. **Results:** GWAS identified twenty associated loci ($p < 5 \times 10^{-8}$). Many of the identified genes are expressed particularly in the liver and are known to be involved in bile acid synthesis (e.g. *CYP7A1*), transportation (several ABC protein coding genes), and related processing of bile acids. 14/20 loci were associated ($P < 1 \times 10^{-6}$) with liver or cardiovascular biomarkers, including albumin, LDL, cholesterol, and gamma glutamyltransferase. Comorbidity analysis suggest association with various pregnancy complications and autoimmune diseases. PheWAS for lead variants show ICP-specific findings, associations with use of statin medication, type 2 diabetes, and cholelithiasis. The latter association was further supported by genetic correlation analysis. **Discussion:** Detailed understanding of the genetic predisposition to ICP has remained elusive, but in this first ICP GWAS we discover 20 associated loci, and outline the genetic architecture predisposing individuals to ICP. We also provide more detailed support for previous findings of impaired bile secretion in ICP.

Complex Traits Posters - Thursday

PB1662. The genetic architecture of the corpus callosum and its subregions

Authors:

M. Campbell; Univ. of Cape Town, Cape Town, South Africa

Abstract Body:

Background: Regional surface area and thickness of the cerebral cortex and volume of subcortical structures are highly heritable brain morphological features with complex, polygenic genetic architectures. However, the genetic architecture of the corpus callosum (CC) and its subregions remains largely unclear. We aim to determine the heritability and investigate the genetic architecture of CC volume and each subregion and the extent to which this overlaps with that of psychiatric disorders.

Methods: MOSTest method was applied to genetic and T1-weighted MRI data of 40,894 individuals from the UK-biobank to conduct a multivariate GWAS aiming to boost genetic discovery and to assess the pleiotropic effects of common variants across the five subregions of the CC (posterior, mid posterior, central, mid anterior and anterior) obtained by running the automatic subcortical segmentation algorithm in FreeSurfer 5.3. MOSTest was run on each subregion, co-varying for total intracranial volume, sex, age squared, scanner and the first 20 genetic principal components. Gene-set enrichment analyses were performed using MAGMA. Linkage disequilibrium score regression (LDSC) was used to determine the SNP-based heritability of the CC and its subregions and the genetic correlation between each subregion and a variety of psychiatric disorders.

Results: Following MOSTest, 70 independent loci were identified across the 5 subregions of the CC ($p < 5 \times 10^{-8}$). Using LDSC, we found evidence to suggest that total CC volume is heritable ($h^2_{\text{SNP}} = 0.38$, $SE = 0.03$). Heritability estimates for volume of the subregions range from 0.22 ($SE = 0.02$) in the mid-posterior subregion to 0.37 ($SE = 0.03$) in the posterior subregion. Significant variants show enrichment in pathways related to regulation of the nervous system and cell development, neurogenesis, and regulation of neuron differentiation. Gene-set analysis revealed 156 significant genes ($p < 2.6 \times 10^{-6}$). Many of the significant SNPs have been previously associated with white matter hyperintensity volume as well as a range of psychiatric disorders. Bipolar disorder (BD) showed significant negative genetic correlation with volumes of the posterior subregion ($r_g = -0.11$; $p = 1.6 \times 10^{-3}$) and total CC volume ($r_g = -0.10$; $p = 4 \times 10^{-3}$).

Conclusions: We provide the first evidence to suggest that total CC volume and volumes of the CC subregions are heritable. Gene set enrichment analyses identified pathways related to neuron development and neurogenesis, suggesting that CC alteration may have an independent developmental origin. We also provide evidence for shared genetic etiology between BD risk and altered CC volumes.

Complex Traits Posters - Wednesday

PB1663. The genetic drivers of young- and early-onset Parkinson's Disease in India

Authors:

S. Andrews¹, P. L. Kukkle^{2,3}, R. Menon⁴, T. S. Geetha⁴, V. Goyal^{5,6,7}, R. M. Kandadai⁸, H. Kumar⁹, R. Borgohain⁸, A. Mukherjee¹⁰, P. M. Wadia¹¹, R. Yadav¹², S. Desai¹³, N. Kumar¹⁴, D. Joshi¹⁵, S. Murugan⁴, A. Biswas¹⁰, P. K. Pal¹², M. Oliver⁴, S. Nair⁴, A. Kayalvizhi⁴, U. Mahadevia⁴, S. Sandaram⁴, M. Deshmukh⁴, A. Bassi⁴, C. Sandeep⁴, N. Mandloi⁴, U. Muthane¹⁶, S. K. Das¹⁰, T. Sandmann¹, R. Gupta⁴, V. L. Ramprasad⁴, Parkinson Research Alliance of India (PRAI); ¹Denali Therapeutics, South San Francisco, CA, ²Parkinson's Disease and Movement Disorders Clinic, Bangalore, India, ³Manipal Hosp., Bangalore, India, ⁴MedGenome Labs Pvt Ltd, Bangalore, India, ⁵All India Inst. of Med. Sci. (AIIMS), New Delhi, India, ⁶Medanta Hosp., New Delhi, India, ⁷Medanta, The Medicity, Gurgaon, India, ⁸Nizams Inst. of Med. Sci. (NIMS), Hyderabad, India, ⁹Inst. of NeuroSci.s Kolkata, Kolkata, India, ¹⁰Bangur Inst. of NeuroSci.s and Inst. of Post Graduate Med. Ed. and Res. (IPGME&R), Kolkata, India, ¹¹Jaslok Hosp. and Res. Ctr., Mumbai, India, ¹²Natl. Inst. of Mental Hlth. and NeuroSci.s (NIMHANS), Bangalore, India, ¹³Shree Krishna Hosp. and Pramukhaswami Med. Coll., Bhaikaka Univ., Karamsad, India, ¹⁴All India Inst. of Med. Sci., Rishikesh, India, ¹⁵Dept. of Neurology, Inst. of Med. Sci., Banaras Hindu Univ., Varanasi, India, ¹⁶Parkinson and Ageing Res. Fndn., Bangalore, India

Abstract Body:

Background: Recent studies have advanced our understanding of the genetic drivers of Parkinson's Disease (PD). Rare variants in more than 20 genes are generally accepted to be causal for PD, and the latest PD GWAS study identified 90 independent risk loci. However, there remains a gap in our understanding of PD genetics outside of the European populations in which the vast majority of these findings occurred. Limited, small-scale studies in East and South Asians have demonstrated that distinct risk factors may exist in these populations, emphasizing the need to characterize the genetic factors driving PD in these groups directly.

Methods: PD subjects with age of onset ≤ 50 years (encompassing juvenile, young, or early-onset PD) were recruited from 10 specialty movement disorder centers across India over a 2-year period. PD cases (N = 675) were genetically profiled via the South Asian Research Genotyping Array (SARGAM) and WES at 60x average coverage. Ancestry-matched controls (N = 1363) were derived from the reference WGS GenomeAsia 100K population, Phase 2 (GASPh2). GWAS of common variants (MAF > 1%) for PD diagnosis and symptom age of onset was performed using merged SARGAM PD case data and GASPh2 WGS control data. PD case WES data was merged with WGS control data, adjusted for platform and/or coverage differences, and used to quantify differential burden of rare (MAF < 1%) and predicted deleterious (per PolyPhen2 and CADD) variants. Finally, WES data was also used to identify PD cases harboring pathogenic/likely pathogenic variants in 21 previously identified monogenic PD genes.

Results: Common variant GWAS of PD diagnosis yielded a GW significant result (lead SNP p-value = 4.1E-11) in a region containing the canonical PD gene *SNCA*. This signal strongly colocalized (posterior probability = 0.88) with *SNCA* region signal from European PD GWAS. Numerous additional loci reached a suggestive significance (p < 1E-5) threshold for both PD diagnosis and age of onset phenotypes; GSEA of genes downstream of associated SNPs (via eQTL and chromatin interaction mapping) implicated synaptic transmission and immune cell migration-related pathways. Pathogenic variants, including homozygous *PRKN* deletions, were identified in 9.8% of PD cases. We will also report on top-ranked genes from WES-based gene burden studies of LoF and deleterious variants.

Conclusions & Future Directions: This study constitutes the largest genetic investigation of PD and first demonstration of *SNCA* association with PD in an Indian population to date. Ongoing work will expand

ASHG 2022 Annual Meeting Poster Abstracts

this cohort by an additional 1,000 PD cases, enabling improved statistical power to detect PD genes in this understudied group.

Complex Traits Posters - Thursday

PB1664. The genetic risk score affects the lifetime blood pressure and obesity modify the effect of genetic risk

Authors:

H-Y. Park¹, Y. Kim², B-J. Kim², N-K. Lim³; ¹Korea NIH, Cheongju, Korea, Republic of, ²Korea NIH, Cheongju-si, Korea, Republic of, ³Korea NIH, Osong-eup, Heungduk-gu, Cheongju, Korea, Republic of

Abstract Body:

Background: Both of obesity and adverse genetic predisposition are known to be major risk factors for hypertension. This study aimed to evaluate the effect of polygenic risk score (PRS) on life-time blood pressure (BP) and whether the genetic risk of hypertension is modified by obesity. **Methods:** Our study was based on a large, prospective, community-based cohort in Korea. Of 10,030, 7,603 participants aged from 40 to 69 at baseline were selected for this analysis. The genotype data were derived from genome-wide association study using the Korea Biobank Array (referred to as KoreanChip) comprised >833,000 markers including >247,000 rare-frequency of functional variants estimated from >2,500 sequencing data in Koreans. Two weighted polygenic risk scores (PRSs) was calculated by summing the number of risk alleles for 86 and 67 single-nucleotide polymorphisms associated with systolic and diastolic blood pressures, respectively. Participants were grouped by body mass index (BMI) as follows: normal, BMI <23; overweight, BMI 23-24.9; obesity, BMI ≥25. And we categorized PRSs into 5 groups based on percentile of PRS (<10%, 10-30%, 30-70%, 70-90%, and >90%). **Results:** Of 7,603 participants, the 4,229 incident hypertension were observed during a median follow-up time of 15.5 years. In blood pressure trajectories using a mixed model, the PRS affects the systolic BP over the life course, and obesity affect the BP in all PRS groups. The diastolic BP showed the similar pattern. Interestingly the diastolic BP was peaked in earlier age in highest PRS group. Compared with having a 20 to 80 percentile of PRS range, the individuals having the PRS in the highest 90% and the lowest 10% had about 2.7 years earlier and 2.7 years later of onset hypertension, respectively. The association between obesity and hypertension was modified by the PRS for systolic blood pressure (P for interaction=0.0053). The effect of obesity on hypertension was greater in individuals with the lowest PRS groups (HR, 1.57; 95% CI, 1.36-1.81 per category of obesity) compared those with the highest PRS group (hazard ratio, 1.38; 95% CI, 1.23-1.54). **Conclusions:** We demonstrated that genetic information can improve prediction of onset age of hypertension and lifetime BP, and the appropriate weight control can modify the onset age of hypertension.

Complex Traits Posters - Wednesday

PB1665. The heritable component of human longevity is highly polygenic and pleiotropic.

Authors:

R. Jabal-Ameli, J-R. Lin, Q. Zhang, J. Mitra, Z. Wang, N. Nguyen, M. Khusidman, G. Atzmon, S. Milman, J. Vijg, N. Barzilai, Z. Zhang; Albert Einstein Coll. of Med., New York, NY

Abstract Body:

The highly polygenic nature of human longevity renders cross-trait pleiotropy an indispensable feature of its genetic architecture. Leveraging the genetic correlation between the aging-related traits (ARTs), we sought to model the additive variance in lifespan as a function of cumulative liability from pleiotropic segregating variants. We tracked allele frequency changes as a function of viability across different age bins and prioritized 22 variants with an immediate implication on the lipid metabolism, body mass index (BMI) and cognitive performance, among other traits. Given the highly complex and non-linear interactions between the genetic determinants of longevity, we reasoned that a composite polygenic score would approximate a substantial portion of the variance in lifespan and developed the integrated longevity genetic scores (*iLGSs*) for distinguishing exceptional survival. We showed that coefficients derived from our ensemble model could potentially reveal an interesting pattern of genomic pleiotropy specific to lifespan. We assessed the predictive performance of our model for distinguishing the enrichment of exceptional longevity among long-lived individuals in two replication cohorts and showed that the median lifespan in the highest decile of our composite prognostic index is up to 4.8 years longer. Finally, using the proteomic correlates of *iLGS*, we identified protein markers associated with exceptional longevity irrespective of chronological age and prioritized drugs with repurposing potentials for gerotherapeutics. Together, our approach demonstrates a promising framework for polygenic modeling of additive liability conferred by ARTs in defining exceptional longevity and assisting the identification of individuals at higher risk of mortality for targeted lifestyle modifications earlier in life. Furthermore, proteomic correlates of *iLGS* highlighted the functional pathway upstream of the PI3K-Akt that can be effectively targeted to slow down aging and extend lifespan.

Complex Traits Posters - Thursday

PB1666. *The impact of copy number variants in the genetic testing yield of LQTS patients.*

Authors:

R. Celeggin¹, G. Meneguzzo², R. Bariani¹, A. Zorzi¹, M. Cason¹, G. Brunetti¹, M. Bueno Marinas¹, C. Basso¹, D. Corrado¹, P. Sarto², B. Bauce¹, K. Pilichou¹; ¹Dept. of Cardio-Thoraco-Vascular Sci. and Publ. Hlth., Univ. of Padua, Padua, Italy, ²Sports Med. Unit, AULSS 2, Treviso, Italy

Abstract Body:

Background. The long QT syndrome (LQTS) is an inherited cardiac disorder characterized by prolonged QT interval on the surface electrocardiogram. It affects 1:2500 individuals, causing lethal ventricular tachyarrhythmias. A causative variant is identified in up to 75% of cases and detection rate for copy number variants (CNVs) among genotype elusive patients by traditional analysis, seem to be around 2-11.5%. Methods. Our cohort of 96 consecutive LQTS index cases, underwent DNA sequencing on the MiSeq platform (Illumina) using the Trusight Cardio panel. CNVs calling was performed by using a modified version of the Exome Depth package. Putative CNVs identified by NGS were confirmed by Multiplex Ligation dependent Probe Amplification. Literature analysis followed the search keyword strings: copy number variants OR CNVs AND long QT syndrome OR LQTS. Results. In our cohort almost 40% (37/96) of LQTS patients carried a putative causative point variant in LQTS related major genes. Noteworthy, our analysis revealed that 5 out of the 59 (8.4%) genotype negative patients carried CNVs, either deletions or duplications. Specifically, we identified 2 deletion on KCNQ1, 1 deletion and 1 duplication on KCNH2 and a duplication of both KCNE1 and KCNE2. Cascade genetic screening revealed the co-segregation of these variants with the clinical phenotype. Conclusion. Genetic screening in LQTS patients should be implement by the research of structural variants such as CNVs. Their prevalence in the disease allow to increase the yield of the genetic testing of about 5%, and offer the chance to better investigate asymptomatic family members.

Complex Traits Posters - Wednesday

PB1667. The impact of exercise on gene regulation in association with complex trait genetics

Authors:

N. Vetr, N. Gay, S. Montgomery; Stanford Univ., Stanford, CA

Abstract Body:

Recurrent exercise can influence risk for a range of complex genetic diseases. However, the molecular basis of exercise's impact on gene function and subsequent disease risk across the body is largely restricted to easily biopsied tissues. Gene expression data across 15 distinct tissues collected through MoTrPAC's rat exercise training experiment provides a unique opportunity to clarify how exercise can affect tissue-specific gene expression and how these adaptations may impact complex disease genes in humans. We integrate this multi-tissue atlas of gene expression changes with gene-disease targets, genetic regulation of expression, and trait relationship data in humans. Consensus from multiple approaches prioritized specific tissues and genes where exercise impacts gene expression and can modify disease risk. Specifically, we identify a total of 5,523 trait-tissue-gene triplets affected by exercise that correspond to effects across a range of 114 cardiometabolic, immunological, endocrine, and other phenotypes.

Complex Traits Posters - Thursday

PB1668. The impact of HIV-1 Vpr on host transcription in macrophages.

Authors:

M. Virgilio, W. Disbennett, T. Chen, J. Lubow, J. D. Welch, K. L. Collins; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Vpr is a highly conserved HIV accessory protein that is necessary for optimal replication in macrophages, but its mechanism of action is poorly understood. Studies using human lymphoid tissue, which are rich in both T cells and macrophages, have found that loss of Vpr decreases virus production but only when the virus strain used is capable of efficiently infecting macrophages. These studies provide evidence that Vpr enhances infection of macrophages and increases viral burden in tissues where macrophages reside. Because Vpr is packaged into the virion and localizes to the nucleus, it may enhance early viral replication events. However, *vpr*-null virus in which Vpr protein is provided by trans-complementation in the producer cells replicates poorly compared to wild-type virus in mononuclear phagocytes. Thus, there is genetic evidence that Vpr's role in the HIV replication cycle continues into late stages. Our group has demonstrated that Vpr counteracts mannose receptor, a macrophage specific restriction factor that targets Env and Env-containing virions for lysosomal degradation. Mannose receptor protein levels are inversely proportional to the amount of viral output from macrophages, affecting spread to neighboring cells. Importantly, Vpr reduces transcription of the gene encoding mannose receptor, *MRC1*, as well as other innate immune genes through an unknown mechanism. Vpr's ability to specifically alter transcription levels of these genes led us to investigate the role of Vpr on global transcription in macrophages from three independent donors using single-cell RNA sequencing. Differential gene expression analysis between *vpr*-WT, and *vpr*-null infected cells revealed thousands of genes were downregulated in the presence of Vpr. Motif scanning of transcriptionally repressed genes revealed several transcription factors (TFs) of interest that Vpr may be targeting to reduce transcription in macrophages. Binding motifs for the master myeloid regulator, PU.1, were found in many suppressed genes, including two known HIV restriction factors - *MRC1* and *IFITM3*. Vpr but not other accessory proteins was able to reduce PU.1 levels in cells that both endogenously and exogenously express PU.1, including primary macrophages. Interestingly, Vpr-dependent degradation of PU.1 was successfully demonstrated with Vpr proteins from several HIV molecular clones as well as from evolutionarily similar Simian Immunodeficiency Virus molecular clones, suggesting degradation of PU.1 may be an evolutionarily conserved function of Vpr. Studies are ongoing to determine whether Vpr targets specific host TFs to alter pathways that influence viral infection and spread.

Complex Traits Posters - Thursday

PB1669. The interacting effects of environmental factors and polygenic risk scores on BMI: application to the multiethnic GENNID family study

Authors:

J. Wan, B. WU, L. Simon, A. Freedland, T. Norden-Krichmar, K. Edwards, American Diabetes Association GENNID Study Group; Univ. of California, Irvine, CA

Abstract Body:

Objective: Polygenic scores (PGS) leverage the effects of genome-wide variants to predict risk of chronic conditions such as obesity and type 2 diabetes (T2D). Environmental factors (EF) such as smoking, diet, physical activity level, and drinking are known risk factors. With the recent developments in PGS as a diagnostic tool, it is important to include EFs and their potential interactions with PGS to enhance the prediction of obesity. We use a multi-ethnic sample of families with T2D to: (1) evaluate the joint effects of a common-variant genetic burden PGS and EF on the body mass index (BMI); (2) to evaluate the prediction of obesity with different models involving the addition of the EFs and PGS-by-EF interaction terms.

Methods: Using 1502 subjects in 259 families from European-American (EA), Mexican-American (MA), African-American (AA), and Japanese-American (JA) families with T2D in the GENNID study, we constructed PGS with the Clumping and Thresholding (C + T) method based on a multi-ethnic GWAS meta-analysis. For each EF, we evaluated the joint effects of PGS and EF on BMI adjusting for age, sex, and self-reported diabetes status using a linear mixed effects model with random intercepts to account for within-family clustering. A pseudo- R^2 summarized the percentage of BMI variance explained by the addition of EF and PGS x EF interaction, and approximate p-values were computed based on a normal distribution. ROC curves and AUC values based on bootstrap resampling compared the ability of different models to predict obesity (BMI>30).

Results: For each ethnic group, the single environmental factor models that explained the most additional variance of BMI beyond PGS were: 2.1% (smoking now, AA), 2.6% (physical activity, EA), 2.5% (liquor drinking now, JA), and 1.4% (low fat diet, MA). The most additional variance of BMI was explained by including these interactions: 1.0% (PGS x diabetic diet, AA), 0.2% (PGS x high fiber diet, EA), 2.4% (PGS x low calorie diet, JA), and 0.5% (PGS x wine now, MA). By adding each EF to the PGS model, prediction beyond PGS improved with at most an AUC increase of the following: 1.1% (high fiber diet, AA), 0.8% (diabetic diet, EA), 2.2% (high fiber diet, JA), and 1% (low fat diet, MA). Adding these interactions to the joint PGS and EF model at most increased the AUC by: 1.4% (PGS x packs per day, AA), 0.8% (PGS x diabetic diet, EA), 1.5% (PGS x low calorie diet, JA), and 1.2% (PGS x low fat diet, MA).

Conclusions: In general, each ethnic group had a distinct environmental risk factor profile. The inclusion of particular environmental factors and their interactions with PGS moderately increased the explained proportion of BMI variance and prediction.

Complex Traits Posters - Wednesday

PB1670. The kynurenine pathway in Alzheimer's disease: A systematic review and meta-analysis.

Authors:

M. Inam, B. S. Fernandes, Z. Zhao; Univ. of Texas Hlth.Sci. Ctr. at Houston, Sch. of BioMed. Informatics, Houston, TX

Abstract Body:

IntroductionThe kynurenine pathway has been increasingly attracting attention as a relevant pathway in neurological disorders, including Alzheimer's disease (AD). Early detection of such metabolic alterations may allow for timely intervention and management, potentially decreasing treatment failure rates. Here, we conducted a systematic review and meta-analysis of the kynurenine pathway metabolites from blood samples in AD.

MethodsPubMed and EMBASE databases were searched from journal inception to April 2022 to identify peer-reviewed case-control studies that assessed kynurenine metabolites in AD compared to healthy controls (HC) in peripheral blood, namely, tryptophan (TRP), kynurenine (KYN), kynurenic acid (KA), quinolinic acid (QA), and 3-hydroxykynurenine (3-HK). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed. The random effects model parameter was selected when comparing the standardized mean differences (SMD) between groups.

ResultsData were extracted from 18 articles that met the inclusion criteria. The mean sample size for each study for individuals with AD and HC was 32 ± 25 and 37 ± 36 , and the mean age was 74.1 ± 10.5 and 72.7 ± 9.2 , respectively. TRP levels ($n=17$) were decreased in AD (SMD: -0.83, CI95% [-1.15 to -0.50], $p<0.001$). KA levels ($n=8$) were also decreased (SMD: -0.27, CI95% [-0.46 to -0.08], $p=0.005$). There were no significant differences in KYN ($n=10$), 3-HK ($n=7$), and QA ($n=5$) levels, but a significant increase in the KYN/TRP ratio ($n=7$) was observed in AD compared to HC (SMD: 0.55, CI95% [0.04 to 1.05], $p<0.030$).

ConclusionTryptophan, the precursor of the kynurenine pathway, is decreased in individuals with AD. KA, which has neuroprotective effects, is also decreased in AD. The KYN/TRP ratio is increased, suggesting a shift in tryptophan degradation toward the kynurenine pathway. Drugs that target the kynurenine pathway, particularly KA, might be useful in AD treatment. Future studies should validate and explore the utility of this peripheral biomarker as a treatment target.

Complex Traits Posters - Thursday

PB1671. The Largest Asian GWAS for Systemic Sclerosis Identified a Novel High Risk Candidate Causal SNP in the Fcγ-Receptor Gene Region.

Authors:

Y. Ishikawa¹, Systemic Sclerosis Working Group of Japan Ministry of Health, Labor and Welfare, C. TERAO^{1,2,3}; ¹Riken, Ctr. for Integrative Med. Sci., Yokohama, Japan, ²Shizuoka Gen. Hosp., Shizuoka, Japan, ³Sch. of Pharmaceutical Sci., Univ. of Shizuoka, Shizuoka, Japan

Abstract Body:

Objectives: While GWASs have clarified genetic architectures of systemic sclerosis (SSc), the Eurocentric nature of the studies have limited application of the findings to Asians. This study was conducted aiming to identify novel causal SNPs for SSc specific to Japanese as well as those shared with Europeans and to clarify mechanistic effects of the SNPs on SSc pathogenesis. **Methods:** A total of 1,499 cases and 112,609 controls were enrolled in the study leading to the ever-largest Asian GWAS for SSc. The imputation was conducted using the reference panel of the phase 3v5 1,000 genome project data combined with a high-depth WGS data of 3,256 Japanese subjects. We conducted logistic regression analyses and combined the Japanese GWAS results with those of Europeans by an inverse-variance fixed-effect model. Polygenicity and enrichment of functional annotations were evaluated by LDSC, Haploreg and IMPACT programs. We also constructed PRS to predict SSc development. **Results:** We identified three (*FCRLA-FCGR*, *TNFAIP3*, *PLD4*) and four (*EOMES*, *ESR1*, *SLC12A5*, *TPIIP2*) novel loci in the Japanese GWAS and the trans-ethnic meta-analysis, respectively. One of the Japanese novel risk SNPs, rs6697139, located within *FCGR* gene clusters had a strong effect size (OR 2.05, P=4.9×10⁻¹¹). We also found the complete LD variant, rs10917688, was positioned in cis-regulatory element and binding motif for an immunomodulatory transcription factor IRF8 in B cells, another genome-wide significant locus in our trans-ethnic meta-analysis and the previous European GWAS. Notably, the association of risk allele of rs10917688 was significant only in the presence of the risk allele of IRF8. Intriguingly, rs10917688 was annotated as one enhancer-related histone marks, H3K4me1, in B cells, implying that *FCGR* gene(s) in B cells may impact on SSc pathogenesis. Furthermore, significant heritability enrichment of active histone marks was found in B cells both in European and Japanese populations by LDSC, implying a shared disease mechanism where abnormal B-cell activation may be one of the key drivers for the disease development. Finally, PRS using effects sizes of the trans-ethnic meta-analysis moderately fit in the development of Japanese SSc (AUC 0.604) with further improvement by adding IMPACT-annotated SNPs for IRF8-binding in the B-cell line (AUC 0.610), paving a path to personalized medicine for SSc. **Conclusion:** Our study identified seven novel susceptibility loci in SSc. Downstream analyses highlighted a novel disease mechanism of SSc where an interactive role of *FCGR* gene(s) and IRF8 may accelerate the disease development and B cells may play a key role on the pathogenesis of SSc.

Complex Traits Posters - Wednesday

PB1672*. The Pan-UK Biobank project improves locus discovery, facilitates genetic architecture comparisons, and increases the resolution and generalizability across diverse populations

Authors:

M. Kanai¹, K. Karczewski¹, R. Gupta¹, K. Tsuo¹, Y. Wang¹, N. Baya¹, R. Walters¹, P. Turley¹, S. Callier², D. Palmer¹, J. Goldstein¹, G. Sarma¹, M. Solomonson¹, N. Cheng¹, W. Lu¹, S. Bryant¹, C. Churchhouse¹, C. Cusick¹, D. King¹, T. Poterba¹, J. Compitello¹, W. Zhou¹, C. Seed¹, M. Daly¹, H. Finucane¹, B. Neale¹, E. Atkinson³, A. Martin¹; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²George Washington Univ., Washington, DC, ³Baylor Coll. of Med., Houston, TX

Abstract Body:

The UK Biobank (UKB) is a collection of a half million individuals that has enabled large-scale genome-wide association studies (GWAS) for common diseases. However, most GWAS in UKB primarily use the individuals of European ancestry. Analyzing more diverse participants improves locus discovery, fine-mapping resolution, and generalizability of polygenic prediction, and facilitates cross-ancestry comparison of genetic architectures. To this end, we initiated the Pan-UKB project, a multi-ancestry analysis of 7,228 biomarker, categorical, continuous, electronic health record, and prescription phenotypes across 441,331 individuals from 6 continental ancestry groups, providing 16,553 GWAS and ancestry-specific LD matrices freely available to the community.

Here, we will present the latest results of the Pan-UKB, including heritability-based phenotype QC, cross-ancestry meta-analysis of high-quality phenotypes, polygenic prediction and fine-mapping. First, we conducted extensive heritability analyses using S-LDSC and RHE-mc. We identified 317 traits with significant out-of-bounds RHE-mc heritabilities; these were enriched for likely confounded traits (e.g. UK geographic location, ethnic background). We incorporated this filter into a systematic QC pipeline to prioritize high-quality phenotypes and prevent noise from low heritability GWAS in downstream analyses.

Using 527 high-quality phenotypes passing QC in EUR and at least 1 non-EUR ancestry, we performed meta-analyses that only incorporated QC-passing ancestry-trait pairs to reduce noise while maintaining power. We identified 87,217 genome-wide significant associations, of which 3,840 were unique to non-EUR ancestries. We identified 39,416 and 5,909 associations that were not previously reported for the same Experimental Factor Ontology term or GWAS Catalog category respectively, highlighting the importance of diverse ancestries for novel discovery.

Finally, we will present a polygenic prediction using leave-one-ancestry-out meta-analysis to investigate how increasing diversity affects the generalizability of GWAS. We will also demonstrate fine-mapping of cross-ancestry meta-analysis to illustrate how diverse populations improve resolution without loss of calibration when samples are from the same cohort. Together, these efforts vastly extend the available resources for interpretation of disease-associated variants and highlight the importance of diversity in genetic studies. In addition, our new analytical pipelines provide the best practices for post-GWAS analyses across multiple ancestries, facilitating studies in underrepresented populations.

Complex Traits Posters - Thursday

PB1673. The *PPARG1* gene regulates the aging retina by the modulation of mitochondrial functions, autophagy and senescence in a mouse model of age-related macular degeneration

Authors:

J. Blasiak¹, M. Maria², A. Koskela², S. Felszeghy², K. Kaarniranta²; ¹Univ. of Lodz, Lodz, Poland, ²Univ. of Eastern Finland, Kuopio, Finland

Abstract Body:

Age-related macular degeneration (AMD), an eye disease is an important global issue concerning vision loss. We have created a model of AMD pathogenesis in which mitochondrial dysfunctions in the retinal pigment epithelium (RPE) cells interplay with senescence and declined autophagy caused by impaired PGC-1 α , a master regulator of mitochondrial biogenesis and an important factor in antioxidant defense, encoded by the *PPARG1* (peroxisome proliferator-activated receptor gamma coactivator 1-alpha) gene. The aim of this work was to evaluate the expression of genes involved in mitochondrial functions, autophagy and senescence in homo- and heterozygotic *PPARG1*-knockout (KO) mice as compared with control animals. We also assessed the influence of age on the effects induced by *PPARG1* knockout. The 3-months- or 1-year-old *PPARG1*-KO mice were derived from the C57BL/6J strain. The expression of the following genes: *TP53*, *HMGB1* and *CDKN2A* (senescence), *SDH*, *MT-COI* (mitochondria), *MTOR*, *ULK1*, *MAP1LC3B* (*Microtubule Associated Protein 1 Light Chain 3 Beta*), *SQSTM1* (*Sequestosome 1*) and *PRKN* (*Parkin RBR E3 Ubiquitin Protein Ligase*) (*macroautophagy and mitophagy*) in the retinal pigment epithelium (RPE) cells was evaluated by RT-PCR and immunohistochemistry. We observed an increased expression of the *MT-COI* gene in RPE cells in both homo- and heterozygous, 3-months- and 1-year-old animals as compared with corresponding controls. Also, an increased expression of the *HMGB1* gene in 1-year-old homozygous KO mice was observed. The *PPARG1* knockout in both alleles resulted in a decreased expression of the *MTOR* gene, but only in 1 year old animals. The expression of the *MAP1LC3B*, *SQSTM1* and *PRKN* in homozygotes was either at borderline significance or displayed a tendency to decrease. These effects were more pronounced in 1-year- than in 3-months-old animals. In conclusion PGC-1 α deficiency results in accelerated decline of the expression of autophagy-related genes with age and alterations in mitochondria- and senescence-related genes in RPE cells. The mice with knockout in the *PPARG1* gene can be useful as a model to investigate the role of impaired mitochondrial and autophagic functions associated with cellular senescence in AMD pathogenesis. This work was supported by National Science Centre, Poland grant number 2017/27/B/NZ3/00872.

Complex Traits Posters - Wednesday

PB1674. The prevalence, penetrance, and expressivity of mitochondrial disorders in a population-based cohort.

Authors:

T. Hall¹, S. Cannon¹, K. Colclough², M. Weedon¹, K. Patel¹; ¹Univ. of Exeter, Exeter, United Kingdom, ²Royal Devon and Exeter NHS Fdn. Trust, Exeter, United Kingdom

Abstract Body:

The prevalence and penetrance of heteroplasmic pathogenic mitochondrial variants that cause a range of severe multisystem diseases, including diabetes, is unknown in the population. Here, we used 118,365 whole genome sequenced individuals of unrelated European ancestry from UK Biobank to assess the prevalence, penetrance and expressivity of known pathogenic mitochondrial variants in the population. We used Mutect2 to call 95 previously reported pathogenic mitochondrial variants from whole genome sequenced individuals from UK Biobank. Individuals with heteroplasmy $\geq 2\%$ were considered carriers for that variant in line with the clinical diagnostic reporting. We performed association analysis with 30 biomarkers and 7 phenotypes commonly associated with mitochondrial disorders including deafness and diabetes, adjusting for age, sex and principal components. We used $p < 1.42 \times 10^{-5}$ as a significance threshold to correct for multiple testing.

Sixty-eight of 95 variants were seen in at least one individual with $> 2\%$ heteroplasmy in the UKbiobank. The prevalence of pathogenic variants ranged from 1 in 118,346 to 1 in 330. None of the pathogenic variants were associated with mitochondrial related phenotype and 30 blood biomarkers with $p < 1.42 \times 10^{-5}$ except *m.3243A>G*. The well-known *m.3243A>G* variant was present in 0.068% (81/118,346) individuals and was associated with diabetes (OR=3.01, 95%CI:1.79-5.05, $P < 4 \times 10^{-5}$). This association with diabetes increased with higher heteroplasmy $\geq 10\%$ (OR=10.41, 95%CI:3.99-27.21, $P < 2 \times 10^{-6}$; OR=3.38, 95%CI:1.12-9.22, $P = 0.017$ respectively).

Mitochondrial mutations are common in the population and have greatly reduced penetrance with low health-related burden. The variant *m.3243A>G* is associated with diabetes and its penetrance is dependent on heteroplasmy levels.

Complex Traits Posters - Thursday

PB1675. The relationship of Alzheimer disease, genetic risk, and education in individuals of African Ancestry.

Authors:

B. Akgun¹, F. RAJABLI¹, L. D. Adams¹, T. D. Starks², R. Laux³, P. Whitehead-Gay¹, B. W. Kunkle^{1,4}, A. Caban-Holt², K. F. McInerney⁵, M. L. Cuccaro^{1,4}, J. M. Vance^{1,4,5}, J. L. Haines³, G. S. Byrd², G. W. Beecham^{1,4}, M. A. Pericak-Vance^{1,4,5}, A. Seixas⁶; ¹John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, ²Maya Angelou Ctr. for Hlth.Equity, Wake Forest Sch. of Med., Winston-Salem, NC, ³Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, ⁴Dr. John T Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, ⁵Dept. of Neurology, Univ. of Miami Miller Sch. of Med., Miami, FL, ⁶Dept. of Psychiatry and Behavioral Sci., Univ. of Miami Miller Sch. of Med., Miami, FL

Abstract Body:

Background: Alzheimer Disease (AD) is the most common cause of dementia in the elderly with a significant genetic risk. However, recent evidence highlight a more complex etiology, one that includes genetic, psychosocial, environmental, and behavioral risk factors. With the growing prevalence of AD globally, especially among racial ethnic groups, there is dire need to identify factors that reduce risk. Higher levels of educational attainment associate with a lower risk for AD, possibly due to increased cognitive reserve. It is unclear whether protective factors like education modifies AD genetic risk. Polygenic risk scores (PRS) provide a potential method to identify individuals with low and high genetic risk of developing AD. In this study, we assessed the relationship between educational attainment and the risk of developing AD in African Ancestry (AA) individuals with low and high AD PRS risk.

Methods: We analyzed 537 AA individuals (AD=156, cognitively unimpaired (CU)=381). First, we classified individuals into two educational categories: individuals with a “high school” education and individuals without a “high school” education. Second, we calculated PRS using summary statistics from the AA genome-wide association study (Kunkle et al. 2021). Based on the PRS scores, we created two groups: individuals with low genetic risk of AD (lower quartile, N=138) and individuals with high genetic risk of AD (upper quartile N=128). Within these two groups, we assessed the relationship between education and AD using the Chi-Square test (once for the low risk group and once for high).

Results: We found a significant association between higher educational attainment and reduced risk of AD in individuals with low genetic risk (OR=0.16, p-value=0.005). However, higher educational attainment was not significantly associated with decreasing risk of AD in the sub-group of individuals with high genetic AD risk (OR=0.75, p-value=0.63).

Conclusions: Our results showed that high educational attainment in AA individuals with low genetic risk reduced AD risk more than in those with a high genetic risk, thereby diminishing education’s protective effect. AD is a heterogeneous disease in which genetic, environmental and lifestyle factors contribute to risk. Thus, understanding not only the individual contribution of these factors to risk but also the impact of their coexistence can lead to more accurate risk identification and intervention strategies for AD.

Complex Traits Posters - Wednesday

PB1676. The role of host genetics on weight gain following ART initiation in people living with HIV

Authors:

T. Jia¹, K. M. Jordahl¹, S. A. Ruderman¹, R. M. Nance¹, C. B. Haas¹, A. L. Willig², A. R. Webel¹, B. M. Whitney¹, J. A. Delaney³, H. M. Crane¹, I. Peter⁴, S. Lindstrom¹; ¹Univ. of Washington, Seattle, WA, ²Univ. of Alabama at Birmingham, Birmingham, AL, ³Univ. of Manitoba, Winnipeg, MB, Canada, ⁴Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Studies suggest that certain classifications of antiretroviral therapy (ART) contribute to weight gain in people with HIV (PWH). The underlying mechanism by which ART interacts with host genetics to affect weight change is unclear. Using data from the multi-ancestry CFAR Network of Integrated Clinical Systems (CNICS) cohort, we assessed the role of host genetic factors in ART-related weight gain among PWH. We assessed weight change in 836 ART-naïve PWH initiating common ART regimens (non-nucleoside reverse transcriptase inhibitors [NNRTIs], N=256, reference; protease inhibitors [PIs], N=108; first-generation integrase inhibitors [II1], N=295; second-generation integrase inhibitors [II2], N=177) between 2012 and 2020. We examined weight change in the immediate period (first 6 months) after ART initiation. We obtained an average of 3.27 weight measurements per PWH for a total of 2,731 observations available for analysis. We generated genome-wide genotype data using the MEGA array. We assessed if ART-related weight gain differed by (i) *CYP2B6* or *UGT1A1* metabolizer genotypes or by (ii) a polygenic risk score (PRS) for body mass index (BMI). We used linear mixed models with exchangeable correlation matrices and robust standard errors to account for within-subject correlations, random intercepts, and random slopes for time on ART. All models were adjusted for nadir CD4 count (cells/mm³), CNICS site, time on ART, and the interaction of time on ART and regimen. Differences in weight gain by host genetics were assessed by including interaction terms. Compared to individuals on NNRTI, individuals who were on a PI regimen had the largest weight gain of 4.82 kg/6 months (95% CI: 2.22-7.42, $p < 0.001$) followed by II2: 2.73 kg/6 months (95% CI: 0.89-4.57, $p = 0.004$). There was no significant interaction with either *CYP2B6* or *UGT1A1* metabolizer genotypes with ART regimens (all $p > 0.05$). For individuals on a PI regimen, we observed the largest weight gain among those in the lowest BMI PRS tertile (8.41kg, 95% CI: 4.81-11.99) compared to those in the highest tertile (3.65kg, 95% CI: -0.54-7.85), although this difference was not significant ($p = 0.12$). Similarly, for individuals on II2 regimen, those in the lowest BMI PRS tertile gained the most weight (4.75kg, 95% CI: 0.52-8.98) compared to those in the highest tertile (1.15kg, 95% CI: -1.32-3.62), but this difference was not significant ($p = 0.21$). Overall, we found limited evidence that host genetic variation involved in drug metabolism and adiposity play a major role in ART-related weight gain.

Complex Traits Posters - Thursday

PB1677. The role of *POLG* in the genetic etiology of Amyotrophic Lateral Sclerosis

Authors:

N. Russell, K. Russell, L. Jorde; Univ. of Utah, Salt Lake City, UT

Abstract Body:

Amyotrophic lateral sclerosis (ALS) is a debilitating and fatal neurodegenerative disease that results in the death of motor neurons of the brain and spinal cord. Approximately 5,000 people in the U.S. are diagnosed with ALS each year, and no effective treatments exist. Previous studies have estimated the heritability of ALS to be approximately 60%. However, only 17% of ALS cases can be attributed to known ALS-causing genes. This leads to a large potential to uncover novel genes that contribute to ALS. Using two publicly available ALS cohorts and ALS patients from our University of Utah cohort, we used VAAST and Phevor to identify genes burdened with rare deleterious variants in ALS patients compared to a control cohort. Among the three cohorts, we analyzed 943 whole-genome sequences and 140 whole-exome sequences. Our highest ranked gene for both public ALS cohorts and the third highest rank gene in the Utah cohort was DNA Polymerase Gamma (*POLG*). Within *POLG*, there were 28 rare (MAF < 0.001) nonsynonymous single nucleotide variants (SNVs) found across the three cohorts. *POLG* encodes the catalytic subunit of the mitochondrial DNA polymerase. *POLG* has been implicated in numerous mitochondrial DNA syndromes and neurological diseases, but never before in ALS. However, studies from our lab have implicated mitochondrial function in ALS etiology, and *POLG*'s function in mitochondria DNA replication provides support for its role in ALS. We also acquired RNA-sequencing data from one of the public ALS cohorts and evaluated the expression of mitochondrial expressed genes. We found that ALS patients with a *POLG* mutation had, on average, higher expression for every mitochondrial expressed gene. This is significant because of the role of *POLG* in mitochondrial DNA maintenance. Additionally, we examine the read depth of mitochondrial DNA in the ALS patients to determine if there is a difference in these cases compared to a control cohort. *POLG* represents a promising candidate in the genetic etiology of ALS and warrants future functional studies to support the role of *POLG* in disease cause and progression.

Complex Traits Posters - Wednesday

PB1678. The variant in *FLNA* identified in a patient with progressive supranuclear palsy

Authors:

K. Kume¹, I. Yuishin², M. Oda³, K. Komatsu⁴, M. Takahashi⁵, Y. Tada¹, M. Kamada⁶, H. Kawakami¹; ¹Res. Inst. for Radiation Biology and Med., Hiroshima Univ., Hiroshima, Japan, ²Tokushima University, Tokushima, Japan, ³Vihara Hananosato Hosp., Mihara, Japan, ⁴Kitano Hosp., Osaka, Japan, ⁵Kitano Hosp., Osaka, Japan, ⁶Kagawa Univ., Kita-gun, Japan

Abstract Body:

Background: Progressive supranuclear palsy (PSP) is a neurodegenerative disorder characterized by tufted astrocyte and accumulation of 4-repeat tau. Because the clinical symptoms are broad, it is difficult to differentiate PSP from other tauopathies including corticobasal degeneration and Alzheimer's disease. In addition, the pathophysiology of PSP remains unknown. Recently, it has been reported that the variants in *FLNA*, encoding filamin-A, are associated with PSP. **Purpose:** The aim of this study is to confirm the significance of the variants of *FLNA* in PSP. **Methods:** Exome sequencing was performed for nine patients with PSP and five patients with corticobasal syndrome (CBS). The data analysis was performed according to the GATK best practices. The identified variant was validated by Sanger sequencing. **Results:** One patient with PSP had a variant (NM_001110556:c.7055C>T,p.S2352F). The variant is reported as a rare variant in the gnomAD (allele frequency: 0.000011) and predicted as pathogenic by several tools (CADD score: 25, PolyPhen-2: possibly damaging, MutationTaster: disease causing). In our variant database, a patient with spinocerebellar ataxia had the same variant. No variants in *FLNA* were identified in patients with CBS. **Discussion:** Our findings showed that the variant in *FLNA* may contribute to the development of PSP.

Complex Traits Posters - Thursday

PB1679. The whole blood transcriptional regulation landscape in 465 COVID-19 infected samples from Japan COVID-19 Task Force

Authors:

Q. Wang¹, R. Eda¹, H. Namkoong², T. Hasegawa³, M. Ishii², R. Koike³, A. Kimura³, S. Imoto⁴, S. Miyano³, S. Ogawa^{5,6}, T. Kanai^{2,7}, K. Fukunaga², Y. Okada^{1,8,4}, the Japan COVID-19 Task Force; ¹Osaka Univ., Osaka, Japan, ²Keio Univ., Tokyo, Japan, ³Tokyo Med. and Dental Univ., Tokyo, Japan, ⁴the Univ. of Tokyo, Tokyo, Japan, ⁵Kyoto Univ., Kyoto, Japan, ⁶Karolinska Inst., Stockholm, Sweden, ⁷Japan Agency for Med. Res. and Dev., Tokyo, Japan, ⁸RIKEN, Yokohama, Japan

Abstract Body:

Studies of gene expression dynamics and regulation landscape in Coronavirus disease 2019 (COVID-19) infected individuals are limited. Here, we report on a thorough analysis of whole blood RNA-seq data from 465 genotyped samples from the Japan COVID-19 Task Force, including 359 severe and 106 non-severe COVID-19 cases.

We discovered 1,169 putative causal ($0.9 < \text{posterior inclusion probability} = \text{PIP}$) expression quantitative trait loci (p-causal eQTLs) including 34 possible colocalizations with biobank fine-mapping results of hematopoietic traits in a Japanese population, 1,549 p-causal splice QTLs (sQTLs; e.g. two independent sQTLs at *TOR1AIP1*), as well as biologically interpretable trans-eQTL examples (e.g., *REST* and *STING1*).

We validated our cis-eQTL fine-mapping result by comparing it with that of another cohort (GTEx) and observing near half (46%) replication rate of p-causal eQTLs, as well as 100% concordance in the effect size direction when replicated. In addition, we show that integrating fine-mapping results from two cohorts improves the power (396 p-causal eQTLs are missing in GTEx and are as functionally enriched as the ones existing) and accuracy (eQTLs with inconsistent PIP between two cohorts are functionally depleted).

We performed differential gene expression analysis to elucidate 198 genes with increased expression in severe COVID-19 cases, enriched for innate immune-related functions (adjusted $p < 10^{-10}$ for the term "Immune System"). We also evaluated the non-zero but limited effect of COVID-19 on eQTL discovery, where only six out of >100K variant-genes tested were off by more than 2 in $-\log_{10}(p)$ scale when including COVID-19 severity as a covariate in the eQTL call.

13 genes, including the ones with known relevance to viral infection (e.g. *CLEC4C*), showed eQTL effects of different magnitudes for different disease severity ($\text{FDR} < 0.05$). We characterized such COVID-19 severity interaction-eQTLs (ieQTLs) in the context of cell type specificity by performing cell type decomposition and testing ieQTL effects for each of the inferred cell type composition. This suggested dynamics of cell type composition such as an increase in neutrophils along with the COVID-19 severity as a major mechanism (eQTL effects in 10/13 genes interacted with inferred neutrophil counts; Bonferroni $p < 0.05$).

Altogether, our study provides a comprehensive catalog of whole blood regulatory variants in Japanese together with a reference for transcriptional landscapes in response to COVID-19 infection, highlights the presence of ieQTLs, and serves as a showcase of improving eQTL fine-mapping performance by including diverse populations.

Complex Traits Posters - Wednesday

PB1680. Tissue-associated examination of risk factors uncovers genetic risk profiles that vary across patients of inflammatory bowel disease.

Authors:

A. Gaité¹, U. Marigorta^{1,2}, M. Sanchez-Mayor¹, Z. Mars¹, L. Bujanda^{3,4}; ¹Integrative Genomics Lab, Ctr. for Cooperative Res. in BioSci.s (CIC bioGUNE), Basque Res. and Technology Alliance (BRTA), Bizkaia Technology Park, Derio, Spain, ²IKERBASQUE, Bilbao, Spain, ³Biodonostia, San Sebastian, Spain, ⁴Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas CIBERehd, Barcelona, Spain

Abstract Body:

Polygenic risk scores (PRSs) are emerging as a tool of choice to summarize the risk of disease in each individual. The calculation of PRSs is based on genetic variants and effect sizes obtained from large studies using many different cohorts of patients. In inflammatory bowel disease (IBD), meta-analysis based on over 75.000 IBD patients from different multi-ethnic cohorts has discovered around 300 independent associated variants. This information can be used to estimate genetic risk and detect individuals with several higher probability of developing the disease during their lifetime. However, PRSs usually do not incorporate functional information about causal variants, and the potential for heterogeneity in effect sizes among subgroups of individuals is disregarded entirely. Hence, PRSs might be ill-suited to capture subcomponents of genetic risk that may account for key differences among specific profiles of patients. We use a pipeline of methodologies based on functional genomics to characterize the regulatory implications of genetic signals associated with IBD by Genome-Wide Association Studies. Using genetically-determined levels of gene expression as proxy of risk, we explore the different patterns of association of causal genes across tissues that participate in IBD etiology. We recapitulate sharing of regulatory signals across tissues for the genes involved in IBD susceptibility. Of note, a fraction of modules of co-expressed genes is enriched for disease heritability, although this correlates with the number of genes in each module, which goes in accordance with the known polygenicity of IBD risk. Besides providing novel mechanistic understanding of IBD risk, we use this information to carry out individualized multi-tissue evaluation of the presence of risk factors in each patient. This analysis serves to check the proposition that a fraction of IBD patients is characterized by enrichments for particular components of genetic risk.

Complex Traits Posters - Thursday

PB1681. To annotate the neuroprotective role of Camel α Lactalbumin and Oleic acid (CLOA) complex against Parkinson's induced model.

Authors:

S. Ubaid¹, S. Pandey¹, M. Akhtar²; ¹King George's Med. Univ., Lucknow, India, ²Central Drug Res. Inst., Lucknow, India

Abstract Body:

Parkinson's disease is a neurodegenerative disorder that affects 2-3% of the population ≥ 65 years of age. Currently, the therapeutic interventions that are available for PD have several adverse effects. Therefore, the development of novel therapies for PD based on biosimilar compounds is urgently needed. Camel α -lactalbumin is known to act as a potential therapeutic candidate for brain disorders via regulation of inflammatory and apoptotic pathways, while oleic acid is a monounsaturated fatty acid and is involved in the synthesis of myelin phospholipids as well as triggering the synthesis of dopamine. This study aims to formulate the camel α -lactalbumin and oleic acid complex and evaluate the underlying neuroprotective efficacy of the CLOA complex in MPTP-induced in vitro and in vivo models. Our result demonstrates that EDTA removes Ca^{++} from camel α -LA, which was confirmed by ICP-MS, and the arsenazo III complex was further validated by TEM, which confirms that it attains the nanostructure. The results also revealed that the viability of the SH-SY5Y cell line and the locomotor activity of the rat model were increased. Furthermore, the CLOA complex alleviated oxidative stress, and inflammation increased the level of dopamine and decreased metabolites in the MPTP-induced model. The results were further validated at the proteomic and genomic levels, where the expression of SIRT1 and FOXO3a was upregulated. Taken together, the CLOA complex shows a neuroprotective effect in MPTP-injured rat and SH-SY5Y cells via activation of the SIRT1 signaling pathway, while the neuroprotective efficacy of individual camel α -lactalbumin and oleic acid is less than that of the CLOA complex.

Complex Traits Posters - Wednesday

PB1682. Trans-ancestry genome-wide analysis of kidney disease in individuals with type 2 diabetes.

Authors:

E. Richard, S. Cao, W. Gu, A. Shadyab, R. Salem; Univ. of California San Diego, La Jolla, CA

Abstract Body:

Background: Diabetic kidney disease (DKD) is a common microvascular complication of type 2 diabetes mellitus (T2D) and is the most frequent cause of end-stage kidney disease (ESKD). However, the genetic basis of kidney disease in those with T2D remains largely unknown. **Methods:** Using data from 15 studies from dbGaP and the UK Biobank, we conducted genome-wide association studies (GWAS) for a comprehensive set of 12 dichotomous kidney disease phenotypes based on albuminuria and estimated glomerular filtration rate (eGFR) in up to 58,228 individuals with T2D. GWAS results were combined via ancestry-specific (42,140 European, 8,834 African, 4,346 Hispanic, 2,562 South Asian, and 346 East Asian- ancestry) and trans-ancestry meta-analysis with fixed-effects. MR-MEGA was also used for trans-ancestry analysis to account for heterogeneity in allelic effects between ancestries. Gene prioritization and gene-set enrichment analyses were performed using DEPICT. **Results:** We identified previously reported associations for *PDILT* and *NR1P1* with chronic kidney disease (CKD; eGFR<60ml/min vs eGFR>=60ml/min), and *CUBN*, *GCKR* and *FOXD2* with albuminuria and DKD (albuminuria and ESKD). A novel genome wide significant (GWS, $p < 5 \times 10^{-8}$) signal for a late DKD phenotype (severe albuminuria or ESKD) was found in the 8p23.1 region (rsID-effect allele: rs28549051-T, OR=0.54) near a β -defensin gene cluster. Analysis using MR-MEGA identified novel GWS associations between an intronic variant (rs76989357) in *CNTN4* and a CKD-DKD phenotype (DKD and eGFR<45ml/min) and between an intergenic variant (rs10238973) closest to *CDC14C* and severe albuminuria. In ancestry-specific analyses, we detected two novel GWS variants, rs7007458-A (OR=0.70) in the intergenic region between the *ZMAT4* and *SFRP1* genes for CKD-DKD in those of European ancestry and rs62364704-C (OR=0.40) in an intronic region of *MCTP1* for severe albuminuria in African ancestry individuals. Among European ancestry participants, gene-set analysis of ESKD (vs no ESKD) identified an enrichment for phospholipase A2 activity (GO:0004623, $p < 0.01$), and gene-based tests identified suggestive associations (FDR <0.20) for *GJB5*, *GJB4*, *HAPLN1* and *OSMR* loci with CKD, and for *UACA*, *DPYSL5*, *TRIM54*, *EPB41L5* with DKD. **Conclusion:** Large-scale GWAS analyses coupled with comprehensive kidney disease definitions in subjects with T2D replicated previously reported loci and identified several novel loci that warrant further investigation.

Complex Traits Posters - Thursday

PB1683. Trans-ancestry PRSs of lipid traits: characterization and evaluation of clinical utility in the diverse PAGE study

Authors:

Z. Wang¹, Y. Cai², M. Kim², R. Smit¹, S. Graham³, Y. Hu⁴, S-A. Love⁵, J. Judin⁶, S. Buyske⁷, K. Young⁸, C. Willer⁹, K. North⁸, C. Kooperberg⁴, R. Loos¹⁰, U. Peters¹¹, M. Graff⁸, The Global Lipids Genetics Consortium, the PAGE Study (ZW and YC contribute equally to this work); ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²Fred Hutch Cancer Res. Ctr., Seattle, WA, ³Univ MICHIGAN, Ann Arbor, MI, ⁴Fred Hutchinson Cancer Res. Ctr., Seattle, WA, ⁵The Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁶Fred Hutchinson Cancer Ctr., Seattle, WA, ⁷Rutgers Univ., Piscataway, NJ, ⁸Univ North Carolina, Chapel Hill, NC, ⁹Univ. of Michigan, Ann Arbor, MI, ¹⁰Univ. of Copenhagen, Copenhagen, Denmark, ¹¹Fred Hutchinson CA Res Ctr, Seattle, WA

Abstract Body:

Lipid profiles are heritable risk factors of cardiovascular disease (CVD), a leading cause of mortality worldwide. Accordingly, lipid polygenic risk scores (PRSs) have the potential to predict future risk of CVD. However, PRSs derived from mostly European ancestry individuals often have poor trans-ancestry portability. Systematic assessment of lipid PRSs in diverse populations is critical for a more accurate and equitable application of PRSs to clinical practice. Here, we leverage multi-ancestry PRSs for lipid traits (derived from a population of which is of 78% European - EUR, 15% of South and East Asian, 5% of African - AA, and 1% of Hispanic - HA background) constructed by the Global Lipids Genetics Consortium (GLGC) using a clumping and thresholding method and apply the PRSs to the Population Architecture through Genomics and Environment (PAGE) study, which was excluded from GLGC. The PAGE study includes 71K diverse participants with 33% HA, 32% EUR, 25% AA, 4% Asian, 3% Native Hawaiian - HI, and 1% Native American - NA. We evaluated PRSs for lipids levels (HDL, LDL, TC and TG) using incremental R² after accounting for age, sex, first 10 PCs and self-reported race and ethnicity. The GLGC multi-ancestry PRSs included 10-12K genetic variants, depending on the lipid traits, with 2-3K rare variants (MAF <1% in GLGC multi-ancestry samples). Overall, the PRSs' explained variance (R²) decreased with increasing genetic divergence from populations included in the original GWAS. For LDL, HDL and TC, the PRS performance was best in EUR, followed by HA and AA, with much lower R² in Asian and HI participants (e.g., LDL R²: 16%, 13%, 10%, 7%, 5% in EUR, AA, HA, Asian, and HI). For TG, the highest R² was still in EUR (10%), followed by Asian, HI, and HA (7%, 6%, 5%), and lowest in AA participants (2%). Results were imprecise in NA because of small sample sizes. Stratified analyses showed slightly higher R² for LDL and TC in women (women vs men: 11% vs 8% for LDL and 12% vs 9% for TC), and no difference by age strata (range 17-50, 50-60, 60-70, and 70-80 years). Although the LDL-PRS was significantly associated with incident CVD risk (HR = 1.07, P = 2.2E-6), after adjusting for traditional risk factors, summarized by the Pooled Cohort Equation, we did not observe significant improvement in prediction of 10-year risk of CVD (AUC 0.71 with or without LDL-PRS). Overall, this study highlights the challenges of implementing multi-ancestry PRS in diverse populations. We are expanding this effort by collaborating with additional biobanks to increase sample size in currently underrepresented populations.

Complex Traits Posters - Wednesday

PB1684*. Transcriptome analysis of familial dysautonomia reveals tissue-specific gene expression disruption in the peripheral nervous system

Authors:

R. Harripaul¹, E. Morini¹, E. Logan¹, E. Kirchner¹, J. Bolduc¹, A. Chekuri¹, M. Salani¹, B. Currall¹, R. Yadav², S. Erdin¹, M. Talkowski¹, D. Gao¹, S. Slauchaupt¹; ¹Massachusetts Gen. Hosp., Boston, MA, ²MGH/Harvard Med. Sch., Cambridge, MA

Abstract Body:

Familial Dysautonomia (FD) is a rare, recessive neurodegenerative disease caused by a splicing mutation in intron 20 of the Elongator complex protein 1 (*ELP1*, alias *IKBKAP*). This mutation leads to tissue-specific skipping of exon 20 and reduction of ELP1 protein, mainly in the central and peripheral nervous systems (CNS and PNS respectively). FD is a complex disorder affecting the development and survival of sensory and autonomic neurons. Little is known about why some neuronal tissues are more susceptible to ELP1 loss. We developed a phenotypic FD mouse model that recapitulates the tissue-specific *ELP1* splicing defect and the disease manifestations in patients including impaired proprioception and nociception. This model was generated by introducing the human *ELP1* gene with the FD mutation into a hypomorphic mouse expressing low amounts of endogenous *Elp1*. We performed transcriptomic profiling in dorsal root ganglion (DRG), trigeminal ganglion (TG), cortex, spinal cord, and medulla collected from the FD phenotypic mouse. We found that overall changes of gene expression were most prominent in medulla and lowest in the spinal cord. We discovered much greater sharing of differentially expressed genes (DEGs) across the TG and DRG tissues than expected by chance (191; $p = 1.6e-111$, hypergeometric test), and these genes are involved in pathways important for the regulation of synaptic signaling and transmission ($FDR < 0.1$). Such convergent transcriptomic dysfunction in the PNS was supported by gene co-expression analysis. We found that co-expressed genes strongly correlated with *ELP1* expression and were highly preserved in PNS tissues. We hypothesized unique cell types shared by PNS tissues might be the main contributor to the convergent PNS dysregulation in FD etiology. By comparing these DEGs in PNS to publicly available single-cell transcriptomic data from human and mouse DRG and TG, we found that the dysregulated PNS genes were significantly enriched for proprioceptor ($p=1e-3$), and nociceptor ($p=3.4e-3$) markers. For the first time, our data reveals the cell-type specific signatures underlying the loss of proprioceptors and nociceptors characteristic of FD. This study is the first to 1) identify tissue-specific DEGs whose expression correlates with ELP1 2) demonstrate a convergent gene dysregulation in FD mouse PNS tissues and 3) explore the cell-specific molecular pathways underlying FD neuronal degeneration. The identification of neuronal-specific gene networks underlying neuronal loss in FD provides a better understanding of disease etiology and identifies potential biomarkers for future clinical studies and new targetable pathways for therapy.

Complex Traits Posters - Thursday

PB1685. Transcriptome-wide analysis in Cameron County Hispanic Cohort (CCHC) provides novel insights into gene expression signatures of blood pressure

Authors:

X. Zhang¹, M. Graff¹, M. Yaser¹, J. Seo¹, S. Okello¹, H-H. Chen², W. Zhu², M. Lee³, J. B. McCormick³, S. P. Fisher-Hoch³, A. D. Gutierrez⁴, J. E. Below², K. E. North¹; ¹Dept. of Epidemiology, Gillings Sch. of Global Publ. Hlth., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³The Univ. of Texas Hlth.Sci. Ctr. at Houston, Sch. of Publ. Hlth., Brownsville, TX, ⁴Div. of Endocrinology, Diabetes and Metabolism, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract Body:

Background: Globally, the cardiovascular burden of hypertension is large - with ~1.4 billion adults with hypertension reported in 2010 and 8.5 million premature deaths in 2015. In particular, Hispanic/Latino Americans are disproportionately affected, with high prevalence but low rates of awareness, treatment, and control. While genome-wide association studies (GWAS) have identified > 1,400 loci associated with hypertension, functions underlying most of these loci remain unknown. To further realize the potential of GWAS, gene expression, which lies in between genetic variation and disease, can highlight pathways for targeted therapeutic intervention. However, to date, limited studies have sought molecular signatures of hypertension through gene expression, particularly in ancestrally diverse participants.

Methods: We leveraged whole blood RNA sequencing data in 806 Mexican Americans (age 51.9 ± 15.9 , 34.7% men) from the Cameron County Hispanic Cohort (CCHC) to identify differentially expressed genes. We used established protocols and alignment, yielding 15,694 protein coding genes after quality control. We adopted a two-step linear regression model to assess differentially expressed genes associated with systolic and diastolic blood pressure (S/DBP). In the first step, each outcome was adjusted for age and sex; the resulting residuals were rank normalized and then regressed on gene expression and 10 probabilistic estimation of expression residuals (PEER) factors.

Results: We identified 122 and 88 genes whose expression was significantly associated with SBP and DBP (FDR $p < 0.05$). Among our statistically significant results, we (1) replicated previous associations between gene expression and BP, including *CRIP1*, *MYADM*, *TIPARP*, *TSC22D3*, *LMNA*, *TPPP3*, *FOS*, and *TAGAP*, (2) identified possible causal genes underlying mapped BP GWAS, including *ATP2B1* and *GOSR2*, and (3) identified novel genes such as *KCNQ1* which has been shown to be associated with glomerular filtration rate and other cardiometabolic measures.

Conclusions: Our results demonstrated not only how transcriptome-wide analysis can bridge GWAS signals and hypertension, but also the importance of ancestrally diverse populations in the replication of previously published findings and in the identification of novel genes. Collectively, our findings fill the gaps in the genetics of hypertension and offer a more comprehensive picture of pathway-level alternations which ultimately could lead to actionable therapeutic targets with biological relevance.

Complex Traits Posters - Wednesday

PB1686*. Transcriptome-wide and Proteome-wide association study of Tourette's Syndrome

Authors:

S. Shekhar¹, P. JAIN², P. Paschou¹; ¹Purdue Univ., West Lafayette, IN, ²Purdue Univ., WEST LAFAYETTE, IN

Abstract Body:

Tourette's Syndrome (TS) is a neurodevelopmental disorder that is characterized by motor and phonic tics. A recent genome-wide association study (GWAS) identified *FLT3* gene as a genome-wide locus significantly linked to TS. However, determining the biological mechanism and pathways of GWAS signals remains difficult, limiting the understanding of TS. Importantly, the current iteration of GWAS does not identify *FLT3* as a significant locus and thus has an ambiguity regarding the role of *FLT3* in TS. To characterize the effect of genetic variation mediated gene expression in TS and to understand the biological underpinnings of the disease, we perform a global and unbiased transcriptome-wide and proteome-wide association study (TWAS and PWAS respectively) in the largest cohort of TS patient samples of general European ancestry consisting of 6133 TS cases and 13565 healthy controls using FUSION tool. Our TWAS analysis using an individual tissue-based prediction matrix of gene expression validates that the increased expression of *FLT3* in the dorsolateral prefrontal cortex (DLPFC) is associated with TS, as reported in a recent publication. Additionally, the analysis identifies eleven novel genes whose transcript expression is significantly associated with TS. Next, we performed TWAS analysis using a cross-tissue-based prediction matrix of gene expression and identified *FLT3* and *EP300* to be significantly associated with TS. As transcriptome expression sparingly correlates with proteome expression, we performed PWAS to complement TWAS analysis. Our PWAS analysis, based on protein expression from DLPFC tissue, identified three novel genes. For these genes, genetic variant-mediated change in protein expression is significantly associated with TS. In conclusion, results from our TWAS and PWAS analysis allow us to identify novel genes associated with TS and identify biological pathways that can be validated via biological experimentation to strengthen our analysis.

Complex Traits Posters - Thursday

PB1687. Transcriptome-wide association analyses in CD4 T-cells identify genes dysregulated in frequent cocaine users and associated with HIV latent reservoir.

Authors:

B. Quach¹, C. Willis¹, D. Hancock¹, D. Konkle-Parker², H. Bolivar³, C. Lahiri⁴, E. Topper⁵, M. Cohen⁶, S. Kassaye⁷, J. DeHovitz⁸, M. Kuniholm⁹, N. Archin¹⁰, P. Tien¹¹, K. Xu¹², B. Aouizerat¹³, E. Johnson¹; ¹RTI Intl., Research Triangle Park, NC, ²Univ. of Mississippi Med. Ctr., Jackson, MS, ³Univ. of Miami-ACRU, Miami, FL, ⁴Emory Univ., Atlanta, GA, ⁵Johns Hopkins Univ., Baltimore, MD, ⁶Cook County Hlth.System, Chicago, IL, ⁷Georgetown Univ., Washington, DC, ⁸Downstate Hlth.Sci. Univ., Brooklyn, NY, ⁹Univ. at Albany, Rensselaer, NY, ¹⁰Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ¹¹Univ. of California San Francisco, San Francisco, CA, ¹²Yale Univ., New Haven, CT, ¹³New York Univ., New York, NY

Abstract Body:

Background: Cocaine use is known to impact HIV treatment and progression and is common among people with HIV. Understanding and eliminating the HIV latent reservoir (HLR), the replication competent but repressed HIV provirus integrated into host cells' DNA, is key to an HIV cure. CD4 T-cells harbor the largest known HLR of human cell types. We examined the association of gene expression with cocaine use and HLR in CD4 T-cells and tested genes as mediators between the two.

Methods: To identify genes associated with cocaine use and HLR, we generated RNA-seq data and quantified HLR size using the intact proviral DNA assay from CD4 T-cells isolated from blood samples. Samples were contributed by HIV-infected, combination anti-retroviral therapy (cART) adherent women enrolled in the Women's Interagency HIV Study (WIHS) between 1993 and 2017 with self-reported cocaine use or no cocaine use in the 6 months prior to blood draw. We performed transcriptome-wide association studies (TWAS) of HLR (N=246) and cocaine use status (N=259) using negative binomial regression. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) biological pathway terms were tested for enrichment of HLR- or cocaine use-associated genes. Genes associated with both phenotypes were further evaluated as potential mediators of the effect of cocaine use on HLR size using causal mediation analysis.

Results and discussion: We discovered 101 and 1,608 genes associated with HLR and cocaine use (false discovery rate [FDR] <0.1), respectively, with 47 significant for both. HLR-associated genes were enriched in 10 GO terms (FDR <0.1) related to cell division, protein-DNA complex formation, and chromatin organization. Cocaine use-associated genes were enriched in 652 GO terms and 4 KEGG pathways (FDR <0.1) largely related to immune system processes (immune cell homeostasis, cytokine signaling, inflammatory response, immune cell activation and proliferation), inter- and intra-cell transport, chromatin modification, and nucleotide and amino acid metabolism. Of the 47 significant genes from both TWAS, 31 showed nominal evidence of being mediators between cocaine use and HLR ($p < 0.05$) with *EPSTII* mediation being statistically significant (Bonferroni $p < 0.05$). *EPSTII* is known to be activated by HIV-1 Tat protein in CD4 T-cells and plays a role in the translocation of NF- κ B to the nucleus. Our results provide evidence of widespread dysregulation of gene expression in CD4 T-cells from cocaine users and a link between some of these genes to HLR. Women are underrepresented in studies of HLR, and these findings contribute to better understanding HLR and cocaine use effects among women.

Complex Traits Posters - Wednesday

PB1688. Transcriptome-wide association study identifies novel candidate susceptibility genes for migraine.

Authors:

T. Meyers¹, J. Yin¹, A. Avins¹, T. Hoffmann², C. Schaefer¹, A. Pressman³, H. Choquet¹; ¹Kaiser Permanente Northern California Div. of Res., Oakland, CA, ²UCSF, San Francisco, CA, ³PRECISIONheor, Bethesda, MD

Abstract Body:

Purpose: Genome-wide association studies (GWAS) have identified 123 genetic susceptibility loci for migraine, however, how the majority of these loci impact migraine etiology is unknown. To identify novel genes associated with migraine and interpret the transcriptional effect of those genes, we conducted a transcriptome-wide association study (TWAS). **Methods:** We performed tissue-specific and multi-tissue TWAS analyses to assess associations between imputed gene expression from 53 tissues with migraine susceptibility using FUSION software. Meta-analyzed GWAS summary statistics from 26,052 migraine cases and 487,214 controls, all of European ancestry and from two cohorts (the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) and the UK Biobank) were used. We evaluated associations for genes after conditioning on variant-level effects from GWAS, and we tested for colocalization of GWAS migraine-associated loci and expression quantitative trait loci (eQTLs). **Results:** In tissue-specific analyses, 47 genes were associated with migraine after correcting for multiple testing and three of these genes (KTN1-AS1, VAT1, and PNKP) did not overlap any previously identified migraine-associated loci. Of these three genes, KTN1-AS1 and VAT1 were nominally associated with migraine ($P < 0.05$) after conditioning on the lead GWAS variant in the region. Moreover, cardiovascular tissues represented the highest proportion of the 47 Bonferroni-significant gene-tissue pairs (22 genes; 47%), followed by brain tissues (8 genes; 17%), and gastrointestinal tissues (4 genes; 9%). Furthermore, the posterior probability for colocalization was high (>0.9) in 20 of the 47 genes. Finally, our multi-tissue analyses identified 1,555 migraine-associated genes outside of known loci, including 37, 124, and 41 genes for the subsets of cardiovascular, brain, and gastrointestinal tissues, respectively. **Conclusions:** Our TWAS suggests novel genes for migraine and highlights the important contribution of brain, cardiovascular, and gastrointestinal tissues in migraine susceptibility etiology.

Complex Traits Posters - Thursday

PB1689. Transcriptomic analysis of severe obesity identifies novel genes in Hispanic/Latino populations with a high burden of disease.

Authors:

P. Sharma¹, H-H. Chen², W. Zhu³, H. Polikowsky³, P. McCormick⁴, K. Karastergiou⁵, S. Fried⁵, A. Justice⁶, M. Lee⁷, S. Puppala⁸, M. Graff⁹, Y. Mohammad¹⁰, J. McCormick⁷, S. Fisher-Hoch⁷, K. North⁹, J. Below¹¹; ¹Univ. of North Carolina, Chapel Hill, NC, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Vanderbilt Univ., Nashville, TN, ⁴Queen Mary Univ. of London, London, United Kingdom, ⁵Mount Sinai Med. Ctr., New York, NY, ⁶Geisinger, Danville, PA, ⁷The Univ. of Texas Hlth.Sci. Ctr., Houston, TX, ⁸Wake Forest Sch. of Med., Winston Salem, NC, ⁹Univ North Carolina, Chapel Hill, NC, ¹⁰Uni. of North Carolina, Chapel Hill, NC, ¹¹Vanderbilt Univ Med Ctr., Nashville, TN

Abstract Body:

The US prevalence of severe obesity (SevO, body mass index [BMI] ≥ 40 kg/m²) is increasing at an alarming rate, with highest prevalence among adult men (6.9%) and women (11.5%), and highest increase in Mexican-American men. While genome-wide association studies (GWAS) have identified >1000 loci associated with body mass index, the function of much of this variation is unknown. Gene expression measures can illuminate the link between genetic variation and disease highlighting pathways for targeted therapeutic intervention, but to-date, only a handful of studies has examined the role of gene expression to identify molecular signatures associated with SevO. To this end, we leveraged extant whole blood (WB) RNA sequencing (RNAseq) data in 75 SevO cases and 116 controls (with BMI = 18-25) collected from randomly selected Mexican Americans in the Cameron County Hispanic Cohort (CCHC) to identify patterns associated with SevO. We used established protocols and alignment, yielding 18,565 genes after quality control. We applied DESeq2 to assess DE associated with SevO, using a negative binomial regression model with a gene-specific dispersion parameter, adjusted for sex, age, T2D, hypertension, hypercholesterolemia, and 10 probabilistic estimation of expression residual (PEER) factors. After FDR correction, 124 genes were significantly DE, including top genes *C1RL*, *IL4R*, and *RGS16*. We identified several replications of the 124 genes in our ongoing GWAS of SevO (vs normal weight controls) in a large meta-analysis of 50,000 HL participants, including the CCHC. We additionally identified several replications of the 124 genes in subcutaneous adipose tissue (SAT) from 19 community volunteers, including for *RGS16*, *C1RL*, and *IL4R*. Collectively, these data demonstrate how transcriptomic studies may elevate understanding of SevO and inform efforts to reduce health disparities associated with SevO in HL populations.

Complex Traits Posters - Wednesday

PB1690. Transcriptomic analysis of whole blood in ancestrally diverse Alzheimer Disease cohorts implicates convergent immune and lipid processing molecular pathways

Authors:

T. Gu¹, N. Nelligan¹, D. Van Booven¹, P. L. Whitehead¹, K. L. Hamilton-Nelson¹, M. Contreras¹, J. J. Sanchez¹, S. Tejada¹, L. D. Adams¹, P. R. Mena¹, T. Starks², C. Silva³, M. R. Cornejo-Olivas⁴, M. Illanes-Manrique⁴, M. L. Cuccaro^{1,5}, J. M. Vance^{1,5}, W. S. Bush^{6,7}, G. S. Byrd², B. E. Feliciano-Astacio³, J. L. Haines^{6,7}, G. W. Beecham^{1,5}, M. A. Pericak-Vance^{1,5}, A. J. Griswold^{1,5}; ¹John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL, ²Maya Angelou Ctr. for Hlth.Equity, Wake Forest Univ., Winston-Salem, NC, ³Dept. of Internal Med., Univ. Central Del Caribe, Bayamón, PR, ⁴Neurogenetics Res. Ctr., Inst. Natl. de Ciencias Neurológicas, Lima, Peru, ⁵Dr. John T Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami, Miami, FL, ⁶Dept. of Population & Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, ⁷Cleveland Inst. for Computational Biology, Cleveland, OH

Abstract Body:

Background: Significant work has identified genetic variants conferring risk and protection for Alzheimer disease (AD), but study of downstream functional effects including modulation of gene regulation is lacking, particularly in individuals of diverse ancestries. To explore transcriptional changes between clinically diagnosed AD and cognitively intact age-matched controls, we analyzed RNA sequencing data from peripheral blood collected from Hispanic individuals of admixed genetic backgrounds and identified molecular pathways altered in these cohorts compared to existing data for European and African Americans. **Methods:** Total RNA was extracted from peripheral whole blood stored in PAXGene tubes from 47 Cubans (22 AD and 25 controls), 85 Peruvians (41 AD and 44 controls), and 168 Puerto Ricans (88 AD and 80 controls). PolyA mRNA was sequenced to 40 million paired end read per sample on the Illumina NovaSeq 6000. The bioinformatic pipeline included mapping to the human reference genome (GRCh38), gene quantifications against the GENCODE v35 annotation set, and differential expression calculated using DESeq2 with sex, age at blood draw, and count of *APOEε4* alleles as covariates. These data were then combined with data we previously generated from European (121 AD and 119 controls) and African American (115 AD and 119 controls) individuals (Griswold et al, 2018, PMID: 32597797) for functional categorization performed by gene set enrichment of gene ontology and KEGG pathways. **Results:** Across the new Hispanic cohorts, a total of 358 protein-coding genes ($FDR \leq 0.05$, Fold change ≥ 1.25) were differentially expressed with 238 down-regulated and 120 up-regulated in AD relative to controls. Few genes overlapped between the cohorts and across the European and African American data. However, pathway analysis revealed common pathways including up-regulation of genes involved in inflammation and RNA processing and down-regulation of genes involved in cellular detoxification and lipid transport, among others. These pathways converged with the existing European and African American data pathways as well. **Conclusion:** Identifying the genes and biological pathways involved in AD is critical in the effort to develop effective therapies. Our analysis reveals a signature of gene expression that implicates increased inflammation and decreased cellular detoxification based on gene expression analysis in genetically admixed individuals. Convergence of pathways across these and African American and European cohorts supports the idea of distinct genes but similar underlying pathological processes contributing to AD across individuals of diverse ancestries.

Complex Traits Posters - Thursday

PB1691. Transcriptomic and Epigenomic Consequences of Heterozygous Loss-of-function Mutations in *AKAP11*, a Bipolar Disorder Risk Gene, and their Effects on Responsiveness to Lithium

Authors:

N. Farhangdoost¹, A. Khayachi¹, Q. He², C. Jiao², V. Vuokila¹, B. Chaumette², P. Dion¹, G. Rouleau³; ¹McGill Univ., Montreal, QC, Canada, ²INSERM, Paris, France, ³Montreal Neurological Inst.-Hosp., Montreal, QC, Canada

Abstract Body:

Bipolar disorder (BD) is a polygenic psychiatric disorder that is characterized by recurrent episodes of mania and depression. Although BD has been shown to be a highly heritable disorder, a significant amount of the genetic risk is not yet detected. Bipolar Exome (BipEx) sequencing studies of 13,933 BD cases and 14,422 controls showed that the strongest case-control enrichment of rare protein-truncating variants is in the gene *AKAP11*. *AKAP11* is now considered a rare-variant large-effect risk gene in both BD and Schizophrenia (Palmer, D.S et al., 2022). This gene encodes A-Kinase Anchoring Protein 11 which acts as a scaffold protein to phosphorylate Glycogen Synthase Kinase 3 β (GSK3 β), the hypothesized target of lithium (Li) mainly prescribed to BD patients, through PKA. This complex has been shown to inhibit GSK3 β 's activity and affect its regulation by increasing its serine 9 phosphorylation levels. Therefore, the primary hypothesis of my study is that when *AKAP11*-Knockout (*AKAP11*-KO) induced pluripotent stem cell (iPSC)-derived neurons are treated with Li, Li would compensate for this deficit and could rescue some of the consequences of *AKAP11*'s loss of function by inhibiting GSK3B. Furthermore, since *AKAP11* has been shown to have direct protein-protein interactions with a histone lysine methyltransferase, SMYD2 (SET and MYND domain-containing protein 2), it is also my goal to investigate the chromatin dysregulation, more specifically DNA methylation and histone modifications, associated with loss of *AKAP11* in iPSC-derived neurons. To achieve these goals, I utilized a control induced pluripotent stem cell (iPSC) to knock out *AKAP11* using the RNP-CRISPR Cas9 genome editing technique. Three independent clones, each with a frameshift mutation in *AKAP11*, were obtained. Once differentiated to neurons, I performed RT-qPCR, RNA-seq, and Li treatment on the isogenic *AKAP11*-KOs. The differentially expressed genes in Li+ vs. Li- KOs, Li+ KOs vs. Li+ parental, and the Li- KOs vs. Li- parental line are currently being analyzed. Subsequently, chromatin immunoprecipitation sequencing of SMYD2-deposited histone marks, H3K36me2 and H3K4me3, and whole-genome bisulfite sequencing have are being performed on these isogenic clones and their parental, both before and after Li treatment, to determine genome-wide modifications of these histone marks along with DNA methylation level changes. This study will lead to gaining biologically meaningful information, useful for better understanding, diagnosis, and treatment of complex psychiatric disorders, such as BD and Schizophrenia, in which *AKAP11* has been shown to be a risk gene.

Complex Traits Posters - Wednesday

PB1692*. Trans-ethnic meta-analysis in a multi-ethnic population refines multiple sclerosis susceptibility loci and identifies novel locus

Authors:

A. Beecham¹, L. Gomez¹, S. Caillier², C. Manrique¹, P. Calabresi³, K. Fitzgerald³, N. Patsopoulos⁴, D. Woo⁵, S. Delgado¹, A. China⁶, L. Amezcua⁷, J. Oksenberg², J. McCauley¹; ¹Univ. of Miami, Miami, FL, ²Univ. of California San Francisco, San Francisco, CA, ³Johns Hopkins Univ., Baltimore, MD, ⁴Brigham & Women's Hosp., Harvard Med. Sch., Boston, MA, ⁵Univ. of Cincinnati, Cincinnati, OH, ⁶San Juan MS Ctr., Guaynabo, PR, ⁷Univ. of Southern California, Los Angeles, CA

Abstract Body:

In Northern European populations, 232 autosomal genetic variants have been identified for association with multiple sclerosis (MS) and explain 39% of the genetic component of MS. Despite the rising incidence of MS in African Americans and Hispanics, individuals from these communities remain underrepresented in genetic research. The genetic admixture and unique linkage disequilibrium structure inherent to these populations affords the opportunity to identify novel loci and increase power for fine-mapping. Our goal was to identify novel loci and fine-map previously identified susceptibility loci in African Americans (1625 MS cases, 1506 controls) and Hispanics (2046 MS cases, 2114 controls) ascertained from more than 10 sites across the United States. Samples were genotyped on a customized Illumina genome-wide array with targeted fine-mapping content and subsequently imputed with TopMed. MR-Mega was used for trans-ethnic meta-analysis after controlling for population substructure. We identified four loci outside of the Major Histocompatibility Complex with genome-wide significance in our combined sample of 7291 individuals; including three within 500KB of a previously identified susceptibility variant (*LCK* in 1p35.2, *CD58* in 1p13.1, and *CD86* in 3q13.33) and one which has not previously been reported (*TAF A5* in 22q13.32). The most strongly associated variant (OR = 1.63, p = 3.14E-08 and posterior probability of causality = 0.95) in *TAF A5* was intronic rs17176694; having the highest allele frequency in Europeans (MAF = 0.13 gnomAD) and Native Americans (MAF = 0.08 PAGE consortium) but rare in Africans (MAF = 0.02 gnomAD). Further investigation is ongoing to determine whether ancestral origin of allele plays a role. *TAF A5* is an important biological candidate for MS given that *TAF A* proteins are predominantly expressed in the brain and act as regulators of immune and nervous cells. An ancestral specific effect may indicate a role for *TAF A5* in the diverse manifestation of MS across populations. The *LCK* signal was fine-mapped to 5 variants in the 99% credible set, spanning 126,194 BP. The most strongly associated variant (OR = 1.85, p = 1.76E-13 and posterior probability of causality = 0.61) is missense mutation rs145088108 which is not present in Europeans (MAF = 0 gnomAD), rare in Africans (MAF = 0.01 gnomAD), and common in Americans (MAF = 0.1 gnomAD). This indicates that novel population-specific risk alleles are yet to be identified even within known susceptibility loci. These findings highlight the utility of a trans-ethnic approach to variant discovery and fine-mapping in a way that will ultimately facilitate treatment and prevention in diverse populations.

Complex Traits Posters - Thursday

PB1693. *TTN* truncating variants in cardiac-expressed exons show high penetrance for cardiomyopathy in carriers with atrial fibrillation

Authors:

K. Barrett¹, E. Cirulli¹, A. Bolze¹, C. Rowan², G. Elhanan³, J. Grzymalski³, W. Lee¹, N. Washington¹; ¹Helix, San Mateo, CA, ²Renown Inst. for Hlth.Innovation, Reno, NV, ³Ctr. for Genomic Med., Desert Res. Inst., Reno, NV

Abstract Body:

Background: Truncating variants in *TTN* (*TTN*tv) represent the largest known genetic cause of dilated cardiomyopathies (DCM). At the population level, even when limited to *TTN*tv in cardiac-specific exons (hiPSI *TTN*tv), penetrance estimates for DCM are low. Recent work shows that individuals harboring *TTN*tv have a high prevalence of other cardiac conditions aside from heart failure, in particular, atrial fibrillation (Afib).

Objectives: (i) Identify exons in *TTN* that are statistically associated with cardiac conditions in the overall population. (ii) Assess whether using additional significantly-associated phenotypes better stratifies cardiomyopathy (CM) risk across *TTN* carriers.

Methods: We leverage longitudinal EHR and exome sequencing data from two cohorts, the UK Biobank and the Healthy Nevada Project, to determine the extent of association between different sets of *TTN*tv and a spectrum of previously implicated cardiac diagnoses using a statistical power-based sliding window analysis technique (“Power Window”). We then assess the penetrance of these variants for CM in the context of the other statistically-confirmed cardio diagnoses.

Results: Controlling for CM and Afib, additional related cardio phenotypes retain only nominal association with *TTN*tv. An unbiased sliding window analysis of *TTN*tv across the locus confirms the association is specific to hiPSI exons for both CM and Afib, with no meaningful associations in lowPSI exons nor improvements from LOFTEE designations. We find 34% of hiPSI *TTN*tv carriers with early Afib have a CM diagnosis - a 5-fold increase in risk over non-carriers with early Afib and 47-fold increase over population controls.

Conclusion: CM and Afib are often coincident in hiPSI *TTN*tv carriers, which represent varying and progressive manifestations of structurally-based heart failure. We provide statistical support for a hiPSI variant interpretation model for *TTN*tv and evidence for the first population-level screening method with clinical utility for cardiomyopathies, especially in relation to an Afib finding.

Complex Traits Posters - Wednesday

PB1694. Understanding the effect of non-coding de novo mutations within craniofacial enhancers in trios with orofacial clefts

Authors:

S. Curtis¹, K. Paraiso², P. Kumari³, S. Chung¹, M. Bishop¹, K. Diaz Perez¹, H. Brand⁴, J. Murray⁵, E. Feingold⁶, T. Beaty⁷, S. Weinberg⁶, M. Marazita⁸, M. Epstein¹, D. Cutler¹, R. Cornell³, J. Cotney⁹, A. Visel², E. Leslie¹; ¹Emory Univ., Atlanta, GA, ²Lawrence Berkeley Natl Lab., Berkeley, CA, ³Univ of Washington, Seattle, WA, ⁴MGH, Wilmington, MA, ⁵U of Iowa, Iowa City, IA, ⁶Univ of Pittsburgh, Pittsburgh, PA, ⁷Johns Hopkins Univ, Sch PubHlth, Cockeyville, MD, ⁸Univ Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA, ⁹UConn Hlth., Farmington, CT

Abstract Body:

Orofacial clefts (OFCs) are a common birth defect, affecting 1 in 700 births, with a strong genetic basis and a high recurrence risk within families. Previous studies have associated many common variants with OFCs, primarily in non-coding elements, but the contribution of de novo mutations (DNMs) in these non-coding elements is not well understood. Therefore, we analyzed two datasets of 2,168 trios with OFCs and hypothesized that non-coding DNMs in craniofacial enhancers are associated with OFC risk. The first dataset included 66 DNMs in 1,409 trios with targeted sequencing of 13 known OFC-associated loci. We annotated variants to predicted craniofacial enhancers based on epigenetic marks in human embryos during craniofacial development. 17 DNMs (24.2%) were in a predicted enhancer region, and 2 DNMs were in the same enhancer (hs1617), which is more than expected by chance ($p = 0.0038$). These DNMs are predicted to lead to binding sites for PAX6 and ZBTB7A transcription factors (TFs) and to disrupt binding sites for STAT1 and STAT3. The enhancer is in a large topologically-associated domain containing HHAT, SERTAD4, and IRF6, all of which are involved in craniofacial development, suggesting a potential mechanism where mutations in an enhancer could disrupt expression of craniofacial genes. We next looked genome-wide at 49,620 DNMs called in a cohort of 759 trios with whole-genome sequencing data from the Gabriella Miller Kids First Research Program. Compared to DNMs called in the 1000 Genome Project, trios with an OFC had a higher frequency of DNMs in enhancer regions, with 234 DNMs being in strong craniofacial enhancers, which was more than expected by chance ($p < 2e-16$). Looking at individual craniofacial enhancers, we found 2 enhancers in OFC trios with an excess burden of DNMs: one upstream of SPIN1 ($p = 3.7e-6$), a chromatin reader that regulates the Wnt signaling pathway; and one within MPPED2 and downstream of ARL14EP ($p = 5.9e-5$), which is transcribed during facial development. When testing for enrichment of TF binding sites near DNMs genome-wide, binding sites for 295 TFs were enriched in trios with OFCs (after Bonferroni correction, $p < 6.8e-5$), and were more likely to be involved with embryo development ($p = 2.8e-3$) and regionalization ($p = 1.9e-2$). By contrast, the 106 TFs enriched in 1000G trios were more likely to be involved in hemostasis ($p = 1.3e-4$) and coagulation ($p = 7.4e-4$). Taken together, this suggests that non-coding DNMs could potentially affect developmental pathways leading to OFCs, and that investigating non-coding DNMs, both near known OFC-associated loci and genome-wide, adds to our understanding of the genetic mechanisms of OFC formation.

Complex Traits Posters - Thursday

PB1695. Understanding the etiology of Alzheimer's disease using genetically predicted gene expression and brain feature models

Authors:

E. Wu, Y. Liang, F. Nyasimi, H. Im; Univ. of Chicago, Chicago, IL

Abstract Body:

Treatment of Alzheimer's disease (AD) is one of the foremost public health concerns in the modern era. The genetic component of AD remains under-explained and is a promising direction of inquiry to further understand the complex mechanisms underlying the disease and develop effective treatment methods. In light of this, we applied cutting edge genetic association techniques, S-PrediXcan and S-BrainXcan, which leverage large-scale omics data to enhance the detection of AD-associated genes and brain features. S-PrediXcan is a transcriptome-wide association study method that uses publicly available GWAS summary statistics and imputes the unobserved transcriptome to associate with a trait of interest using 49 tissue-specific reference prediction models built from GTEx (v8) eQTL data. This analysis allows us to identify significant genes that are predicted to be differentially regulated under normal and AD conditions. We performed S-PrediXcan and subsequently S-MultiXcan - an aggregation method that meta-analyzes S-PrediXcan results across different tissues - on the most recent published AD GWAS and detected 166 genes associated with AD, with 7 genes not found in a previous GWAS or TWAS locus: *RP11-138I18.2*, *AC012370.3*, *SIGLEC11*, *LRP4*, *ZNF652*, *SUSD3*, and *PITPNA*. Additionally, we performed S-BrainXcan of AD, which is a similar method to S-PrediXcan that uses GWAS summary statistics and reference MRI imaging data from the UK Biobank to predict and identify significant brain features associated with disease in the form of image-derived phenotypes (IDPs). We identified changes in 25 T1 and diffusion MRI IDPs that were significantly associated with AD, which overall suggested that structural white matter, but not grey matter atrophy predisposes disease development. While these are exciting steps towards elucidating the causes of Alzheimer's disease, additional mechanistic studies will be needed to fully understand the biological signals captured by these results.

Complex Traits Posters - Wednesday

PB1696. Understanding the molecular basis of SARS-CoV-2 pathogenesis at a gene expression level using Applied Biosystems™ TaqMan™ Flexible Array Panels

Authors:

A. Gupta¹, R. Yang¹, K. Li¹, A. Chen², P. Kilgas³, K. Kirchmeier³; ¹Thermo Fisher Scientific, south san francisco, CA, ²Thermo Fisher, Carlsbad, CA, ³Thermo Fisher Scientific, Carlsbad, CA

Abstract Body:

The broad spectrum of clinical manifestations from SARS-COV-2 infection and observed risk factors for severe disease highlight the importance of understanding molecular mechanisms underlying SARS-CoV-2 associated disease pathogenesis. Research studies have identified a large number of host proteins that play roles in viral entry, innate immune response, or immune signalling during infection. The ability to interrogate subsets of these genes simultaneously within SARS-COV-2 infected samples is critical to understanding how their expression contribute to phenotypic variability of the disease caused by the pathogen. To bridge this gap, we obtained nasopharyngeal swabs from 30 SARS-CoV-2 infected individuals with asymptomatic or symptomatic disease, and 30 healthy controls and used flexible TaqMan array panels designed to interrogate SARS-CoV-2 associated genes for virus entry, cellular restriction factors as well as cytokines, chemokines, and growth factors shown to be dysregulated in SARS-CoV-2 infection. We observed that several genes were differentially modulated between asymptomatic and symptomatic SARS-CoV-2 infected samples, as well as between infected and healthy controls, shedding light on how host-pathogen interactions influence disease development. For instance, the gene expression of ACE2 was found to be significantly downregulated in symptomatic samples compared to asymptomatic samples from infected subjects, a finding supported by other studies and shown to be associated with increased inflammation. Furthermore, we observed sex-related differences in the expression levels of specific entry receptors and inflammatory mediators, which provides insights on the relationship between sex-bias and SARS-CoV-2 associated disease severity. Ultimately, investigation of such virus and host-associated factors in various cell and sample types can greatly increase our understanding of disease pathogenesis which may help in the development of novel therapeutics against the virus and the disease caused by it.

Complex Traits Posters - Thursday

PB1697. Undiagnosed complex cases with neurodevelopmental disorders (NDD) after exome sequencing-what's next?

Authors:

S. Trajkova¹, G. Battista Ferrero², E. Sukarova-Angelovska³, T. Jovanoski⁴, C. Giovenino⁵, A. Mussa⁶, S. Cardaropoli⁶, T. Pippucci⁷, P. Dimartino⁸, G. Spirito⁹, F. Musacchia¹⁰, S. Gustincich¹¹, M. Vecchi¹², S. De Rubeis¹³, J. Buxbaum¹³, S. Haghshenas¹⁴, H. McConkey¹⁵, J. Kerkhof¹⁶, B. Sadikovic¹⁷, A. Brusco¹⁸; ¹Univ. of Turin, Turin, Italy, ²Dept. of Clinical and Biological Sci., Univ. of Turin, Turin, Italy, ³Dept. of Endocrinology and Genetics, Univ. Clinic for Children's Diseases, Faculty of Med., Univ. "Sv. Kiril I Metodij", Skopje, Macedonia, The Former Yugoslav Republic of, ⁴Dept. of Plastic Surgery, Faculty of Med., Univ. "Sv. Kiril I Metodij", Skopje, Macedonia, The Former Yugoslav Republic of, ⁵Università degli Studi di Torino, Turin, Italy, ⁶Dept. of Publ. Hlth. and Pediatrics, Univ. of Turin, Turin, Italy, ⁷Sant'Orsola Malpighi Univ. Hosp., Bologna, Italy, ⁸Dept. of Med. and Surgical Sci., Univ. of Bologna, Bologna, Italy, ⁹Area of Neurobiology, SISSA, Trieste, Italy, ¹⁰Ctr. for Human Technologies, Istituto Italiano di Tecnologia, Genova, Italy, ¹¹Dept. of NeuroSci. and Brain Technologies, Istituto Italiano di Tecnologia, Genova, Italy, ¹²Istituto Italiano di Tecnologia, CMP3VdA, Aosta, Italy, ¹³Icahn Sch. of Med. at Mount Sinai, New York, NY, ¹⁴Molecular Diagnostics Program, Verspeeten Clinical Genome Ctr., London Hlth.Sci. Ctr., London, ON, Canada, ¹⁵Molecular Diagnostics Program, Verspeeten Clinical Genome Ctr., London Hlth.Sci. Ctr., London, ON, Canada, ¹⁶London Hlth.Sci. Ctr., London, ON, Canada, ¹⁷LHSC, London, ON, Canada, ¹⁸Univ. of Torino, Torino, Italy

Abstract Body:

Major advances in sequencing technologies have greatly increased diagnostic rate of neurodevelopmental disorders (NDDs). Still, a molecular diagnosis is reached in only 30%. Epigenetic marks as X-chromosome inactivation (XCI), and recently defined "episignatures" can both guide clinical re-evaluation, and further genetic analysis in unsolved NDD cases. We explored skewed XCI, episignature profiles (EpiSign™ classifier), whole genome sequencing (WGS) and integrated a-CGH/exome sequencing (ES) data in a deeply phenotyped cohort of unsolved NDD cases, negative at FRAXA, array-CGH and trio-ES [pathogenic(P)/likely pathogenic (LP) variants: 29% (173/589 cases); variant of uncertain significance (VoUS) :13% (78/589); novel genes:7% (39/589); unsolved:51% (299/589)]. We found a skewed XCI in 14/201 (7%) of mothers of undiagnosed males, and 12/98 (12%) female patients. We possibly solved 8/26 (31%) whose causative gene was previously missed. Four occurred in genes encoding chromatin remodeling proteins; in *KDM5C* and *BRWD3*, we confirmed the corresponding episignature [(i) *KDM5C*:D402N LP variant (female; 90:10 XCI); (ii) *BRWD3* c.1233-7_1233-3del LP(male; mother 90:0 XCI); (iii) *OTUD5*:P509L LP in recently described gene (two brothers, mother 100:0 XCI); (iv) *ZMYM3*: R441Q, a strong novel gene candidate (male; mother 100:0 XCI)]. We exploited episignatures to further interpret VoUS in chromatin remodeling genes (17/78; 22%). Twelve did not show a methylation profile compatible with the mutant gene, suggesting that variant was benign. Interestingly, a case with *ARID1B*:D1074N classified as VoUS, was not confirmed as Coffin-Siris syndrome 1 by EpiSign™ classifier, but confidently positive for Cornelia de Lange Syndrome (CdLS). So far, exome data did not reveal causative variant in CdLS genes. Deep phenotyping was especially useful in a boy with *ATRX*-like phenotype where subsequently was found an *ATRX* deletion of exons 3-4 (confirmed with episignature and mother 100:0 XCI), family with two different *de novo* variants in two similarly affected brothers (*TRIP12*:L1044Ffs*3, *FBN1*:A1728V), and case of Apert syndrome, negative to *FGFR2* point mutations, where was found an *Alu* insertion in intron 7 of the gene, using WGS. Combining array-CGH and ES, we solved a female case with an autosomal recessive intellectual disability associated with *HNMT*; the disease was due to a paternal missense p.Y147H and a 760kb

ASHG 2022 Annual Meeting Poster Abstracts

maternal deletion. Our work suggests that trio-ES analyses is greatly improved by deep phenotyping, XCI analysis to pinpoint X-linked genes, functional tests to verify VoUS in genes associated with epesignatures, as well as combining array-CGH and ES data.

Complex Traits Posters - Wednesday

PB1698. Universal clinical Parkinson's disease axes identify a major influence of neuroinflammation. Universal clinical Parkinson's disease axes identify a major influence of neuroinflammation.

Authors:

C. Sandor¹, S. Millin², A. Dahl², A-K. Schalkamp¹, M. Lawton³, L. Hubbard¹, N. Rahman¹, N. Williams¹, Y. Ben-Shlomo³, D. Grosset⁴, M. Hu², J. Marchini⁵, C. Webber⁶; ¹Univ. of Cardiff, Cardiff, United Kingdom, ²Oxford Univ., Oxford, United Kingdom, ³Univ. of Bristol, Bristol, United Kingdom, ⁴Queen Elizabeth Univ. Hosp., Glasgow, United Kingdom, ⁵Regeneron Genetics Ctr., Tarrytown, NY, ⁶Cardiff Univ., Cardiff, United Kingdom

Abstract Body:

There is large variation in the clinical presentations and progression between Parkinson's disease patients. The generation of deeply and longitudinally phenotyped patient cohorts has enormous potential to identify disease subtypes for prognosis and therapeutic targeting. Replicating across three large cohorts of Parkinson's disease patients with extensive clinical observational data repeatedly collected over many years, we developed a Bayesian multiple phenotypes mixed model incorporating the genetic relationships between individuals which was able to reduce a large number of diverse clinical measurements into a smaller number of continuous underlying factors ("phenotypic axes"). The most influential of three principal axes of Parkinson's disease patient phenotypic variation identified was specifically associated with the genetic risk of Alzheimer's disease and CSF A β 1-42 level. As observed previously for Alzheimer's disease genetic risk and in contrast to Parkinson's disease genetic risk, the loci influencing this primary axis were associated with microglia-expressed genes. These results suggest that neuroinflammation influences the development of more aggressive forms of Parkinson's disease. Finally, by integrating the individual Alzheimer's disease genetic risk into our model, we could significantly more accurately predict patient clinical progression as compared to using the full genotype. Our results propose that Parkinson's disease patients with a higher genetic risk for Alzheimer's disease are more likely to develop a more severe and rapidly progressing form of Parkinson's disease including, but not limited to, dementia.

Complex Traits Posters - Thursday

PB1699. Unraveling the genetic basis of autoimmune hypothyroidism

Authors:

M. Bujnis¹, R. Sterenborg^{2,3}, B. Åsvold^{4,5}, L. Brčić⁶, J. Freudenberg⁷, L. Fritsche⁸, M. Kanai^{9,10}, G. Leese¹¹, E. Marouli¹², A. Metspalu¹³, B. Mukherjee^{8,14}, Y. Okada¹⁵, C. N. A. Palmer¹⁶, M. R. Moksnes¹⁷, A. Papadopoulou¹⁸, V. B. Perica¹⁹, S. Srinivasan²⁰, E. Soto-Pedre²⁰, M. Siddiqui¹⁶, M. Teder-Laving²¹, L. Jorde²², M. Medici^{2,3}, A. Teumer²³; ¹Univ. of Utah, Sale Lake City, UT, ²Dept. of Endocrinology, Radboud Univ. Med. Ctr., Nijmegen, Netherlands, ³Dept. of Internal Med. and Academic Ctr. for Thyroid Diseases, Erasmus Univ. Med. Ctr., Rotterdam, Netherlands, ⁴Dept. of Publ. Hlth. and Nursing, Faculty of Med. and Hlth.Sci., K.G. Jebsen Ctr. for Genetic Epidemiology, NTNU, Norwegian Univ. of Sci. and Technology, Norway, Trondheim, Norway, ⁵Dept. of Endocrinology, St. Olavs Hosp., Trondheim Univ. Hosp., Trondheim, Norway, ⁶Dept. of Med. Biology, Sch. of Med., Univ. of Split, Split, Croatia, ⁷Regeneron, Tarrytown, NY, ⁸Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, ⁹Broad Inst. of MIT and Harvard, Cambridge, MA, ¹⁰Lab. for Statistical Analysis, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, ¹¹Ctr. for Pharmacogenetics and Pharmacogenomics, Ninewells Hosp. & Med. Sch., Univ. of Dundee, Dundee, United Kingdom, ¹²Barts and The London Sch. of Med. and Dentistry Queen Mary Univ. of London, London, United Kingdom, ¹³Universaity of Tartu, Inst. of Genomics, Tartu, Estonia, ¹⁴Dept. of Epidemiology, Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, ¹⁵Osaka Univ. Graduate Sch. of Med., Suita, Japan, ¹⁶Univ Dundee, Dundee, United Kingdom, ¹⁷Dept. of Publ. Hlth. and Nursing, Faculty of Med. and Hlth.Sci., K.G. Jebsen Ctr. for Genetic Epidemiology, NTNU, Norwegian Univ. of Sci. and Technology, Trondheim, Norway, ¹⁸Queen Mary Univ. of London, London, United Kingdom, ¹⁹Dept. of Med. Biology, Sch. of Med., Univ. of Split, Split, Croatia, ²⁰Div. of Population Hlth. & Genomics, Ninewells Hosp. & Med. Sch., Univ. of Dundee, Dundee, United Kingdom, ²¹Estonian Genome Ctr., Inst. of Genomics Univ. of Tartu, Tartu, Estonia, ²²Univ. of Utah, Salt Lake City, UT, ²³Univ. Med. Greifswald, Greifswald, Germany

Abstract Body:

Despite its high prevalence and high heritability, very little is known about the genetic basis of autoimmune hypothyroidism¹. For this reason, we have performed the largest GWAS meta-analysis of autoimmune hypothyroidism to date to gain more insight into the genetic basis of this disease. We obtained GWAS summary statistics from nine cohorts, including 48,149 cases and 863,954 controls. Cases were defined as individuals with an ICD9/10 code for autoimmune thyroiditis and unspecified hypothyroidism. Individuals without any thyroid disease were used as controls. Reference genomes were used to impute missing genotype data^{2,3}. Covariates adjusted for in our analysis include age, sex, population structure, and related individuals³. Using METAL, we performed the largest GWAS meta-analysis to date on hypothyroidism⁴. We identified 75 independent SNPs passing the GWAS significance threshold ($p=5 \times 10^{-8}$) after Bonferroni test correction to be associated with autoimmune hypothyroidism. Preliminary pathway analyses suggest involvement of known immune pathways as well as enrichment of associations in chromatin structure regulation. Identification of novel genetic determinants is a key step in advancing the understanding of this common disease. This knowledge could be applied to help predicting disease onset, personalized management, and treatment of autoimmune hypothyroidism.

1. Chiovato, L., et al. *Advances in Therapy*. 36, 47-58 (2019). 2. Auton, A. et al. *Nature*. 526, 68-74 (2015). 3. McCarthy, S. et al. *Nat. Genet.* 48, 1279-1283 (2016). 4. Willer, C., et al. *Bioinformatics* 26, 2190-2191 (2010).

Complex Traits Posters - Wednesday

PB1700*. Unsupervised learning revealed metabolic syndrome sub-groups with differing phenotypic and genotypic traits

Authors:

A. Lim^{1,2,3}, E. U. Lim¹, P-L. Chen^{4,5}, C. S. J. Fann^{1,2}; ¹Inst. of BioMed. Sci., Academia Sinica, Taipei, Taiwan, ²Taiwan Intl. Graduate Program in Molecular Med., Natl. Yang Ming Chiao Tung Univ. and Academia Sinica, Taipei, Taiwan, ³ASUS Intelligent Cloud Services (AICS), Taipei, Taiwan, ⁴Natl. Taiwan Univ. Hosp., Taipei, Taiwan, ⁵Natl. Taiwan Univ., Taipei, Taiwan

Abstract Body:

Metabolic syndrome (MetS) is a collection of cardiovascular risk factors of hypertension, hyperglycaemia, obesity and dyslipidaemia that overlap and highly heterogenous. Through unsupervised learning, five MetS subgroups were identified in UK Biobank which were C1: non-descriptive (n=33,707), C2: hypertensive (n=23,215), C3: obesity (n=30,089), C4: lipodystrophy (n=13,116) and C5: hyperglycaemia (n=3,869). Association of 21 clinical outcomes were compared through multivariate logistic regression. Overall combined CVD outcomes odds ratio of MetS was 5.517 (CI=5.207, 5.847). Some MetS clusters had higher CVD risks such as C1 (OR=6.765, CI=6.360, 7.192) and C5 (OR=9.486, CI=8.680, 10.360). Despite being undescriptive across all traits, C1 had higher risks for most clinical outcomes and highest CVD risk after adjustment for T2D. Intriguingly, the C2 overall had lower disease risk (combined CVD OR=3.783, CI=3.540, 4.043). GWAS of each MetS clusters with healthy control revealed different genetic variants, genes and biological pathways. *LPCAT2* was associated with all clusters and expression is specific to innate immune cells, implying an important underlying shared pathophysiology of chronic inflammation. BMI-associated genes *NUDT21* & *OGFOD1* were also associated with all clusters. C1 GWAS revealed novel findings such as *TRIM63* (rs35738294), *MYBPC3* (rs11039155, rs77509279), *MYLPP* (rs8054556) and *RAPSN* (rs77509279, rs4603265). *TRIM63* & *MYBPC3* are specifically upregulated in cardiac tissues while *MYLPP* & *RAPSN* in skeletal muscles. *MYLPP* was unique to only C1. C1, C3 & C4 were associated with genes highly expressed in brain tissues such as *CNIH2*, *TMEM151A*, *MT3* & *CIQTNF4*, highlighting the involvement of hypothalamus & pituitary in these clusters. C5 GWAS identified a novel regulatory SNP (rs142827301) which interacts with *ZPHHC6*, *SMC3*, *SHOC2*, *TCF7C2*, *VTI1A* and *BBIP1*; most are not known cardiometabolic genes. Other genes identified by C5 GWAS were known T2D genes such as *TCF7L2*, *IRS1*, *BBIP1* & *GIN1*. GWAS C2 revealed novel genetic variants in known cardiometabolic genes such as rs1260326 (*GCKR*), & rs633185 (*ARHGAP42*). When analysing the cluster-specific genes, C1 had 156 distinct genes while C2 had 18, C3 had 93, C4 had 123 and C5 had 16. C5 gene-set analysis was associated with pancreas while other clusters were associated with mostly lipoprotein and lipid homeostasis. MetS subgroups are semi-distinctively different in terms of phenotypic and genotypic traits, representing a key step towards precision medicine. GWAS of clinically relevant subgroups can reveal novel cardiometabolic genotypes which might be masked by heterogeneity of MetS.

Complex Traits Posters - Thursday

PB1701. Untargeted metabolomics on nutrient density and their genomic interaction among Mexican Americans

Authors:

S. Chung¹, C. Evans², C. F. Burant², D. Aguilar^{1,3}, E. L. Brown¹, C. L. Hanis¹, G. Jun¹; ¹Dept. of Epidemiology, Human Genetics & Environmental Sci. (EHGES), Sch. of Publ. Hlth., Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ²Michigan Regional Comprehensive Metabolomics Resource Core, Univ. of Michigan, Ann Arbor, MI, ³Div. of Cardiovascular Med., Univ. of Kentucky, Lexington, KY

Abstract Body:

About one-third of U.S. adults suffer from prediabetes and diabetes mellitus. Since diabetes can cause damage to multiple organs resulting in diverse complications, understanding how this condition develops in its early stages is essential for efficient preventive intervention. Although evidence suggests that interactions between genetics and diet contribute to worsening glycemia little is known about gene-diet interactions due to the sheer number of possible hypotheses, limited power, and absence of appropriate data. This study aims to identify genetic variants interacting with diet by utilizing untargeted metabolomics data coupled with genome-level genotyping and extensive phenotyping. The metabolomic data are much more *proximal* to both genes and the diet than typical endophenotypes. Among 616 participants in Starr County, Texas, an initial set of 486 were available with requisite data for preliminary analyses. Genetic variants were obtained from previously imputed GWAS data. Nutrient intake was assessed by food frequency questionnaires and transformed to nutrient densities based on standardized recipes and the amount of energy intake. Metabolomic data were standardized and log-transformed. 325 identified plasma metabolites were evaluated for associations with genetic variation, glycemic traits, and nutrient intake. Age, gender, and BMI were adjusted in all analyses. Eighty metabolites were significantly associated with genetic variants via Generalized linear Mixed Model Association Tests (GMMAT) (p-value < 5E-08). Twenty metabolites showed statistically significant associations with derived nutrient intake variables by linear regression models (p-value < 1.5E-04). These included 3-hydroxybutyric acid associated with protein intake, fasting plasma glucose, 2-hour plasma glucose, HbA1c, and glucose intolerance status. To investigate gene-nutrient density interactions, Mixed-model Association tests for Gene-Environment interactions (MAGEE) are used. We analyze interactions between genetic variations and major nutrient intake, and preliminary analyses identified several significant interactions with protein and carbohydrate intake. For example, more than ten significant genetic variants were associated with 3-hydroxybutyric acid in the gene-protein and gene-carbohydrate interaction analyses, and these were not founded on the GMMAT. Our preliminary results inform which metabolites are affected both by gene and specific nutrient intakes, suggesting that more personalized approaches could be developed for preventive interventions for prediabetes and diabetes in Mexican Americans.

Complex Traits Posters - Wednesday

PB1702*. Using genetics to uncouple highly correlated metabolic phenotypes and test their separate role in disease.

Authors:

S. Martin¹, J. Tyrrell¹, E. L. Thomas², M. J. Bown^{3,4}, A. R. Wood¹, R. N. Beaumont¹, L. C. Tsoi⁵, P. E. Stuart⁵, J. T. Elder^{5,6}, P. Law⁷, R. Houlston⁷, C. Kabrhel^{8,9}, N. Papadimitriou¹⁰, M. J. Gunter¹⁰, C. J. Bull¹¹, J. A. Bell¹¹, E. E. Vincent¹¹, N. Sattar¹², M. G. Dunlop^{13,14}, I. P. M. Tomlinson¹³, S. Lindstrom^{15,16}, INVENT consortium, J. D. Bell², T. M. Frayling¹, H. Yaghoobkar^{17,1,2}; ¹Univ. of Exeter, Exeter, United Kingdom, ²Univ. of Westminster, London, United Kingdom, ³Univ. of Leicester, Leicester, United Kingdom, ⁴NIHR Leicester BioMed. Res. Ctr., Leicester, United Kingdom, ⁵Univ. of Michigan, Ann Arbor, MI, ⁶Ann Arbor VA Hosp., Ann Arbor, MI, ⁷The Inst. of Cancer Res., London, United Kingdom, ⁸Havard Med. Sch., Boston, MA, ⁹Massachusetts Gen. Hosp., Boston, MA, ¹⁰Intl. Agency for Res. on Cancer, Lyon, France, ¹¹Univ. of Bristol, Bristol, United Kingdom, ¹²Univ. of Glasgow, Glasgow, United Kingdom, ¹³Univ. of Edinburgh, Edinburgh, United Kingdom, ¹⁴Western Gen. Hosp., Edinburgh, United Kingdom, ¹⁵Univ. of Washington, Seattle, WA, ¹⁶Fred Hutchinson Cancer Res. Ctr., Seattle, WA, ¹⁷Brunel Univ. London, London, United Kingdom

Abstract Body:

Genetic variants are usually associated with traits that mirror observational associations - for example, alleles associated with higher LDL-Cholesterol are also associated with higher cardiovascular disease risk. However, some variants do not follow expected patterns and can be useful to understand the separate roles of different risk factors in disease that would be difficult to test with other study designs. We used genetics to uncover whether the aspect of higher adiposity causing several obesity-driven diseases is the accompanying adverse metabolic profile or a secondary (e.g. mechanical) effect.

We selected two clusters of variants - a set that follow expected patterns (alleles associated with higher adiposity and an adverse metabolic profile - metabolically unfavourable adiposity (UFA)), and a set that follow unexpected patterns (alleles associated with higher adiposity and a favourable metabolic profile - favourable adiposity (FA)). We performed two-sample Mendelian randomisation (MR) with these variants and 37 obesity-associated diseases using genome-wide association study (GWAS) summary statistics from the largest available published GWAS meta-analyses and the latest release of FinnGen consisting of 309,154 individuals, allowing real-time updates with FinnGen releases.

Our MR analysis identified two groups of diseases. First, 8 conditions where the FA effect was protective of disease but the UFA effect was a risk factor, and so the metabolic component is the likely cause, including type 2 diabetes (FA: 0.12 OR [95% CI: 0.09, 0.17], P = 3E-34; UFA: 5.14 OR [4.08, 6.48], P = 1E-43), polycystic ovary syndrome (FA: 0.46 OR [0.21, 0.99], P = 0.046; UFA: 5.72 [3.16, 10.34], P = 8E-9) and colorectal cancer (FA: 0.67 OR [0.53, 0.84], P = 6E-4); UFA: 1.23 OR [1.02, 1.49], P = 0.034). Second, 10 diseases where both FA and UFA effects were risk factors for disease and so some non-metabolic (e.g. mechanical) aspect is likely causal, including venous thromboembolism (FA: 2.50 OR [1.89, 3.31], P = 2E-10; UFA: 1.55 OR [1.23, 1.96], P = 2E-4), deep vein thrombosis (FA: 2.83 OR [1.67, 4.80], P = 1E-4; UFA: 2.04 OR [1.26, 3.32], P = 0.004) and osteoarthritis (FA: 1.43 OR [1.04, 1.96], P = 0.027; UFA: 1.85 OR [1.41, 2.44], P = 1E-5).

We found evidence for similar contributions of metabolic and non-metabolic components for heart failure (FA: 0.95 OR [0.74, 1.22], P = 0.695; UFA: 2.22 OR [1.82, 2.72], P = 9E-15) and atrial fibrillation (FA: 0.91 OR [0.61, 1.37], P = 0.656; UFA: 1.89 OR [1.46, 2.43], P = 1E-6).

Our results represent clear examples of how using genetic variants with unexpected effects on related phenotypes can be used to understand the role of different risk factors in disease.

Complex Traits Posters - Thursday

PB1703. Using Independent Component Analysis to disentangle obesity heterogeneity discovers genetic drivers of clinical disease signatures and their variation with BMI status.

Authors:

M. Shuey¹, E. Farber-Eger², T. Lasko¹, Q. Wells³; ¹Vanderbilt Univ. Med. Ctr., Murfreesboro, TN, ²Vanderbilt Univ, Nashville, TN, ³Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Attempting to understand the heterogeneity of a condition such as obesity by assigning each patient to exactly one subgroup, depending on their pattern of clinical variables, neglects the fact that patients usually express the clinical signatures of multiple conditions simultaneously. Instead of this *hard clustering* approach, a *soft clustering* approach that disentangles multiple clinical signatures appearing in a population by allowing each patient to express any number of those signatures can more effectively identify disease variants, and potentially enabling the discovery of precise disease sources, prognoses, and optimal treatments.

We used Independent Component Analysis (ICA), a soft clustering approach that relies on the probabilistic independence of unrelated disease sources, to disentangle the clinical heterogeneity of obesity from data in electronic health records (EHRs). A set of 46 obesity-relevant clinical variables derived from demographics, laboratory measures, medications, and disease phenotype codes (Phecodes) was extracted, which we divided into those with high body mass index (BMI ≥ 30 kg/m², 16,030 patients), and low BMI (27,551 patients). Mutually independent clinical signatures were disentangled by applying ICA to the high BMI subgroup, and the records of the low BMI subgroup were projected into this space. We used the expression of each patient for each clinical signature as a phenotype for GWAS and performed gene-set and genetic correlation analyses to evaluate their genetic architecture conditioned on BMI state (high vs. low vs. unconditioned).

The analysis identified 46 clinical signatures across the spectrum of all disease, most of which were clinically recognizable, and eight (17%) of which included BMI as an important variable. Most signatures had expected genetic associations, but 11 (23.9%) signatures had associations that varied dramatically with BMI conditioning, suggesting that some genetic variants have different clinical effects on obese vs. non-obese individuals. For example, the results for an elevated liver enzyme signature demonstrated significant gene-set associations for *PNPLA3* ($p=4.1E-14$) and *SAMM50* ($p=5.0E-10$) genes in the high-BMI, but in the low-BMI cohort the most significant association was with *SLC39A12* ($p=0.8E-10$).

Our results support this data-driven approach to disentangle simultaneously-expressed clinical signatures of heterogenous disease. Some of the identified signatures demonstrate variable associations based on clinical state (e.g. BMI > 30 kg/m²) which may suggest the need for novel strategies for clinical care and intervention for these signatures based on weight.

Complex Traits Posters - Wednesday

PB1704*. Using Stimulation of B cells to uncover novel disease-associated QTLs

Authors:

D. Murphy, Y. Fu, J. Kelly, R. Pelikan, G. Wiley, J. Lynn, P. Gaffney; Oklahoma Med. Res. Fndn., Oklahoma City, OK

Abstract Body:

Although millions of disease-associated variants have been discovered using GWAS, the bulk of these occur in non-coding regions and their mechanistic link to disease is unclear. Further, separating causal from neutral variants within clusters of genetically linked loci remains challenging. To date, analyses of expression quantitative trait loci (eQTLs) in healthy tissues only account for a subset of GWAS hits, as there is limited overlap between the mechanisms driving variant discovery in the two approaches. Epigenome-based QTL phenotypes, such as chromatin accessibility (caQTL) or histone modification (hQTL), may capture distinct regulatory mechanisms that are not detected using eQTL analysis. Moreover, latent QTLs that more accurately regulate disease-specific mechanisms may be revealed by the comparison of healthy “resting” to stimulated tissue within the same set of samples. To this end, we performed ATAC-seq on resting and stimulated B cells purified from 81 genotyped healthy individuals of European ancestry using a general immune response “danger” stimulus consisting of BCR crosslinking, CD40 co-stimulation, and TLR7. We identified 42,746 differentially accessible (DA) chromatin regions when comparing pairs of resting and stimulated samples using DESeq2. We then used RASQUAL with normalized ATAC-seq read counts to identify 67,165 chromatin accessibility QTLs (caQTLs) either within the DA regions or 1kb upstream or downstream of these regions. Thousands of the discovered QTLs are stimulus-dependent (stQTLs), i.e., their effects are only revealed in the stimulated state. Relative to the total set of QTLs, stQTLs are enriched in enhancer-like sequences, transcription factor binding domains, autoimmune risk variants, and are distributed further from transcription start sites. Our approach also reveals different types of stQTLs, including variants with dominant and recessive effects, and stQTLs whose effects are lost under stimulation. Collectively, these results show the promise of stimulation for discovery of novel, conditionally regulated mechanisms associated with disease using a multi-omics approach beyond expression data alone. Incorporation of stimulus-response study designs into future QTL discovery efforts could identify QTLs that better explain risk for complex genetic diseases.

Complex Traits Posters - Thursday

PB1705. Using whole genome sequencing to understand the genetic architecture of cerebral palsy in a Canadian cohort.

Authors:

R. Wintle¹, D. L. Fehlings², M. Zarrei¹, W. Engchuan¹, B. Thiruvahindrapuram¹, E. J. Higginbotham¹, R. Thapa², T. Behlim³, J. Wei¹, P. S. Danthi¹, G. Pellicchia¹, T. Nalpathamkalam¹, S. Ghaffari¹, R. Patel¹, R. Shaath¹, B. Trost¹, S. Knights⁴, D. Samdup⁵, A. McCormick⁶, C. Hunt⁴, A. Kirton⁷, A. Kawamura², R. Mesterman⁸, J. W. Gorter⁸, N. Dlamini¹, D. Merico⁹, R. K. C. Yuen¹, M. Shevell³, D. J. Stavropoulos¹, M. Oskoui³, S. W. Scherer¹; ¹The Hosp. for Sick Children, Toronto, ON, Canada, ²Holland Bloorview Kids Rehabilitation Hosp., Toronto, ON, Canada, ³McGill Univ., Montreal, QC, Canada, ⁴Grandview Children's Ctr., Oshawa, ON, Canada, ⁵Queen's Univ., Kingston, ON, Canada, ⁶Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, ⁷Alberta Children's Hosp., Calgary, AB, Canada, ⁸McMaster Univ., Hamilton, ON, Canada, ⁹Deep Genomics, Inc., Toronto, ON, Canada

Abstract Body:

Cerebral palsy (CP) is a permanent, non-progressive brain injury leading to movement and posture disorders. It can involve comorbidities, including other neurodevelopmental disorders (NDD). Genomic changes have increasingly been found to play a role in etiology of CP. We applied Whole Genome Sequencing (WGS) to uncover a more complete landscape of genetic variation in both nuclear and mitochondrial DNA (CNVs, structural variations, SNVs, indels, tandem repeat (TR) variations, and mitochondrial DNA (mtDNA) variations) on a representative sample of 327 individuals with CP recruited from CP-NET (Ontario) and the Canadian CP Registry. The cohort included 308 complete trio and four quartet families. We also analyzed genomes from two paediatric control cohorts (100 quartets, 203 trios). These latter data were used for comparison of *de novo* rates, and for statistical analyses.

We found clinically relevant alterations (CNVs, SNVs, indels, SVs, and mtDNA variants) in 30% of cases. Roughly one-third of these (12% of the total) were classified as pathogenic/likely pathogenic. This includes pathogenic mtDNA variants in 1.5% of cases, 1q21.1 distal duplication, deletions of 17p12 and 2q13, and single-gene deletions of *NRXN1*, *ARID2*, and *DLG2*. *De novo* or rare-inherited protein truncating variants (PTVs) were found in *SPAST*, *TRIP12*, *ERLIN2*, and *MEF2C*, and *de novo* or rare-inherited missense variants in *COL4A1*, *PIK3R2*, *TUBA1A* and *GNAO1*. Two aneuploidies were detected (trisomy 21 and 47,XXX). The *de novo* rate for CNVs and PTVs was higher in CP as compared with controls (CNVs: 5.8% vs 2.7%, PTVs: 14.7% vs 11.2%). While the *de novo* variants identified in the CP cohort accounted for a portion of clinically relevant variants, all *de novo* variants in paediatric controls were assessed as not clinically relevant to CP or NDD.

We performed pathway analysis of over-transmission of large TRs from parent to child, and burden of *de novo* SNVs in both CP cases and paediatric controls. Other classes of variants were not similarly analyzed due to a lack of power. We did not find enrichment of SNVs or over-transmission of large TRs in the paediatric controls. However, in the CP cohort, we observed an enrichment of *de novo* PTVs and missense variants in neurological pathways associated with nervous system development and function. Missense variants were also enriched in genes with high expression in the brain. We also found an over-transmission of large TRs in genes related to brain function in individuals with CP. Our analysis of WGS in CP highlights the importance of searching for clinically relevant variants in all classes of genomic aberrations in this complex neuromotor condition.

Complex Traits Posters - Wednesday

PB1706. USP53 variant associated with psychosis in a consanguineous pedigree.

Authors:

J. V Pardo¹, A. Kanwal², S. Sheikh³, A. Iftikhar², F. Aslam⁴, S. Yasin², S. Naz²; ¹Univ. of Minnesota & MVAHCS, Minneapolis, MN, ²Univ. of the Punjab, Lahore, Pakistan, ³Hawkes Bay Hosp., Hastings, New Zealand, ⁴Univ. of the Punjab, Lahore. Pakistan, Lahore, Pakistan

Abstract Body:

Mendelian variants for common psychiatric disorders such as schizophrenia are rarely sought. Yet, such variants are the most direct link between disease and its pathophysiology are as well foundational to many medical fields. The paucity of understanding the underlying biology impedes translational progress. To find such rare variants where they are most likely to be identified, we recruited multiplex consanguineous families with apparent Mendelian recessive segregation and an extreme phenotype: early onset, severe, treatment-resistant psychosis without comorbidity. A Pakistani family involving first cousin parents had six children. All were healthy except for two sisters with the extreme psychosis phenotype. No intellectual disability or substance abuse was present. Whole-exome sequencing was used to determine the cause of the disorder by evaluating the DNA of the affected sisters and one parent. All rare homozygous variants (allele frequency < 0.01) absent as homozygous in normative databases were considered. The allele frequency was $(10)^{-5}$ in south Asians. Candidate variants were analyzed by computational tools and Sanger sequencing of DNA from all participants. A ubiquitin specific protease USP53 variant, p.(Cys228Arg), predicted in silico as damaging segregated with psychosis. The variant was not a polymorphism (absent in 400 ethnically-matched, indigenous controls). The affected amino acid was absolutely conserved among vertebrates. The functions of USP53 remain unclear, but it has been associated with cholestasis, hearing loss, cancer, neural tissue, and bone development. USP53 lacks deubiquitinase catalytic activity; its function must rely on protein-protein interactions. Symptoms of reported human USP53 variants center on cholestasis with or without hearing loss—both absent in the affected participants. Murine USP53 Cys228 when mutated to serine causes hearing loss—highlighting its significance in neural processing. USP53 is implicated in several pathways relevant to schizophrenia such as ubiquitin processing, folate metabolism, neurotransmission, and tight junction physiology. The variant is unlikely to be incidental or false positive: several studies using the same approaches and parameters have produced either none or single novel variants of interest per family for a variety of different disorders including patients with the psychosis phenotype. This work suggests this USP53 variant may be a potential cause of psychosis and motivates functional studies to define its pathogenicity.

Complex Traits Posters - Thursday

PB1707. Validation of a rapid TTR genotyping assay as a point-of-care tool for cardiac amyloidosis diagnosis in low-income settings

Authors:

C. Anyika¹, N. Gandotra², C. Scharfe², E. Miller², M. Murray², T. 54gene¹, D. Attiogbe Attipoe¹, A. Ene-Obong¹, C. O'Dushlaine¹, O. Popoola¹; ¹54gene, Lagos, Nigeria, ²Yale Univ., New Haven, CT

Abstract Body:

Transthyretin cardiac amyloidosis (ATTR-CA) is a form of heart failure, characterized by progressive accumulation of misfolded transthyretin (TTR) protein in tissues, predominantly the nerves and heart. The *TTR* gene variant *c.424G>A, p.V142I* (a.k.a. *V122I*) is hereditary, pathogenic, and predominant in patients of West-African origin. However, ATTR-CA is typically unrecognized in this population due to the limited information on the clinical relevance of the variant, as well as expensive diagnosis. In response to this unmet need, a low-cost PCR-based TTR assay was developed by the Scharfe Lab at Yale (<https://medicine.yale.edu/lab/scharfe/research/rapidgenotyping/>) to enable rapid genetic diagnosis of the *TTR* variant in at-risk patients.

Using this assay, we evaluated the expression of *TTR V142I* in 100 whole blood samples from the 54gene Biobank Lagos Nigeria. These samples were recruited from participants with clinically-defined cardiovascular disease, as part of an ongoing 54gene research recruitment. We identified the *TTR* wild-type successfully in all samples tested. Additionally, we identified the *TTR V142I* variant in 2 of the 100 samples tested. These findings were corroborated by whole-genome sequencing performed in-house, confirming that this rapid TTR assay is capable of successfully identifying both WT and variant *TTR* in the African population tested.

We also assessed a larger database of ~16,000 individuals with *V142I TTR* genotypes available from our in-house custom GENIISYS microarray. The average minor allele frequency in this set was 0.0196, consistent with earlier reports. These individuals spanned >200 unique self-reported ethnolinguistic groups, 42 of which were represented by at least 20 individuals. We identified significant variation in minor allele frequency of the *TTR* variant in this set, ranging from a maximum of 0.044 to a minimum of 0. For a subset of about half the individuals that had a cardiovascular disease diagnosis, we see some modest enrichment of the *TTR* minor allele frequency in a number of populations. Our resulting data set validates the use of this assay as an affordable point-of-care test for LMIC hospitals and laboratories in West Africa, to better diagnose ATTR-CA.

References:1. Daniel Jacobson, et al. 2016. The prevalence and distribution of the amyloidogenic transthyretin (*TTR*) *V122I* allele in Africa. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5023940/>
2. Nikolaos Papoutsidakis et al, 2020. Enrichment for Cases of African-American Patients with Pathogenic *TTR V142I* Variant in the TOPCAT Trial.

<https://www.medrxiv.org/content/10.1101/2020.10.14.20201046v1.full.pdf>

Complex Traits Posters - Wednesday

PB1708. Variant analysis of disease causing genes and risk factor genes in patients with Parkinson's Disease.

Authors:

L. Andriamboavonjy¹, C. Michaud², M. Labrecque¹, M. Panisset¹, S. Audet¹, M. Tétreault¹; ¹CRCHUM, Montréal, QC, Canada, ²IRCM, Montréal, QC, Canada

Abstract Body:

Parkinson's disease is a complex disorder in which environmental variables and genetic predisposition can lead to disease apparition and progression. To date, 14 highly penetrant disease-causing genes and more than a hundred risk factor genes have been described. The disease-causing genes only explain 30% of familial and 3 to 5 % of sporadic PD. On the other hand, genetic risk factors explain a larger proportion of patients, but cannot predict its apparition since other variables are required. Since most PD patients are without familial history of the disease, clinical genetic testing is not frequently performed. This lack of knowledge of the genetic pathogenicity of PD complicates its positive diagnosis as well as the distinction of atypical PD. Hence, studying the interaction of PD gene polymorphisms could help better understand the molecular pathways involved in the progression of the pathology.

For this study, 13 parkinsonian patients, who did not have access to genetic screening prior to their recruitment had their blood sampled. We conducted a whole-exome sequencing (WES) on the DNA extracted from peripheral blood mononuclear cells and extracted all the high-confidence single nucleotide variants (SNV) of PD-related genes following variant calling. Our approach led to the detection of six SNVs in disease-causing genes, four of which were in the coding region of *PINK1*, *DNAJC6* or *GBA*, and mostly of uncertain significance (VUS). All 13 patients possessed at least one of 54 SNVs identified in risk factor loci, 37 of which were coding variants causing non-synonymous changes, and eight were either a frameshift indel or a splicing site variant. In several cases, patients were carrying the exact same variant, which was the case of two *MAPT* variants. A gene clustering analysis identified an enrichment of mutations linked to pathways such as transcription regulation and maintenance of gastrointestinal epithelium which is particularly interesting due to the gut-brain axis theory of PD.

We were also able to identify 18 homozygous variants unknown to be PD-related in our population, which have an allele frequency inferior to 5% and a CADD score over 20, with all of them of uncertain significance. Those variants can potentially be PD-related and need further experiments for validation. While results are still preliminary, it will be interesting to perform a biological validation of the VUS or increase our cohort size to assess their implication in PD. Another main takeaway from our findings is the apparent clustering of variants that may affect transcription regulation; hence a transcriptomic analysis will be necessary to further our research in the genetic understanding of PD.

Complex Traits Posters - Thursday

PB1709. Variants in *GBA* and *MAPT* influence parkinson disease risk and age at onset in a colombian patient

Authors:

J. Satizabal Soto, D. Arturo-Terranova, L. Moreno Giraldo; Univ. del Valle, Cali, Colombia

Abstract Body:

Mutations in *GBA*, the gene encoding the lysosomal enzyme glucocerebrosidase, are among the most known genetic risk factors for the development of Parkinson disease and related synucleinopathies. Parkinson disease (PD) is a neurodegenerative disorder defined by the presence of motor symptoms and signs. A great deal is known about *GBA1* are causal for the rare autosomal storage disorder Gaucher disease. Over the past decades, significant progress has been made in understanding the genetics and cell biology of glucocerebrosidase. Both patients with Gaucher disease (GD) and heterozygous carriers are at increased risk of developing Parkinson disease, Age at Onset (AAO) and Dementia with Lewy Bodies, although our understanding of the mechanism for this association remains incomplete. The case of a 50 year old female patient with a clinical of generalized tremors in the extremities, with 1 year of evolution with rapid progression; imaging studies that show non-specific involuntional changes, was presented. It was decided to take the molecular study through whole exome sequencing, obtaining a heterozygous variant in the *GBA* gene (c.1448T>C p.L483P exon 10) with pathogenic clinical significance and a heterozygous variant in the *MAPT* gene (c.1270G>A p.A424T exon 5). The interaction network between both genes carried out in the GeneMania program showed a close relationship in molecular functions such as positive regulation of proteolysis, microtubule polymerization or depolymerization, regulation of cell growth, neuron projection extension and regulation of microtubule cytoskeleton organization. The L483P mutation is usually associated with GD type 2 or 3, even when presenting in a compound heterozygous state. It seems that severe *GBA* mutations, such as L483P, are associated with a higher risk of causing PD compared to milder mutations, such as the N370S. Previous studies in multiple genetically diverse cohorts of PD patients have found associations between AAO of PD and variants in multiple genes including *GBA* and *MAPT*. A recent study using polygenic score for PD risk found that individuals with earlier onset also have higher polygenic risk, indicating that some of the genetic variants associated with PD risk also affect AAO. Several groups have previously reported association between variants at the *MAPT* locus and AAO. To our knowledge, our study is the first to report an association between AAO and variants in *GBA* and *MAPT* in a Colombian Patient. Our data confirm the strong effect of *GBA* and *MAPT* on PD risk, more frequent cognitive impairment, and more rapid progression.

Complex Traits Posters - Wednesday

PB1710. Variant-to-gene mapping in human microglial cell models with clonal CRISPR validation implicates *RTFDC1* and *CASS4* at the Alzheimer's disease 'CASS4' locus.

Authors:

E. Burton¹, M. Argenziano², K. Cook³, S. Lu³, C. Su³, E. Manduchi¹, M. Leonard³, K. Hodge³, L-S. Wang¹, G. Schellenberg¹, M. Johnson³, J. Pippin³, A. Wells³, S. Anderson¹, C. Brown¹, S. Grant³, A. Chesi¹; ¹The Univ. of Pennsylvania, Philadelphia, PA, ²Univ. of South Florida, Tampa, FL, ³The Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Late-onset Alzheimer's disease (LOAD) is the most common neurodegenerative disease among the elderly population, affecting nearly 6 million Americans over the age of 65. Despite being the 6th leading cause of death in the US, there are still no effective therapies that can slow or halt disease progression. LOAD research has principally focused on characterizing neurons as the primary causal cell type due to their production of aggregating amyloid beta ($A\beta_{1-42}$), which is thought to contribute to neurodegeneration. In contrast, recent studies suggest that genetic mechanisms drive microglia, the resident macrophages of the central nervous system, to prolonged inflammation in LOAD brains in the presence of $A\beta_{1-42}$, exacerbating neurodegeneration. Indeed, recent GWAS of LOAD have identified multiple loci near genes related to microglial function, such as *TREM2* and *CRI*. However, GWAS does not have the sensitivity to identify causal variants or effector genes. In order to map interactions between GWAS-implicated variants and their putative effector genes in LOAD, we used a combination of ATAC-seq and high-resolution promoter-focused Capture-C in two human microglial cell models (both the immortalized HMC3 microglia-like cell line and iPSC-derived iMicroglia) to map interactions between GWAS-implicated variants and their putative effector genes. Our promoter-focused Capture-C method improves upon the relatively low resolution of available Hi-C data by utilizing a 4-cutter enzyme (DpnII) instead of the traditionally used 6-cutter (HindIII). This variant-to-gene mapping in both microglial cell models identified 67 putative effector genes (51 coding) across both cell types, with 14 observed in both models. We identified a novel proxy SNP, rs6024870 ($r^2 = 0.93$ to sentinel SNP rs6014724), at the 'CASS4' locus, which coincided with open chromatin and directly contacted the promoter of *RTFDC1*, a gene not previously implicated in LOAD. Deletion of the putative enhancer region harboring rs6024870 by lentiviral CRISPR-Cas9 in HMC3 reduced the expression of *RTFDC1* at both the mRNA and protein level ($p < 0.01$ and $p < 0.05$ for mRNA and protein, respectively). It was also noted that *CASS4* levels were modestly influenced by this CRISPR-mediated perturbation. Our results implicate *RTFDC1* as a novel effector gene at the LOAD 'CASS4' GWAS locus. Further efforts will characterize the phenotypic effect of this variant in microglial cell models, including on inflammation and phagocytic activity.

Complex Traits Posters - Thursday

PB1711. Variation and impact of polygenic hematological traits in monogenic sickle cell disease

Authors:

G. Lettre^{1,2}, T. Pincez¹, K. Lo², A-L. Pham Hung d'Alexandry d'Orengiani³, M. Garrett⁴, C. Brugnara⁵, A. Ashley-Koch⁴, M. Telen⁴, F. Galactéros³, P. Joly⁶, P. Bartolucci³; ¹Université de Montréal, Montréal, QC, Canada, ²Montreal Heart Inst., Montréal, QC, Canada, ³Hôpital Henri-Mondor, Paris, France, ⁴Duke Univ., Durham, NC, ⁵Children's Hosp. Boston, Boston, MA, ⁶CHU-Lyon, Lyon, France

Abstract Body:

Several complications observed in sickle cell disease (SCD) are influenced by variation in hematological traits (HT), such as fetal hemoglobin (HbF) level and neutrophil count. Previous large-scale genome-wide association studies carried out in largely healthy individuals have identified 1000s of variants associated with HT, which have then been used to develop multi-ancestry polygenic trait scores (PTS). Here, we tested if these PTS associate with HT in SCD patients and can improve statistical models associated with SCD-related complications. In 2056 SCD patients, we found that the PTS predicted less HT variance than in non-SCD African-ancestry individuals. This was particularly striking at the Duffy/*DARC* locus, where we observed an epistatic interaction between the SCD genotype and the Duffy null variant (rs2814778) that led to a two-fold weaker effect on neutrophil count. PTS for these routinely measured HT were not associated with complications in SCD. In contrast, we found that a simple PTS for HbF that includes only six variants explained a large fraction of the phenotypic variation (17.1-26.4%), associated with acute chest syndrome and stroke risk, and improved the predictive model for vaso-occlusive crises. Using Mendelian randomization, we found that increasing HbF by 4.8% reduces stroke risk by 36% ($P = 0.0008$). Taken together, our results highlight the importance of validating PTS in large diseased populations before proposing their implementation in the context of precision medicine initiatives.

Complex Traits Posters - Wednesday

PB1712. Whole exome sequencing identifies two heterozygous novel variants in CCL21 and ITGA2 as a cause of Hypertrophic cardiomyopathy

Authors:

F. Carlus¹, L. B.Sujatha¹, A. Ganesh Kumar², V. George³, S. Justin Carlus⁴; ¹Dept. of Zoology, Pachaiyappa's Coll., Chennai, India, ²Ctr. for Res. & Dev., Dept. of Microbiol., Hindustan Coll. of Arts and Sci., Chennai, India, ³Consultant Cardiologist, Gen. Hosp.,, Ernakulam, Kerala, India, ⁴Micro Hlth.Lab., Calicut, India

Abstract Body:

Familial hypertrophic cardiomyopathy (HCM) is the most common genetic cardiac disease. While sarcomeric gene mutations explain many HCM cases, the genetic basis of about half of HCM cases remains elusive. Here we aimed to identify the gene causing HCM in a non-consanguineous Indian family with affected family members. Proband's mother and his elder sister expired due to CAD/stroke. Methods: HCM Proband and his 4 siblings screened by the whole exome sequencing. Conservative analysis, protein structural and functional prediction were performed on the identified pathogenic variants. Results: After a careful analysis, two heterozygous variants were detected in the CCL21 and ITGA2 genes in 3 siblings and not detected in 1 case. Both the variants have not been reported in the 1000 genomes and in our internal databases and have a minor allele frequency of 0.002% in the gnomAD database. The in silico predictions of the variant are probably damaging by PolyPhen-2 (HumDiv) and damaging by SIFT, LRT and MutationTaster2. The reference codon is conserved across mammals. Both the variants are a common cause of coronary artery disease. To date, CCL21 and ITGA2 have not previously been associated with HCM; to the best of our knowledge, this is the first report of heterozygous CCL21 and ITGA2 mutations associated with HCM. This study gives a broader spectrum of mutations linked to HCM and an additional scientific basis for diagnosis.

Complex Traits Posters - Thursday

PB1713. Whole Exome Sequencing revealed spectrum of mutations associated with different myopathy in clinically suspected DMD patients of Bangladesh

Authors:

T. Eshaque¹, H. Akter^{2,3}, S. Sarker^{4,5}, M. Basiruzzaman⁴, M. Chowdhury^{6,7}, M. Rahaman¹, R. Mim¹, M. Morshed¹, A. Islam^{1,8}, G. Kundu⁹, S. Kanta¹⁰, R. Biswas¹¹, K. Uddin¹, M. Uddin^{12,8}; ¹Genetics and Genomic Med. Ctr., NeuroGen Hlth.care, Dhaka, Bangladesh, ²NeuroGen Hlth.care Ltd., Dhaka, Bangladesh, ³Dept. of Biochemistry and Molecular Biology, Univ. of Dhaka, Dhaka, Bangladesh, ⁴Dept. of Child Neurology, NeuroGen Hlth.care, Dhaka, Bangladesh, ⁵Dept. of paediatric neuroSci., Dhaka Shishu Hosp., Dhaka, Bangladesh, ⁶NeuroGen Hlth.care, Dhaka, Bangladesh, ⁷Natl. Inst. of NeuroSci.s, Dhaka, Bangladesh, ⁸Cellular Intelligence Lab, GenomeArc Inc., Toronto, ON, Canada, ⁹Dept. of Child Neurology, Bangabandhu Sheikh Mujib Med. Univ., Dhaka, Bangladesh, ¹⁰Dept. of Paediatric NeuroSci., Dhaka Shishu Hosp., Dhaka, Bangladesh, ¹¹Dept. of Paediatric Endocrinology Dhaka Shishu (Children) Hosp. Shere Bangla Nagar, Dhaka, Bangladesh, ¹²Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates

Abstract Body:

BackgroundDuchenne muscular dystrophy (DMD) is an inherited genetic disorder resulting progressive skeletal, respiratory and cardiac muscle weakness. DMD is caused by point mutations along with deletions and duplications in DMD gene located on the X chromosome. Clinical phenotypes of DMD often overlaps with other kind of muscle disorders. In this study, we aimed to identify several types of mutations associated with different myopathy in clinically DMD suspected patients of Bangladesh.

MethodWe have conducted multiplex PCR test for the deletion analysis of 30 DMD suspected patients. Out of 30, Whole Exome Sequencing (WES) was done for 8 negative patients to identify single nucleotide variants (SNVs), small insertion and deletion. We have used human genome build GRCH38/UCSC hg38 as reference. Variant classification analysis was conducted based on the American College of Medical Genetics (ACMG) guidelines.

ResultsThe cohort comprises 30 male patients. From multiplex PCR analysis, we have found pathogenic deletion in 56.6% (17) patients. And no clinically relevant variants were found in 43.3% (13) patients. The diagnostic yield for this test is 56.6%. Out of 13 negative cases, WES was performed for 8 patients. Out of 8, we have found 6 pathogenic variants in a range of 4 disease-associated genes such as *DMD* associated Duchenne muscular dystrophy, *COL6A3* associated Bethlem myopathy, *LAMA2* associated Muscular dystrophy and *DOK7* associated Myasthenic syndrome. Our analysis pipeline detected a pathogenic frameshift deletion c.2604delA (p.Val869Serfs*) and a stopgain mutation c.967C>T (p.Arg323Ter) in *DMD*; a frameshift insertion c.2137_2138insGTGT (p.Ser713Cysfs*) and a splice site mutation c.4389+1G>A in *COL6A3*, a stopgain mutation c.5476C>T (p.Arg1826Ter) in *LAMA2* and a frameshift deletion c.204delG (p.Ala70Profs*) in *DOK7*. We have also found a variant of uncertain significance (VOUS) c.3220C>G (p.Gln1074Glu) in *MYH7* gene associated with Laing distal myopathy. No clinically relevant variants were found in one patient. So, the diagnostic yield of WES from PCR negative patients is 75% (6/8).

ConclusionsDMD symptoms may create confusion when overlapped with other muscle related disorders and creates trouble in future management. Our results show the utility of using both multiplex PCR test and WES for precise genetic diagnosis and its integration into the diagnosis therapeutics and management of DMD suspected cases.

Key WordsWhole Exome Sequencing (WES); Duchenne muscular dystrophy (DMD); single nucleotide variants (SNVs); American College of Medical Genetics (ACMG)

Complex Traits Posters - Wednesday

PB1714. Whole exome sequencing suggests a role for rare genetic variation in the *NLRX1* gene in HIV disease progression in pediatric African populations of Botswana and Uganda.

Authors:

M. Amujal¹, J. Mukisa², T. Diphoko³, G. Sebetso³, S. Mwesigwa⁴, G. Mboowa¹, S. Kyobe¹, E. Katagirya⁵, M. Joloba⁶, D. Kateete⁶, M. Matshaba⁶, G. Mardon⁷, N. Hanchard⁸; ¹Makerere Univ., Kampala, Uganda, ²Makerere Univ. Coll. of Hlth.Sci., Kampala, Uganda, ³Baylor Coll. of Med., Houston, TX, ⁴MAKERERE Univ., Kampala, Uganda, ⁵Coll. of Hlth.Sci., Makerere Univ., Kampala, Uganda, ⁶MAKERERE Univ., KAMPALA, Uganda, ⁷Baylor Col Med, Houston, TX, ⁸NIH, Bethesda, MD

Abstract Body:

HIV-1 remains a significant public health concern in Africa, with an estimated 26 million people living with HIV-1 in the region (UNAIDS, 2019). Children perinatally infected with HIV exhibit extremes in their temporal progression to AIDS. Rapid Progressors (RPs) have at least one documented CD4+ T-cell decline (<15%) and/or AIDS-defining illness within 3 years of HIV infection. In contrast, Long-Term Non-Progressors (LTNPs) show no signs of AIDS ≥10 years after infection, even without therapy. Common genetic variation in HLA alleles; HLA-B*5701, HLA-B27 and the *CCR5-Δ32* deletion only explain a minority of the LTNP phenotype. In this study, we determine the role of rare genetic variation in HIV disease progression in African pediatric populations of Uganda and Botswana.

Approximately, 1,000 participants were recruited in HIV care centres in Uganda and Botswana; accordingly, these participants were stratified into RPs and LTNPs. To identify rare missense/nonsense variants with MAF ≤1% contributing to the LTNP phenotype, we performed WES on 388 LTNPs and 419 RPs and findings from this analysis were further validated in 400 WGS from Uganda and Botswana. Single variant enrichment analysis revealed a role for a rare missense variant (rs145985036) in the NOD-like receptor X1 (*NLRX1*) gene. The variant was reported to be either deleterious/damaging by 11 out of 14 different insilico prediction tools including SIFT, Polyphen2, LRT, MutationTaster, and MutationAssessor. Additionally, the variant site was highly conserved across different insilico conservation tools including GERP++, SiPhy, phastCons and phyloP. Genotype-Tissue Expression (GTEx) analysis showed significant expression of the *NLRX1* gene in the oesophagus-mucosa, minor salivary glands and vagina. Furthermore, analysis of the molecular structure of the *NLRX1* gene revealed the variant allele to occur at the central nucleotide-binding domain termed (NOD domain) responsible for the activation of the receptor for an immune response. Additionally, the *NLRX1* gene has been reported to facilitate HIV-1 replication by 1) sequestering STING thereby negatively regulating the interferon response and 2) enhancing oxidative phosphorylation and glycolysis during HIV -1 Infection of CD4 T cells to promote viral replication.

Findings from this study reveal the first novel/non-HLA gene directly implicated in Pediatric HIV disease progression unique to the African population. We, therefore, recommend studies to understand the functional mechanisms of this gene that would provide potential targets for HIV immunotherapeutics and vaccines.

Complex Traits Posters - Thursday

PB1715. Whole genome sequence analysis of long non-coding RNAs for plasma lipid traits

Authors:

Y. Wang¹, M. Selvaraj^{2,3,4}, P. Natarajan^{2,3,4}, G. Peloso¹, TOPMed Lipids Working Group; ¹Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, ²Cardiovascular Res. Ctr., Massachusetts Gen. Hosp., Boston, MA, ³Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, ⁴Dept. of Med., Harvard Med. Sch., Boston, MA

Abstract Body:

Background Elevated blood lipids are heritable risk factors and major modifiable cause of cardiovascular disease. While long non-coding RNAs (lncRNAs) have important regulatory functions for lipid metabolism in model systems, the relationship between genetic variation in lncRNAs and blood lipid levels in humans is not well understood. We now utilize large-scale whole genome sequencing (WGS) studies and new statistical methods for variant set tests to assess the association between lncRNAs across the genome and plasma lipid traits. **Methods** We analyzed 66,329 individuals with TOPMed freeze8 WGS data and lipid levels (LDL-C, HDL-C, TC and TG). We defined lncRNA testing units by integrating annotations from four different genome annotation projects: GENCODE (v38), FANTOM CAT(robust), NONCODE (v6), and lncRNAKB (v7). We aggregated rare (MAF < 1%) variants for each lncRNA based on the lncRNA genomic locations and conducted the rare variants aggregate test using the STAAR framework incorporating multiple functional annotations. We further performed conditional analyses adjusting for previously reported common variants that associated with lipids. Since there are overlapping regions between the lncRNAs, we estimated the effective number of aggregate-based tests (M_{eff}) for multiple testing correction. **Results** In total, we conducted RV aggregate tests in 166k lncRNA regions with 113,587 effective number of aggregate-based tests. We identified 40, 31, 30, and 30 genome-wide significant ($p < 0.05/113587 = 4.4e-07$) lncRNAs with LDL, TC, HDL, and TG, respectively, in 11 loci. After conditioning on known lipid-associated variants, 21, 16, 15, and 13 associations remained significant. Of the significant lncRNAs in the conditional analysis, 19, 15, 13, and 12 associations were near at least a known lipid mendelian gene, including ENSG00000233271.1 near *PCSK9* associated with LDL-C, NONHSAG026009.2 near *APOE* associated with TC, NONHSAG108446.1 near *CETP* associated with HDL-C, and NONHSAG009700.3 near *APOA5* associated with TG. The remaining associations were all in lipid GWAS regions, except ENSG00000260441.5, which is an antisense to *PLA2G15* that is associated with HDL-C. **Conclusions** We discovered several associations between lncRNAs and plasma lipid traits, which provide insights into potential lipid regulatory mechanisms of GWAS loci. We will further seek replications in UK Biobank WGS and investigate the effects of lncRNAs on gene expression.

Complex Traits Posters - Wednesday

PB1716. Whole Genome Sequencing Analyses of 87,652 Individuals Reveal Rare Variants in Promoter of *HMGAI* Associated with Height

Authors:

X. Li¹, Z. Li¹, J. A. Brody², K. Ferrier³, D. D. Burkardt⁴, N. L. Heard-Costa⁵, J. E. Manson⁶, J. I. Rotter⁷, J. Hirschhorn⁸, C-T. Liu⁹, L. A. Lange³, X. Lin¹, TOPMed Anthropometry-Adiposity Working Group; ¹Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ²Univ of Washington, Seattle, WA, ³Univ. of Colorado Denver, Denver, CO, ⁴ACMG/NIH-NHGRI, Bethesda, MD, ⁵Boston Univ Sch Med., Boston, MA, ⁶Brigham and Women's Hosp., Boston, MA, ⁷Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, ⁸Boston Children's Hosp., Boston, MA, ⁹Boston Univ. SPH, Boston, MA

Abstract Body:

Introduction

Height is heritable and provides insights into the genetic architecture of human traits. Thousands of common and low-frequency variants have been identified with height using GWAS. However, these common variants only explain a limited fraction of heritability. A recent study shows that rare variants (RVs) are a major source of the missing heritability of height. Large-scale whole-genome sequencing (WGS) studies, such as the multi-ethnic NHLBI Trans-Omics Precision Medicine (TOPMed) Program, enable the assessment of associations between height and rare variants across the genome, especially for the noncoding genome.

Hypothesis

Rare variant aggregations are associated with height.

Methods

We applied our newly developed STAARpipeline workflow to rare variant (MAF < 0.01) association analyses using 87,652 individuals from TOPMed Freeze 8 WGS data, including gene-centric analysis and non-gene-centric analysis using a variety of coding and noncoding masks. The gene-centric analysis provides five coding and eight noncoding functional categories. The non-gene-centric analysis includes sliding window analysis with fixed sizes and dynamic window analysis with data-adaptive sizes.

Results

In the gene-centric analysis of coding RVs, we identified 7 genome-wide significant associations at the Bonferroni-corrected level $5.00E-07$ ($=0.05/20,000/5$). After conditioning on known height-associated variants, the association between missense RVs in *ACAN* remained significant at the same level $5.00E-07$. In the gene-centric analysis of noncoding RVs, we identified 8 genome-wide significant associations at the Bonferroni-corrected level $3.12E-07$ ($=0.05/20,000/8$). The association of RVs in the promoter of *HMGAI* remained significant at the same level $3.12E-07$ in conditional analysis by adjusting for known height-associated variants. In 2-kb sliding window analysis, we identified 25 genome-wide significant associations at the Bonferroni-corrected level $1.88E-08$ ($=0.05/2.66E06$). After conditioning on known height-associated variants, the strengths of all associations reduced and 4 of these associations remained significant at level $1.00E-05$. Two of them are in the coding region of *ACAN*, and the other two are in the upstream region of *HMGAI*. The results of dynamic window analysis are similar to sliding window analysis. We identified 11 genome-wide significant associations, and 4 of these associations remained significant at level $1.00E-05$ in conditional analysis.

Summary

Two new RV associations, missense RVs in *ACAN* and RVs in the promoter of *HMGAI* with height, were identified using the TOPMed WGS Freeze 8 data through STAARpipeline.

Complex Traits Posters - Thursday

PB1717. Whole genome sequencing enables identification of a genetic susceptibility locus for idiopathic pulmonary fibrosis in the 16p subtelomere.

Authors:

L. Donoghue¹, A. Stockwell¹, M. Neighbors¹, R. Sheng¹, P. Wolters², L. Lancaster³, J. Kropski³, T. Blackwell³, B. Yaspan¹, M. McCarthy¹; ¹Genentech, South San Francisco, CA, ²Univ. of California San Francisco, San Francisco, CA, ³Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Idiopathic pulmonary fibrosis (IPF) is a severe lung disease characterized by progressive interstitial fibrosis and inflammation. Even with available treatments, the average survival following diagnosis is ~3-5 years, highlighting the significant need for greater understanding of disease etiology and pathogenesis. Genome-wide association studies (GWAS) from large, increasingly ancestrally-diverse cohorts have identified >20 common variants associated with IPF, including a *MUC5B* promoter variant with large effect, and rare variant analyses have implicated genes involved in telomere integrity. Here, we report the first genome-wide search for common and rare IPF risk variants fully based on whole genome sequencing (WGS). We replicated 13 previously-identified common variant signals plus rare variant burden associations in *TERT* and *RTEL1* with 1,638 clinically-defined IPF cases and 7,947 controls. Access to WGS data enabled us to detect a new common variant association in the chromosome 16 subtelomere (16p13.3, rs367849850: $P=2.7 \times 10^{-20}$, OR=2.1), a region poorly and inconsistently covered by previous array-based discovery. This association replicated in an independent GWAS from the Global Biobank Meta-analysis Initiative with biobank-defined IPF cases ($P=1.1 \times 10^{-3}$, OR 1.3). Through conditional analysis with these 16p.13.3 subtelomeric variants and a recently-reported association 70kb away near *NPRL3*, we found rs367849850 (an 18-base pair indel in pseudogene *IL9RP3*) accounted for both the subtelomeric and *NPRL3* associations. Furthermore, these subtelomeric IPF-risk alleles were associated with shorter leukocyte telomeres in UK Biobank participants ($P=4.4 \times 10^{-4}$, OR=1.02). Collectively our findings highlight the advantages of using WGS data and clinically-defined IPF cases to more comprehensively evaluate genetic variation in historically ambiguous regions such as the subtelomere and identify this large-effect association missed by previous, larger GWAS. Further investigation into these risk variants may implicate a new pathogenic mechanism for IPF involving subtelomeric sequence variation.

Complex Traits Posters - Wednesday

PB1718. Whole genome sequencing of a diverse Hispanic IBD population in the United States reveals differences in previously identified risk alleles.

Authors:

T. Haritunians¹, J. L. McCauley², E. Torres³, L. Gomez², D. Li¹, M. Daly^{4,5,6}, C. Stevens⁵, J. Leavitt⁷, O. M. Damas², M. A. Quintero², M. T. Abreu², D. P. B. McGovern¹, A. H. Beecham²; ¹Cedars-Sinai Med. Ctr., Los Angeles, CA, ²Univ. of Miami, Miami, FL, ³Univ. of Puerto Rico, San Juan, Puerto Rico, ⁴Massachusetts Gen. Hosp., Boston, MA, ⁵The Broad Inst., Cambridge, MA, ⁶Inst. of Molecular Med. Finland, Univ. of Helsinki, Helsinki, Finland, ⁷GastroHlth., Miami, FL

Abstract Body:

Background: Inflammatory bowel disease (IBD), Crohn's disease (CD) and ulcerative colitis (UC) are immune-mediated diseases causing chronic inflammation of the digestive tract, with increasing incidence in the Hispanic population. Nearly all ~240 IBD risk loci were identified through studies of Northern European ancestry populations. The genetic admixture of Hispanics provides a unique opportunity to examine the ancestral origins of risk and may further elucidate underlying IBD pathophysiology. We sought to test the relevance of common European IBD variants within our diverse Hispanic cohort. **Methods:** 2415 self-identified Hispanics (854 CD; 634 UC; 44 undetermined IBD; 883 controls) and 248 previously identified susceptibility variants passed stringent quality control of whole genome sequence data. Global ancestry was calculated (Admixture) and logistic mixed model association was performed (Regenie) after controlling for population substructure. **Results:** Geographical differences in ancestry were observed across recruitment sites, with Hispanics on the west coast demonstrating higher levels of Native American (NAM) ancestry than those on the east coast. We observed one-sided replication ($p < 0.05$) of 36 of 96 CD risk variants (37.5%; expected 95% CI: 40-47%), 18 of 58 UC risk variants (31%; expected 95% CI: 31-50%), and 22 of 94 general IBD risk variants (23%; expected 95% CI: 34-41%). We found less replication than statistical power suggested for CD and IBD, indicating potential locus heterogeneity or differing linkage disequilibrium patterns between populations. We observed the strongest association ($p < 5 \times 10^{-4}$) for well-established loci including *NOD2* and *IL23R* for CD, as well as *MST1* and *EPHB4* for UC. We identified several susceptibility variants which conferred risk in Europeans but demonstrated protection in Hispanics, including *LRRK2* for susceptibility to CD. Further analysis revealed an interaction with NAM ancestry, where the *LRRK2* susceptibility variant conferred risk of CD in Hispanics with a low proportion of NAM ancestry but protection in Hispanics with a high proportion of NAM ancestry. **Conclusions:** Hispanics represent a genetically heterogeneous group often overlooked in IBD studies. We replicated association in our Hispanic population of key IBD loci originally identified in predominantly European cohorts and identified evidence for global ancestry risk modification at select loci. Work is ongoing to further assess local ancestry risk modification. Identifying differences in genetic effects between populations is crucial to facilitate individualized treatment and prevention across diverse patient populations.

Complex Traits Posters - Thursday

PB1719. Whole genome sequencing study in multiple sclerosis identifies novel CTBP2 association with change in brain lesion burden

Authors:

T. Bhangale¹, C. Yang², N. Sadhu², E. Fisher², E. A. Sartor², N. Creps¹, H. McLaughlin², C. Singh², K. C. Fitzgerald³, F. Briggs⁴, D. Devon⁵, M. Comabella⁶, H. Wiendl⁷, E. M. Mowry³, P. Calabresi³, A. Gafson², M. I. McCarthy¹, A. Herman¹, H. Runz², S. Belachew², S. John², X. Jia⁸, P. Bronson²; ¹Genentech, South San Francisco, CA, ²Biogen, Cambridge, MA, ³Dept.s of Neurology and Epidemiology, Johns Hopkins Univ., Baltimore, MD, ⁴Case Western Reserve Univ. Sch. of Med., Cleveland, OH, ⁵Mellen Ctr. for Multiple Sclerosis, Cleveland Clinic, Cleveland, OH, ⁶Dept. of Neurology and Neuroimmunology, MS Ctr. of Catalonia, Vall d'Hebron Univ. Hosp., Barcelona, Spain, ⁷Dept. of Neurology, Univ. of Münster, Münster, Germany, ⁸Genentech, San Francisco, CA

Abstract Body:

Introduction: Over 200 GWAS loci are associated with Multiple Sclerosis (MS) risk. These loci are largely immunological, corroborating the success of immunomodulatory therapies. Here, we performed whole-genome sequencing (WGS) in MS to interrogate mechanisms driving MS age-at-onset (AAO) and rate of change in total T2-weighted hyperintense lesion volume (T2LV), an MRI measure of brain lesion burden.

Methods: MS PATHS (Partners Advancing Technology and Health Solutions) is a longitudinal study collecting clinical and MRI data and biospecimens in >20,000 persons with MS (pwMS). We performed WGS at 30X in MS PATHS subjects with genetic consent (N=7,938), followed by QC and joint genotyping (BWA-GATK). First, we performed a GWAS for AAO in 6,210 pwMS of European ancestry. We then performed a GWAS for T2LV change (T2LVC) in 2,657 pwMS with 1+ year of MRI. T2LVC was computed as annualized percent change in T2LV from first to last brain MRI (median 2.6 years between MRIs). Both continuous outcomes were rank-based inverse-normal transformed. Covariates were sex, genetic ancestry, and for T2LVC, age. We tested significant loci for shared causal variants (colocalization) with complex traits and UK Biobank phenotypes.

Results: We called >114.5 million variants across 7,938 subjects. We did not detect genome-wide test inflation ($\lambda < 1.02$). Our AAO GWAS replicated a known association with *HLA-DRB1*15:01* ($\beta = -0.10$, $P = 3 \times 10^{-6}$), and identified two novel associations: an intronic variant at *HLA-DRA* (rs3129882*A, $\beta = 0.11$, s.e. = 0.02, $P = 2.6 \times 10^{-10}$), and an intronic variant at *EYA2* (rs3091826*G, $\beta = -0.28$, s.e. = 0.05, $P = 2.7 \times 10^{-8}$). The T2LVC GWAS identified a novel association in a *CTBP2* intron (rs4962725*C, $\beta = 0.16$, s.e. = 0.03, $P = 7.8 \times 10^{-9}$), which interacts with a *CTBP2* promoter. This locus shared a causal variant with a body mass index (BMI) locus (PP4 = 0.98).

Discussion: This is the largest WGS-based GWAS of AAO and T2LVC in MS. We replicated the known association between younger AAO and *HLA-DRB1*15:01*. We identified novel AAO loci at *HLA-DRA*, the sole alpha chain for *HLA-DRB1* (together HLA-DRA and HLA-DRB1 form a heterodimer); and at *EYA2* (a transcriptional coactivator and phosphatase). We also identified a novel locus for T2LVC at the corepressor *CTBP2*. Interestingly, this association signal colocalized with BMI and *CTBP2* plays a role in adipocyte differentiation. BMI is a known risk factor for MS, and for accelerated atrophy in MS. *CTBP2* also harbors known associations with brain cortical thickness and surface area. Future directions include replicating our novel loci, and expanding our GWAS to test longitudinal brain atrophy and clinical disability progression traits.

Complex Traits Posters - Wednesday

PB1720. Whole-exome sequencing in five families with specific language impairment (SLI) suggests novel SLI genes and confirms previous findings.

Authors:

M. Raza, E. Andres, K. Earnest, C. Zhong, M. Rice; The Univ. of Kansas, Lawrence, KS

Abstract Body:

While most children acquire language with no formal instruction, about 7% of the population have specific language impairment (SLI), causing them to struggle with the language despite average or above average nonverbal intelligence (NV-IQ) and no other behavioral diagnoses or hearing loss. Twins studies provided high heritability estimates, and family studies suggest SLI is inherited in families. Genetic studies have begun to identify gene loci through next-generation sequencing, but the genetic basis of SLI remains unknown. In the current study, an age-appropriate standardized omnibus language measure was used to categorically define the SLI phenotype. We performed whole-exome sequencing (WES) in 23 individuals from five families (total $N = 48$) and prioritized 96 rare exonic variants. Co-segregation analysis revealed 31 variants of further interest. Upon Sanger sequencing of these 31 variants in additional unrelated probands with SLI ($N = 160$), seven probands carried seven unique nonsynonymous variants. Two of these variants were present in the previously suggested genes (*UBE4B* and *ATP2C2*). *UBE4B* is located within the chromosome 1p36 region previously linked to reading disorder and SLI, and we previously reported a suggestive linkage to SLI in two of the five families included in this report. In an unrelated SLI proband, we identified a homozygous variant on *ATP2C2* (rs78887288, MAF = 0.09). This variant was previously reported as a heterozygous variant in another family utilized in the current study. The remaining five variants were identified in novel genes: *YLPM1*, *PCBP4* (observed in three probands), with three of these variants identified in one proband: *PRDX5*, *AKAP9*, and *TRIM28*. These results support the utility of family-based investigation of SLI, confirming previous findings and suggesting novel genes. Further study of these proposed genes is needed to understand their role in language acquisition.

Complex Traits Posters - Thursday

PB1721. Whole-genome sequencing analysis of fructosamine and glycated albumin in Black and White participants in the Atherosclerosis Risk in Communities (ARIC) Study.

Authors:

S. Venkataraghavan¹, D. Ray^{1,2}, E. Selvin^{1,3}, J. Coresh^{1,3}, A. Kottgen^{1,4}, J. S. Pankow⁵, E. Boerwinkle⁶, B. Yu⁷; ¹Dept. of Epidemiology, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ²Dept. of Biostatistics, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ³Welch Ctr. for Prevention, Epidemiology, & Clinical Res., Johns Hopkins Univ., Baltimore, MD, ⁴Inst. of Genetic Epidemiology, Faculty of Med. and Med. Ctr., Univ. of Freiburg, Freiburg, Germany, ⁵Div. of Epidemiology and Community Hlth., Sch. of Publ. Hlth., Univ. of Minnesota, Minneapolis, MN, ⁶The UTHlth.Sch. of Publ. Hlth., Houston, TX, ⁷Dept. of Epidemiology, Human Genetics and Environmental Sci., UTHlth.Sch. of Publ. Hlth., Houston, TX

Abstract Body:

Genetic studies of serum glyceamic biomarkers, fructosamine and glycated albumin, can improve our understanding of the genetic architecture of type 2 diabetes and its complications. GWAS on both traits conducted till date have examined associations of common variants with MAF>5%. We sought to explore the associations of variants (SNVs and indels) across the entire MAF spectrum with fructosamine and glycated albumin using whole genome sequencing data on 1456 Black and 5429 White middle-aged participants (mean age 57) without diabetes from Jackson, MS; Forsyth County, NC; Minneapolis, MN and Washington County, MD in the population-based ARIC cohort. We performed single-variant analyses of log transformed glyceamic biomarkers- fructosamine and glycated albumin- for variants with MAF>1%. At a whole-genome significance threshold of 10^{-8} , we detected locus chr19q13.33 near *RCN3*, where the lead variant effect allele T was associated with a 1% decrease in fructosamine levels (rs10419198 [T/C], $p = 2.7 \times 10^{-9}$) replicating previous findings from array-based GWAS. We also identified variants found exclusively in Blacks at a low-frequency (MAF 1-5%) associated with one or both traits at chr7p14.1 (rs151030849 [C/T], fructosamine $p = 7.2 \times 10^{-16}$, glycated albumin $p = 10^{-10}$) and chr17p11.2 (rs566083925 [C/CG], glycated albumin $p = 6.4 \times 10^{-9}$ each leading to 13-15% increase in associated biomarker levels. For rare variants (MAF<1%), we performed gene-centric tests of the log-transformed biomarkers using SMMAT-E, grouping putatively deleterious variants determined using WGSa annotater from coding and regulatory segments. At a Bonferroni-corrected significance threshold of 2.7×10^{-6} , we implicated 23 genes including a novel glyceamic gene, *HIST1H2AA*, which was associated with fructosamine ($p = 5.7 \times 10^{-9}$) and glycated albumin ($p = 1.9 \times 10^{-10}$) in Blacks and showed suggestive evidence of association with fructosamine ($p = 2.6 \times 10^{-5}$) in Whites. We found associations of fructosamine with genes enriched in fatty acid metabolism pathway, *ACOT2* ($p = 1.8 \times 10^{-6}$) and *UGDH* ($p = 1.3 \times 10^{-7}$), driven by frameshift mutations and associations of glycated albumin with *SLIT2* ($p = 9.1 \times 10^{-7}$) and *ROBO3* ($p = 2.0 \times 10^{-7}$) by missense variants increasing glycated albumin levels by 32-37%. *SLIT2* and *ROBO3* were enriched in axon-guidance pathway linked to diabetic neuropathy. Our study identified similarities and differences in genetics of biomarker levels across ancestries, novel susceptibility regions for glyceamia including genes in a well-established type 2 diabetes pathway and a potential genetic basis for increased glycated albumin levels among patients with diabetic neuropathy.

Complex Traits Posters - Wednesday

PB1722. X Chromosome Wide Association Study on Latin American cohort reveals new potential loci associated with Parkinson's Disease

Authors:

T. Peixoto Leal¹, R. C. Shilpa¹, M. Inca-Martinez¹, M. Gouveia², J. French-Kwawu³, V. Borda³, E. Mason¹, D. Loesch³, A. Horimoto⁴, E. I. Sarihan¹, M. Cornejo-Olivas⁵, L. Torres⁶, P. Mazzetti⁵, C. Cosentino⁶, E. Sarapura-Castro⁵, A. Rivera-Valdivia⁵, A. Medina⁷, E. Dieguez⁸, V. Raggio⁸, A. Lescano⁹, V. Tumas¹⁰, V. Borges¹¹, H. B. Ferraz¹¹, C. R. Rieder¹², A. Schumacher-Schuh¹³, B. L. Santos-Lobato¹⁴, C. Velez-Pardo¹⁵, M. Jimenez-Del-Rio¹⁵, F. Lopera¹⁵, S. Moreno¹⁵, P. Chana-Cuevas¹⁶, W. Fernandez¹⁷, G. Arboleda¹⁸, H. Arboleda¹⁷, C. E. Arboleda-Bustos¹⁷, D. Yearout⁴, C. Zabetian¹⁹, T. Thornton⁴, T. O'Connor³, I. F. Mata¹, International Parkinson Disease Genomics Consortium (IPDGC), Latin American Research Consortium on the Genetics of Parkinson's Disease (LARGE-PD); ¹Cleveland Clinic, Cleveland, OH, ²NIH, Bethesda, MD, ³Univ. of Maryland Sch. of Med., Baltimore, MD, ⁴Univ. of Washington, Seattle, WA, ⁵Inst. Natl. de Ciencias Neurológicas, Lima, Peru, ⁶Univ. Natl. Mayor de San Marcos, Lima, Peru, ⁷Univ. Natl. del Altiplano, Puno, Peru, ⁸Univ. de la República, Montevideo, Uruguay, ⁹Univ. de la República, Montevideo, Uruguay, ¹⁰Univ. e de São Paulo, Ribeirão Preto, Brazil, ¹¹Univ. e Federal de São Paulo, São Paulo, Brazil, ¹²Univ. e Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil, ¹³Hosp. de Clínicas de Porto Alegre, Porto Alegre, Brazil, ¹⁴Univ. e Federal do Pará, Belém, Brazil, ¹⁵Univ. de Antioquia, Medellín, Colombia, ¹⁶Univ. de Santiago de Chile, Santiago, Chile, ¹⁷Univ. Natl. de Colombia, Bogotá, Colombia, ¹⁸Univ. Natl. de Colombia, Bogota, Colombia, ¹⁹VA Puget Sound Hlth Care Sys, Seattle, WA

Abstract Body:

Genome-wide association studies of autosomes have identified over 90 independent risk signals for Parkinson's disease (PD), but little is known about the impact of X chromosome (chrX) variants on PD susceptibility. Only one previous study (LeGuen et al., 2021) has investigated the X chromosome in PD, and this was done in a cohort of European ancestry. In this study we performed the first X chromosome-wide association study (XWAS) of PD in an admixed Latino cohort. We used genome-wide data from the Latin American Research consortium on the GENetics of PD (LARGE-PD) (1,294,079 variants, N=1,498), including individuals from Brazil, Chile, Colombia, Peru, and Uruguay. We used the TOPMed as a reference to impute our data. Since the chrX e has different recombination dynamics and is more impacted by evolutionary forces, we adapted and followed LeGuen et al. chrX-specific pipeline to make it viable for admixed populations. We performed association analysis (sex-stratified and not) in two different ways: (i) all samples and (ii) by country, followed by meta-analysis. For all association tests, the models were adjusted for the first 10 chrX PCs + Age. We used chrX PC due to the low correlation between autosomal and chrX PCs observed in our data. We calculated the significance threshold using the number of effective tests based on the pair-wise correlation among SNPs. We obtained 81 statistically significant variants across five loci. To replicate our findings, we used data from Latinos in the International Parkinson Disease Genomics Consortium integrated with the Brazilian Bambuí Cohort Study of Aging. We replicated the association for rs393033 in females (discovery OR = 2.012, 95% CI = 1.387 to 2.292, p= 0.000231 and validation OR = 1.911, 95% CI = 1.051 to 3.475, p = 0.033785). Also, 12 variants identified during the discovery analysis were significantly associated but in a sex-stratified analysis other than from the discovery test. In conclusion, we performed the first XWAS for PD in admixed populations. We also provided a chrX-specific pipeline to facilitate ChrX association analysis using admixed populations. Using LARGE-PD data we nominated a novel intergenic variant, rs393033, located 1.2 Mb from FMR1 gene, which needs further studies.

Mendelian Phenotypes Posters - Wednesday

PB1723. 4-years of Face2Gene in a General Genetics Clinic: Insights from Retrospective Analysis of Diagnosed Cases

Authors:

M. Muriello, N. Xiong, D. Basel; Med. Coll. of Wisconsin, Milwaukee, WI

Abstract Body:

Diagnosing patients with non-specific dysmorphic features is a challenge faced by medical geneticists. Machine learning (“artificial intelligence”) technologies in medicine have the potential to improve the diagnostic yield and shorten the diagnostic odyssey. Face2Gene (F2G) is a facial image analysis software developed using deep-learning algorithms to detect features from thousands of patient images¹. Face2Gene was used as a clinical tool for a subset of undiagnosed cases, being seen in the pediatric genetics clinic at the Medical College of Wisconsin between 2017 and 2021. Among 1747 unique F2G cases created in that timeframe, 113 had a final diagnosis added in F2G; 101 of which were molecularly confirmed and 12 were clinically diagnosed. Data was collected from the electronic medical record including demographics, final diagnosis, primary clinical features, diagnostic test that diagnosed the patient’s molecular diagnosis, other genetic testing, Face2Gene top-10 accuracy, gestalt score, and a subjective assessment of picture quality. The final diagnosis was present in the top-10 syndromes by gestalt score in 68%. Amongst those diagnosed where the diagnosis was not in the top 10, none had higher than a low gestalt score. Amongst those diagnosed where the diagnosis was in the top 10, 46% had a medium or higher gestalt score. Neither age, sex nor ethnicity had a significant impact on top-10 accuracy. Subjective assessment of picture quality rated 55% as high-quality, 40% as medium-quality, and 5% as low-quality. There was no difference in top-10 accuracy based on picture quality. Several diagnoses for which there were non-specific clinical features were made by targeted testing prompted by a medium or high gestalt score. Estimates of the utility F2G have been variable, with top-10 sensitivity of Gestalt analysis ranging from 48-91%¹⁻⁵. We found 68% top-10 accuracy which may reflect the real-world applicability of Face2Gene in a general pediatric genetics clinic. A medium or higher gestalt-score in almost all diagnosed cases suggests this cutoff may be useful. The accuracy was not significantly impacted by age, sex, ethnicity, or picture quality, suggesting Face2Gene may be useful for most patients. This study had several limitations. Not all cases had features entered, so the impact of the feature score could not be analyzed. The sample may be biased by lack of standardized process for updating F2G cases with a final diagnosis.

Mendelian Phenotypes Posters - Thursday

PB1724. A A recurrent de novo mutation in *ZMYND11*

Authors:

T. Carrion, Alba Muñoz Santa María Antonieta Ballesteros Vizoso Laura Torres Juan Alexander Damián Heine Suñe; Hosp. universitario son espases, PALMA DE MALLORCA, Spain

Abstract Body:

Background/Objectives:

A 7-year-old patient initially diagnosed as Silver-Russel syndrome due to mosaicism of chromosome 7 returns to our genetic consultation for diagnostic re-evaluation. Characteristics symptoms are global developmental delay (84%), hypotonia, feeding difficulties, short stature and several facial dysmorphisms (microcephaly, depressed nasal bridge and microretrognathia).

Methods:

Karyotype and CGH array were performed. MLPA assay (MS-MLPA Probemix ME032 DUP7-DUP14) was also done to test if there was maternal uniparental disomy of chromosome 7 (DUP7). Clinical exome was sequenced by Human Whole-Genome Sequencing with the Nextera™-DNA-Flex-Library Preparation Kit (Illumina).

Results:

Karyotype and CGH array were normal. MLPA assay revealed absence of DUP7. The analysis of the clinical exome showed a heterozygous autosomal dominant missense variant: c.1798C>T p.(Arg600Trp) in *ZMYND11* (NM_006624.5), classified according the ACMG guidelines as pathogenic. Genetic testing of both parents showed it was arisen de novo.

Conclusion:

Zinc finger MYND-type, expressed in many human tissues, acts as a transcriptional repressor, playing an inhibitory role in the muscle and neuronal differentiation steps. Specifically, the mutated position in this case Arg600 is very conserved and essential for its binding to ligands (Kateb et al.,2013). *ZMYND11* has also been proposed as a candidate gene for 10p15.3 microdeletion syndrome, which shares common clinical features with our patient (DeScipio et al., 2021). Moreover, other authors have described the same mutation (Cobben et al., 2014), so that it can be considered as definitely pathogenic. Finally, the same SNP has been reported 4 times in Decipher and 8 times in ClinVar thus the variant here reported could be a hotspot mutation.

Mendelian Phenotypes Posters - Wednesday

PB1725. A case of atypical inheritance in late-onset Pompe disease

Authors:

M. Mroczek¹, J. Meienberg¹, P. Rejmer², C. Henggeler¹, M. R. Baumgartner³, G. Matyas¹; ¹Ctr. for Cardiovascular Genetics and Gene Diagnostics, Swiss Fndn. for People with Rare Diseases, Schlieren-Zurich, Switzerland, ²Seegarten Klinik AG, Klichberg, Switzerland, ³Div. of Metabolism and Children's Res. Ctr., Univ. Children's Hosp. Zurich, Univ. of Zurich, Zurich, Switzerland

Abstract Body:

Background and Aims: Pompe disease is a rare autosomal-recessive disorder caused by acid α -glucosidase (GAA) deficiency and characterized by proximal progressive muscular weakness and/or respiratory insufficiency. Carriers of one *GAA* pathogenic variant are usually considered to be asymptomatic, although they can have reduced enzyme activity. Here, we present a case heterozygous for the *GAA* c.-32-13T>G mutation and having a GAA level and clinical symptoms comparable with late-onset Pompe disease (LOPD). **Materials and Methods:** For genetic testing, whole-genome sequencing (60x, PCR-free, PE150) has been applied. The GAA enzyme activity has been measured in leucocytes and cultured fibroblasts. RNA-Seq has been performed from the patient's fibroblasts. **Results:** A 28-year-old patient presented with an adolescent onset, proximal, lower limb weakness and lower leg atrophy. He suffers from severe myalgia and is able to walk only a few steps. Functional lung tests showed mild restriction. In fibroblasts, the activity of GAA was 0.83 (reference 6.04-17.06 nmol/min/mg prot.; 5-13%N). In leucocytes, the activity of GAA was normal, however, quotient +/- acarbose was reduced (0.24, reference 0.45-0.63). The patient inherited the following variants that we considered as possibly significant: from the mother *GAA* c.-32-13T>G (pathogenic); from the father *GAA* c.852G>A p.(=) (benign), *PHKB* c.445C>T p.(Leu149Phe) (VUS), and *AMPDI* c.133C>T p.(Gln45Ter) (VUS). No other variant with suspected significance in the *GAA* gene has been identified. RNA-Seq confirmed abnormal transcripts due to c.-32-13T>G but no other splicing defects in *GAA*. As *GAA* and *PHKB* gene products are parts of the same metabolic pathway, we hypothesize that a digenic inheritance with synergistic heterozygosity could occur. **Discussion:** In LOPD the levels of GAA enzyme activity and the severity of the clinical pictures may be highly variable among individuals. Most of the LOPD patients carry biallelic mutations in the coding region of *GAA*. However, a small number of individuals have (deep) non-coding *GAA* variants or a mosaicism. There are also several cases where only one pathogenic *GAA* variant has been identified, alone or together with other heterozygous pathogenic variants related to neuromuscular diseases, so that digenic/oligogenic inheritance has been suggested. The treatment with enzyme replacement therapy led to a clinical improvement in a few heterozygous LOPD cases. **Conclusion:** There may be a small cohort of LOPD patients where a symptomatic heterozygosity or digenic/oligogenic inheritance can be considered. Reporting further cases could confirm or contradict this hypothesis.

Mendelian Phenotypes Posters - Thursday

PB1726. A case of Fanconi anemia diagnosed due to *BRCA2* gene mutation triggered by cancer genome profiling test.

Authors:

K. Fukuda¹, S. Masuda¹, M. Matsui¹, S. Ito¹, M. Kuroda¹, H. Yamanaka¹, H. Futagawa¹, H. Muramatsu², M. Wakamatsu², H. Yoshihashi¹; ¹Tokyo Metropolitan Children's Med. Ctr., Fuchu-shi, Japan, ²Nagoya Univ. Hosp., Nagoya-shi, Japan

Abstract Body:

Fanconi anemia (FA) is a hematologic disorder characterized by progressive pancytopenia, transition to MDS and AML, abnormal body morphology, and solid tumors against a background of chromosomal instability based on DNA repair defects. There are currently 22 subgroups of FA, and the corresponding causative genes have been identified for each. Among them, FA caused by mutations in the *BRCA2* (*FANCD1*) gene (FA-D1) has a high risk of carcinogenesis and an extremely poor prognosis, and is inherited in an autosomal recessive form. We report a male infant with Glioblastoma who was diagnosed with FA-D1 and required genetic counseling for the parents, including diagnosis of carriers. A 20-month-old boy was brought in by ambulance for a cluster of seizures. He was transferred to our hospital for close examination and treatment because a CT scan of the head showed a left frontal lobe mass. He had no perinatal or medical history other than developmental delay in infancy, short stature, microcephaly, and auricular deformity. Various imaging tests were performed, including right renal hypoplasia seen on abdominal CT and a 6-7 cm intraparenchymal tumor in the left frontal lobe revealed on Head-MRI. Therefore, chemotherapy was started after partial resection. A cancer genome profile test performed to find molecularly targeted drugs detected a nonsense variant of the single-allelic *BRCA2* gene in the tumor cells. He was highly suggestive of FA in combination with multiple dysmorphic features. Chromosome fragility tests showed numerous chromosomal breaks, and whole exome analysis using the peripheral lymphocytes identified variants of the *BRCA2* gene (c.A7969T, p.K2657Ter and c.C7847T, p.S2616F) in a compound heterozygous manner. The uniallelic variant of the *BRCA2* gene was found in each of the parents. It was necessary to comprehensively examine the clinical course and physical characteristics of the affected children, as well as the consistency of the cancer genome profile test results. The results of the whole exome analysis revealed an accurate picture of the genetic background within the family. The pathological variants obtained from tumor cells provided evidence for the diagnosis of an affected child, treatment decisions, and prognosis of life expectancy, as well as for preventive cancer surveillance of hereditary breast and ovarian cancer in the parents. In addition, it was useful for estimating recurrence rates with respect to FA-D1 in the next pregnancy. Genomic practice collaboration among multiple departments was useful, leading to genetic counseling across pediatric rare diseases, hereditary tumors, and prenatal diagnosis.

Mendelian Phenotypes Posters - Wednesday

PB1727. A case of fatal cardiac arrest in a neonate diagnosed with VLCADD postmortem

Authors:

P. Singh¹, D. Amaro², O. Obi², F. Kiran², E. Hediger¹, T. L. Toler¹, P. Dickson¹, D. Grange¹; ¹Washington Univ. Sch. of Med. in St. Louis, Saint Louis, MO, ²Univ. of Missouri Sch. of Med. in Columbia, Columbia, MO

Abstract Body:

Abstract: Newborn screening has become advantageous in providing nutritional intervention and medication treatment to alter long term outcomes for patients. Yet, the interventions can be limited in cases of a more severe presentation. We report on a case of Very Long-Chain Acyl-Coenzyme A Dehydrogenase (VLCAD) Deficiency diagnosed post-mortem. Our patient was born to a 26-year-old G1P0 by uncomplicated vaginal delivery after presenting in labor. Antenatally, the pregnancy was uncomplicated, and mother reports only taking prenatal vitamins. APGARS were 9 and 9 at one and five minutes, respectively, and she was transitioned for routine care neonatally. There was concern for grunting episodes on DOL1 and breath holding yet this resolved spontaneously. On the evening of DOL1 (30 hours of age), she became unresponsive resulting in resuscitation efforts by the neonatal intensive care unit and pediatric cardiology. An echocardiogram showed no abnormalities, and continuous cardiac monitoring showed a pattern of wide QRS complexes with peaked T waves refractory to intervention. Initial venous blood obtained 1 hour into resuscitation showed a pH 7.002 and a base deficit of 10. Initial K was >10 mmol/L. It was after over 2 hours of resuscitation that efforts were halted. Newborn screening results received after death indicated a positive screen for VLCAD Deficiency with a C14:1 level of 3.20 µmol/L (normal <0.60 µmol/L). Post-mortem examination was performed, and family was consented for genetic testing. Examination of the placenta was notable for rare pigmented macrophages concerning for meconium aspiration. Autopsy findings included a structurally normal heart without hypertrophy, lungs with some evidence of meconium aspiration, stomach with early ischemic changes, kidney with tubular vacuolization, and liver with microsteatosis. Genetic testing included a normal chromosomal microarray, and whole exome sequencing revealed biallelic variants in *ACADVL* (maternal c.1375dup, p.(R459Pfs*4) pathogenic variant and a paternal c.1678+3_1678+6del? likely pathogenic variant). The maternal variant is predicted to result in nonsense mediated decay, and the paternal variant supports a deleterious effect on splicing; these are consistent with the loss of function mechanism of disease of VLCAD Deficiency. Our patient suffered from cardiac arrest leading to demise, yet increased awareness of this more severe presentation could alter prognosis. We present a case with both postmortem confirmatory genetic testing and consistent postmortem findings of hepatic steatosis of an acute form of VLCAD deficiency resulting in neonatal demise.

Mendelian Phenotypes Posters - Thursday

PB1728. A case report of an Egyptian patient with a severe neurodevelopmental disorder and a novel biallelic loss-of-function variant in *GOLGA2*.

Authors:

R. Khang¹, A. Kenawy², R. N. El Kharadly², G. Seo¹, H. Lee¹; ¹3billion, Seoul, Korea, Republic of, ²Med. Res. Inst., Alexandria, Egypt

Abstract Body:

GOLGA2 gene encodes GM130 protein that constitutes Golgi apparatus (GA) which plays a critical role in protein secretion and sorting. The symptoms of various Mendelian disorders associated with GA component proteins are heterogeneous, but neurodevelopmental phenotype is one of the common phenotypes in relatively severe disorders. *GOLGA2* gene is not yet associated with a human disorder with only a couple of patients reported with homozygous loss-of-function (LoF) variants. *Golga2* knockout (KO) mice present with autophagy and fibrosis in lung and liver, and *golga2* knocked down zebrafish show severe skeletal muscle disorganization and microcephaly.

Here, we report a patient with a novel homozygous variant in *GOLGA2*. The patient was a 2-year-old female born after a full term uneventful pregnancy. Her birthweight was 2250 grams. On clinical examination, she had growth retardation, global developmental delay, hypotonia and dysmorphic features including microcephaly, deep-set eyes, upturned nasal tip, long philtrum, thin upper lip, and pointed chin. She also had bridged simian creases, hypoplastic nails and overlapping toes. Brain MRI showed hypoplastic corpus callosum and delayed myelination for age. EEG showed generalized epileptiform activity. The ophthalmological fundus examination and hearing assessment were normal. Family history was significant with 3 deceased siblings presenting with similar symptoms. Exome sequencing was performed on the patient using xGen Exome Research Panel v2 (Integrated DNA Technologies) and Illumina NovaSeq 6000. A mean coverage of 231X was achieved and the patient had regions of homozygosity across ~1.4% of the genome, suggesting the parents are distantly related. There were no clinically significant variants that could explain the patient's phenotype. However, there was a homozygous 4-base deletion variant that is predicted to result in LoF (NM_001366244.1:c.2018_2021del; p.Lys673ArgfsTer31), truncating the last 345 amino acids of the protein.

The two previously reported patients with variants in *GOLGA2* were a 10.5-months old female and a female who started showing symptoms at 6 months of age. They had homozygous LoF variants c.1311_1314delAGAG (p.Glu423ArgfsTer6) and c.2296C>T (p.Gln766Ter), predicted to result in similar consequences as the variant found in our patient. All of our patient's symptoms were reported in these two patients with a similar age-of-onset. Although finding additional cases with similar symptoms and variants, and performing functional studies are warranted, our patient augments the possibility of biallelic LoF variants in *GOLGA2* leading to a severe neurodevelopmental disorder.

Mendelian Phenotypes Posters - Wednesday

PB1729. A case with mosaicism pigmentary and Prader-Willi Syndrome: importance of molecular diagnosis in cases of mosaicism

Authors:

Y. Gasparini, E. Moura, M. Montenegro, V. Almeida, B. Wolff, G. Carvalho, A. Mendes, L. Vieira, L. Kulikowski; Univ.e de Sao Paulo, Sao Paulo, Brazil

Abstract Body:

Mosaicism is a usual event in the world population, defined by the presence of two or more genetically different cell lines in the same individual, arising from the same zygote. However, the characterization of mosaicism still poses a great challenge for laboratorial medicine, since variants may not be present in all tissues. This is associated with mosaicism confined to tissue or pigmentary mosaicism, which constitutes a heterogeneous group of skin pigmentation alterations associated with multisystem involvement. We present a case of a three year-old male infant, born from a non-consanguineous healthy couple. The patient displayed hypotonia, sinophre, long eye-lashes, ptosis, simplified helix ear, retrognathia, high palate, congenital heart disease, spots in Blaschko lines and bilateral cryptorchidism. We have collected samples from peripheral blood, saliva and skin fibroblasts. Results from cytogenetic analysis (GTG Band) performed in peripheral blood were normal (46,XY). Unexpected results from skin fibroblasts and saliva using genomic array (cytoSNP850k-Illumina) revealed a trisomy of chromosome 15 in mosaic. Additionally, a methylation test was performed for the *SNURPN* gen, which showed an exclusively maternal methylation pattern, with no evidence of the paternal allele. The most commonly accepted mechanism behind cases of UPD for different chromosomes reported in literature is trisomy rescue. This mechanism may cause different outcomes in somatic cells mitosis: hetero-UPD (rescue), a normal cell (rescue) or trisomy (no or incomplete rescue). Incomplete rescue of trisomy 15 may lead to the development of mosaic mutations and chromosome rearrangements, which are associated with cases of PWS with maternal UPD.

Mendelian Phenotypes Posters - Thursday

PB1730. A Difficult Dual Diagnosis: Dilemmas in the Era of Next Generation Sequencing

Authors:

A. Rekab, K. Anyane-Yeboah; Div. of Clinical Genetics, Dept. of Pediatrics, Columbia Univ. Med. Ctr., New York, NY

Abstract Body:

The advent of next generation sequencing (NGS) in the clinical setting has increased the frequency of identification of dual diagnoses, which can present challenges for clinicians. Here we describe a case involving a dual diagnosis of two rare neurological conditions with overlapping phenotypes.

Our patient is a 2-month-old female with seizures at 6 weeks old. Brain MRI was unremarkable, and EEG demonstrated focal seizures. On physical exam she was hypotonic with asymmetric facies, plagiocephaly and torticollis. SNP microarray was negative, but an epilepsy panel revealed a likely pathogenic variant in *CDKL5* (c.826-1G>A, splice acceptor) and a pathogenic variant in *KIF1A* (c.2362G>T, p.E788*). Both variants are expected to cause loss of function and were de novo.

CDKL5 deficiency disorder (CDD) is an X-linked developmental epileptic encephalopathy caused by pathogenic variants in the *CDKL5* gene and is characterized by epilepsy with onset by three months of life, global developmental delays, intellectual disability, hypotonia and cortical visual impairment. CDKL5 plays a role in brain development and function. KIF1A-associated neurological disorder (KAND) is a neurodegenerative condition characterized by spastic paraplegia, peripheral neuropathy, developmental delays, seizures and optic nerve atrophy. KAND is caused by mono- or bi-allelic pathogenic variants in *KIF1A* which encodes a component of intracellular transport in axonal microtubules, essential for neuronal survival.

The overlapping phenotypes of these two conditions bring to light several management and counseling dilemmas for clinicians. In cases such as ours, a dual diagnosis complicates the ability of clinicians to clearly delineate symptoms, anticipate the possible evolution of phenotype, and has the potential to be a source of confusion and anxiety for families. This patient's clinical picture is well explained by the *CDKL5* variant, yet the possibility of future *KIF1A* contribution to her phenotype cannot be overlooked. A dual diagnosis can also complicate eligibility criteria for some clinical research studies. With KAND treatments on the horizon, our patient's eligibility may be challenged due to her *CDKL5* variant. Thus, raising the question of how clinicians can equitably care for patients with dual diagnoses.

A dual diagnosis can present challenges for clinicians and families alike. Greater structure in clinical management recommendations could alleviate the uncertainty in these cases. Further research on patients with dual diagnoses is paramount to better equip clinicians' management and counseling, as NGS increases the recognition of multiple genetic diagnoses.

Mendelian Phenotypes Posters - Wednesday

PB1731*. A dysmorphology physical examination entry system facilitates structured genetic phenotype capture and natural language processing

Authors:

I. Campbell^{1,2}, A. Magge³, J. C. Priestley¹, K. M. Szigety¹, S. F. Schmidt¹, M. L. McManus¹, S. E. Sheppard⁴, E. M. Lourie^{1,2}, N. Muthu^{1,2}, H. Hakonarson^{1,2}, G. Gonzalez-Hernandez³; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Univ. of Pennsylvania, Philadelphia, PA, ³Cedar Sinai Hosp., Los Angeles, CA, ⁴Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., Rockville, MD

Abstract Body:

The dysmorphology physical examination is a critical component of diagnostic evaluation in clinical genetics, as it directly influences the selection and interpretation of genetic testing. Such information is also extremely useful to researchers attempting to delineate undescribed genetic conditions. However, this key information is nearly always captured within the electronic health record (EHR) as unstructured free text, making it unavailable for downstream computational analysis. To address this issue, we implemented a specialized dysmorphology exam system in our EHR that speeds the clinical workflow while also capturing common examination findings discretely through button clicks and continuous anthropometric measurements through numeric fields. Other findings are documented in free text fields. To extract additional structured phenotypic findings from the free text, we implemented a novel recurrent neural network classifier to predict human phenotype ontology (HPO) terms. We trained the classifier using the synonyms present in the HPO and UMLS Metathesaurus. We generated additional training data by manually annotating 2,013 physical exam text strings describing pediatric patients. Review of inter-annotator agreement among our 4 annotators found 76% complete concordance for the 890 text strings annotated at least twice. We found the classifier had 83% accuracy on the validation data, which was a substantial improvement upon a previously published HPO classifier, PhenoTagger (68% accuracy). Using this system, we captured a median of 1 abnormal HPO phenotype per participant from discrete data and a median of 3 abnormal phenotypes from free text. Finally, we analyzed the most frequently documented findings in free text and used this information to improve options for discrete capture. In conclusion, the development of specialized physical exam systems has the potential to both speed clinical documentation and facilitate capture of discrete human phenotypes. Moreover, natural language processing is a promising approach for secondary EHR data reuse in genetics. We have partnered with the EHR vendor to bring this functionality to other customers.

Mendelian Phenotypes Posters - Thursday

PB1732. A Highly Polymorphic VNTR in the *DRD4-DEAF1* Intergenic Region

Authors:

D. Vandenbergh, C. A. Brogan, A. T. Apsley, A. J. Burich; Penn State Univ, University Park, PA

Abstract Body:

Adolescent Idiopathic Scoliosis (AIS) impacts 4 out of every 100 adolescents (Konieczny et al., 2013), but its etiology is not well understood. In a Genome-Wide Association Study, an association was found with SNP rs11604855 (Liu et al., 2017), which is in the intergenic region between *DRD4* and *DEAF1*. Whether it plays a causative role in AIS, or is linked to a second causative site is unknown. We identified a novel Variable Number of Tandem Repeats (VNTR) locus less than 2 kb from the SNP that is a 28-nucleotide repeat present as 4.9- and 75-copy variants in the reference and alternate genomes. Analysis of 64 long-read sequences revealed 34 variants ranging from 3 to 156 copies of the repeat. Alleles at this VNTR are in linkage disequilibrium with rs11604855 raising the possibility that the association of the SNP and AIS might be due to functional effects of this highly variable tandem repeat.

Mendelian Phenotypes Posters - Wednesday

PB1733. A homozygous variant in a novel gene "*AP2A2*" causes an early-onset hereditary spastic paraplegia in a Malian family.

Authors:

S. Diarra^{1,2,3}, S. Gosh⁴, L. Cissé⁵, T. Coulibaly^{2,5}, A. Baneye⁵, G. Harmison¹, S. Diallo⁶, S. H. Diallo⁶, O. Coulibaly⁷, A. Schindler¹, A. Bocoum², C. A. K. CISSE², A. Yalcouye^{5,8}, S. Lakhani³, E. Mis³, M. Khokha³, O. samassekou², M. Traoré⁹, S. Jacobson¹⁰, C. Blackstone¹¹, C. O. Guinto⁵, J. S. Bonifacino⁴, G. Landoure^{1,2,5}, K. H. Fischbeck¹, C. Grunseich¹; ¹Neurogenetics Branch, NINDS, NIH, Bethesda, MD, ²Université des Sci., Techniques et Technologies de Bamako (USTTB), Bamako, Mali, ³Pediatric Genomics Discovery Program (PGDP), Yale school of Med., New Haven, CT, ⁴NeuroSci.s and Cellular and Structural Biology Div., NICHD, NIH, Bethesda, MD, ⁵Service de Neurologie, CHU du Point "G", Bamako, Mali, ⁶Service de Neurologie, CHU du Gabriel Touré, Bamako, Mali, ⁷Service de Chirurgie pédiatrique, CHU of Gabriel Touré, Bamako, Mali, ⁸Cytogenetic Dept., Univ. of Cape Town, Cape Town, South Africa, ⁹Service de Cytogénétique et la Réproductive Biologique, INRSP, Bamako, Mali, ¹⁰Neuroimmunology Div., NINDS, NIH, Bethesda, MD, ¹¹Massachusetts Gen. Hosp., Charlestown, MA

Abstract Body:

Background: Hereditary spastic paraplegias (HSPs) are a heterogeneous group of neurodegenerative disorders characterized by lower-extremity spasticity with "pure" and "complicated" forms. Several genes and loci have been associated HSPs, but many other clinically entities remain with no molecular diagnosis. **Objectives:** Characterize patients with HSPs and identify the underlying genetic defect. **Methods:** After consent, patients went through a thorough clinical examination. Blood chemistries, brain and spinal imaging were done to exclude common causes. DNA was extracted for WES. Variants were checked in public and local SNP databases, and classified according to (ACMG). Functional studies were performed using iPSC cells differentiated into iPSCs like neurons (iNeur) obtained from patients and relatives' fibroblasts (fib). SgRNA were made for CRISPR/cas9 knockout of the target gene in frog embryos to further confirm variant pathogenicity. **Results:** We identified a consanguineous Malian family with three siblings presenting symptoms consistent with complicated HSP. The age of onset was ~8months-old. Neurological examination found lower limb weakness, equine feet, spastic gait, and cognitive decline. Brain imaging showed thin corpus callosum with cortical and cerebellar atrophy and EEG was abnormal. HSP candidate gene and targeted panel gene were negative. However, WES identified a homozygous missense variant in *AP2A2* gene. The variant segregated with the disease phenotype in the family and not present in SNP databases. *AP2A2* has not been previously associated with any disease, but it is a member of a complex associated with other HSPs. *AP2A2* protein is known to be highly expressed in the spinal cord and fetal brain and has role in endocytosis function. It is also conserved through diverse species and located in appendage C-terminal domain, a platform of protein-protein interactions domain. Functional studies showed no difference in protein and mRNA levels in fibroblasts, however reduction in protein levels was observed in neuronal cells. Immuno-precipitation of mutant AP2-appendage alpha-C construct shows a defect of binding to accessory proteins. Endocytosis function was tested through transferrin receptor (TfR) uptake in fib and iNeur, and showed decreased uptake of iron uptake in patients neuronal cell. In additional, we observed neurite swelling in 11-day-old patient neuronal cultures. **Conclusion:** Our findings suggest a novel candidate gene causing HSP. Further studies are underway to evaluate the functional implication of this variant in *X. tropicalis* model. **Keyword:**HSPs, exome sequencing, *AP2A2*, clathrin-mediated endocytosis, Mali

Mendelian Phenotypes Posters - Thursday

PB1734. A multi-branch family with mutation in *SLC26A4* gene from a village in the southeastern region of Iran.

Authors:

M. Mohseni¹, S. Dorgaleleh², M. Beheshtian¹, F. Keshavarzi¹, F. Ghodratpour¹, F. Zare Ashraf¹, K. Jalalvand¹, S. Arzhangi¹, N. Nikzat¹, K. Kahrizi¹, H. Najmabadi¹; ¹Genetics Res. Ctr., Univ. of Social Welfare and Rehabilitation Sci., Tehran, Iran, Islamic Republic of, ²Student Res. Committee, Golestan Univ. of Med. Sci., Gorgan, Iran, Islamic Republic of

Abstract Body:

In order to reduce the burden of hereditary hearing loss (HHL), the most frequent sensorineural disorder influencing 1 in 500 newborns around the globe, molecular diagnosis is crucial; particularly in countries with higher rates of consanguineous marriages such as Iran. Recently, we identified a village located at the southeastern part of Iran with many people suffering from HHL and decided to determine the causative gene. In present study, investigation of molecular cause of autosomal recessive hearing loss (ARHL) in a multi-branch family originating from a village in southeastern Iran with 31 affected members manifesting moderate to severe HL. After a comprehensive clinical examination on affected individuals with the history of GJB2-negative HL, we subjected an affected individual to whole exome sequencing (WES). Analysis of WES data was mainly performed by filtering out variants with minor allele frequency >0.05, particularly reported in Iranian populations using the Iranome database. Applying Sanger sequencing, we confirmed candidate variants. We detected the known missense mutation, c.716T>A (p.V239D), in the sixth exon of the *SLC26A4* gene, which had been firstly reported in Iran in Kurdish families. Our Sanger sequencing data confirmed the presence of this pathogenic variant in the rest of 30 affected members in this family. Previously, the p.V239D mutation was reported from Pakistan, in the southeast of Iran, causing ARHL. Given that the *SLC26A4* gene as the most common gene after *GJB2* in Iran, causing ARHL, Finding this mutation in a village next to the border of Pakistan might reflect the magnification of a founder effect, and contribute to the final genetic make-up of this region's population.

Mendelian Phenotypes Posters - Wednesday

PB1735. A new case of YARS1 associated multisystemic disorder with compound heterozygous in a patient with Klinefelter syndrome.

Authors:

J. Kuan, A. B. Hansen, H. Wang; Loma Linda Univ. Sch. of Med., Loma Linda, CA

Abstract Body:

Introduction: Aminoacyl-tRNA synthetases are a group of enzymes that catalyze the coupling of amino acids to their respective tRNAs. Tyrosyl tRNA synthetase, encoded by the gene YARS1, is responsible for the aminoacylation of tyrosine to its tRNA. Heterozygous variants in YARS1 are associated with autosomal dominant Charcot-Marie-Tooth type C. Recently, cases of biallelic variants in YARS1 have been reported to result in an autosomal recessive multisystemic disease. Here we report a new case with compound heterozygous variants in YARS1 in a patient with Klinefelter syndrome. **Case presentation:** A 12-week-old Hispanic male born at 37 weeks was admitted for 6 weeks of emesis and 2 weeks of intermittent seizure-like activity. Abdominal ultrasound was suggestive of possible hepatocellular dysfunction. EEG showed focal sites of hyperexcitability. Brain MRI was unremarkable. He failed initial and subsequent hearing screenings in the left ear. The hospitalization was complicated by intermittent hyperkalemia. Microarray revealed Klinefelter syndrome. At 14 months old he presented to genetics with microcephaly, hypotonia and significant delay in milestones. A PEG tube was placed for failure to thrive. Family history revealed that his father has adult-onset hearing loss, wears hearing aids, and has pain and weakness in his feet. Consanguinity was denied. Considering that Klinefelter cannot explain the patient's clinical phenotypes, whole exome sequencing was performed, which revealed one likely pathogenic variant (c.1099C>T, p.R367W, paternally inherited, reported before) and one variant of uncertain significance (c.782T>G, p.L261R, maternally inherited, never been reported) in the YARS1 gene. Based on the clinical phenotypes, the diagnosis of YARS1 related multisystemic disorder was established. **Discussion/Conclusion:** This disorder was recently updated by OMIM #619418 and named as infantile-onset multisystem neurologic, endocrine, and pancreatic disease type 2 (IMNEPD2). It is characterized by global developmental delay, microcephaly, FTT, sensorineural deafness, retinal abnormalities with visual defects, hypotonia, hepatomegaly, endocrine abnormalities, and brain findings of dysmyelination, thin corpus callosum, cerebral atrophy and white matter abnormalities. Unique for this case is that the father likely has Charcot-Marie-Tooth disease, indicating that both dominant and recessive disorders are present in the same family, and that this is the first report of this disorder in a patient complicated with Klinefelter syndrome, demonstrating that comprehensive genetic testing is important for patients with complicated phenotypes.

Mendelian Phenotypes Posters - Thursday

PB1736. A novel de novo nonsense variant in *TBX2* cause osteochondrodysplasia

Authors:

M. Umair, M. Alfadhel; Med. Genomics Res. Dept., King Abdullah Intl. Med. Res. Ctr. (KAIMRC), King Saud Bin Abdulaziz Univ. for Hlth.Sci. (KSAU-HS), Ministry of Natl. Guard Hlth.Affairs (MNG-HA), Riyadh, Saudi Arabia., Riyadh, Saudi Arabia

Abstract Body:

Background: Genetic skeletal disorders are a diverse heterogeneous group of developmental disorders of the skeletal system and cartilaginous tissues. According to the nosology classification, many types have been classified according to molecular analysis and clinical presentation. **Methods:** Using molecular and clinical methods we characterized a proband having features of osteochondrodysplasia. The affected individual's age was 5 years at the time of study and who passed away at 6.5 years exhibited features such as chondrodysplasia phenotypes, short stature, and global developmental delay. Additional clinical investigation revealed platybasia at the skull base delayed myelination, chest showing bilateral subsegmental atelectasis/consolidation in the lower lobes, bilateral hip dislocation, and diffuse central vertebral endplate depression. Molecular diagnosis was performed using Trio-Whole-exome sequencing using standard methods. **Results:** WES identified a novel heterozygous nonsense mutation p.Lys177Ter in exon 2 of the *TBX2* gene, confirmed by Sanger sequencing. The amino acid Lys177 is conserved across different species. **Conclusion:** *TBX2* is important for the development of the skeleton and the brain and four prior reports described *TBX2* association with craniofacial dysmorphism, complex vertebral anomalies, and endocrine dysfunctions. Our findings expand the current spectrum of *TBX2* associated genetic skeletal disorder and support the evidence that disease causing variants in *TBX2* have heterogeneous clinical presentations associated with osteochondrodysplasia.

Mendelian Phenotypes Posters - Wednesday

PB1737. A novel homozygous missense mutation in *ARSK* causes a new subtype of MPS.

Authors:

M. Sun¹, L. Randolph²; ¹Children's Hosp. Los Angeles/Keck Sch. of Med. of USC, Los Angeles, CA, ²Children's Hosp. Los Angeles, Los Angeles, CA

Abstract Body:

Mucopolysaccharidoses (MPS) are a group of rare inborn errors of metabolism caused by defective lysosomal enzymes which prevent cells from recycling certain carbohydrates and fats, causing storage of glycosaminoglycans in cells throughout the body. This leads to multisystem abnormalities including connective tissues, brain, blood, spinal cord and other tissues. A recent study suggested defects in *ARSK* may lead to a new subtype of MPS, MPS X.

Our patient is a 13-year-old cognitively intact boy of Syrian ancestry with Perthes disease and pectus carinatum referred for possible skeletal dysplasia. He was born to nonconsanguineous parents. His height is at the 18th percentile, relatively short for his mean parental height. His skeletal survey showed partially collapsed concave bilateral femoral capital epiphyses, striated trabecular appearance of the ends of the long bones, platyspondyly with anterior beaking, and exaggerated thoracic kyphosis, consistent with skeletal dysplasia, “likely spondyloepiphyseal dysplasia”. At that time he had normal levels of glycosaminoglycans in his urine.

Trio exome sequencing revealed a novel homozygous missense germline variant, c.1067C>A (p.S356Y), in *ARSK* (reference sequence: NM_198150.3). Both parents were confirmed as heterozygous carriers. This variant occurs at a highly conserved amino acid position within conserved sulfatase domain. *In-silico* predictions suggest a deleterious effect of this change. This variant has not been reported either as a benign or disease-causing variant in human. It has not been observed in the gnomAD database (<https://gnomad.broadinstitute.org/>).

To date, only two variants (p.R84C and p.L184X) associated with *ARSK*-related conditions have been reported in the literature. Both showed childhood onset with relatively mild MPS features, such as a mild elevation of dermatan sulfate detected by the more sensitive method of LC-MS/MS. Additional cardiac and ophthalmological abnormalities were noticed on follow-up examination. Our exome findings of the homozygous *ARSK* variant (p.S356Y) combined with the constellation of clinical presentation suggest that our patient likely has the newly defined MPS type X condition. The effect of *ARSK* deficiency due to the p.S356Y variation in *ARSK* are under investigation with functional studies and repeat urine MPS testing at an alternative laboratory, which may help us better understand its pathogenic roles on GAG storage in humans.

Mendelian Phenotypes Posters - Thursday

PB1738. A novel homozygous variant in *HECW2* gene is associated with intellectual disability and epilepsy in a Malian family.

Authors:

M. Sangare^{1,2}, M. Dembele^{1,2}, A. Sissoko^{1,2}, S. Bamba¹, S. Diarra³, S. Mefoung^{1,2}, O. Traore^{1,2}, A. Yalcouye^{1,2}, A. Maiga², S. Diallo^{1,4}, L. CISSE¹, C. Guinto^{1,2}, G. Landoure^{1,2,3}, H3Africa consortium; ¹Faculté de Médecine et d'Odontostomatologie, USTTB, Bamako, Mali, Bamako, Mali, ²Service de Neurologie, Ctr. Hosp.ier Univ.ire du Point G, Bamako, Mali, Bamako, Mali, ³Pediatric Genomics Discovery Program, Yale Univ. Sch. of Med., New Haven, CT, New Haven, CT, ⁴Service de Neurologie, Ctr. Hosp.ier Univ.ire de Gabriel Touré, Bamako, Mali, Bamako, Mali

Abstract Body:

Background: Epilepsy is a chronic neurological condition particularly common in patients with intellectual and developmental disabilities (IDD). To date, several genes are reported to be associated with IDD. Variants in the *HECW2* gene were shown to cause this phenotype in European and North African patients. However, no case was reported in sub-Saharan African population. We report here the first *HECW2* variant associated with epilepsy and intellectual disability in a Malian family. **Objective:** To clinically characterize patients with IDD and identify the underlying genetic defects. **Materials and Methods:** Institutional ethical approval was obtained from the Faculty of Medicine and Dentistry, Bamako, Mali. Patients and available family members were enrolled after given a written consent. They were also carefully examined by a multidisciplinary team including neurologist, neurogeneticist. Blood chemistries including blood glucose, cell counts and ions; and EEG were performed in some patients. DNA was extracted from peripheral blood for whole exome sequencing. Putative variants were checked in all individuals for segregation analysis. We used several *in silico* prediction tools to detect deleteriousness including CADD. **Results:** Three siblings (two males and one female) from a non-consanguineous family were referred to our clinic for familial epilepsy. The age at diagnosis was 8,4 years and 16 months respectively. All patients had a focal tonic seizure with a psychomotor delay (neck tone, seat, crawl, gait) and absent speech. They received antiepileptic drugs which reduced the severity of the seizures. WES identified a homozygous missense variant **c.1856C>G; p. Pro619Arg** in *HECW2* gene in two affected individuals whereas parents were heterozygous for the variant. This variant was absent in various SNP databases including gnomAD and is predicted to be damaging by several tools (CADD = 23, Fathmm = 0.98; Consequence score = 6). Further investigations including functional analysis are ongoing. **Conclusion:** We report here the first case of IDD with epilepsy caused by mutation in the *HECW2* gene in a Malian family, expanding its genetic epidemiology. A large cohort study may help identify several other genetic variants for a better understanding of the pathophysiology of IDD.

Mendelian Phenotypes Posters - Wednesday

PB1739. A Novel Homozygous Variant in Homologous Recombination Repair Gene *ZSWIM7* Causes Azoospermia in Males and Primary Ovarian Insufficiency in Females

Authors:

S. Nawaz, Shah Hussain, Ihsan Khan , Nida Khan, Shabir Hussain, Imran Ullah, Khalid A. Fakhro, Wasim Ahmad; Sidra Med., Doha, Qatar

Abstract Body:

A Novel Homozygous Variant in Homologous Recombination Repair Gene *ZSWIM7* Causes Azoospermia in Males and Primary Ovarian Insufficiency in Females Infertility is a common, clinically heterogeneous reproductive disorder worldwide with a prevalence of about 15%. To date about 80 genes have been discovered to cause non-syndromic infertility, affecting males and females equally, though traditionally the genetic analysis of each group has been conducted separately. Here, we report the clinical and genetic characterization of a consanguineous family of Pakistani origin with multiple individuals, including male and female, affected with infertility. Males exhibited azoospermia whereas females had primary ovarian insufficiency. Whole exome sequencing revealed a missense variant (c.176C>T, p.(Ser59Leu)) in the *ZSWIM7* gene which functions in homologous recombination repair. The variant was found in a homozygous form in all affected males and females. To our knowledge, this is the first mutation in *ZSWIM7* to be shown to cause infertility in both sexes, pointing at the utility of large consanguineous families with multiple affected siblings to reveal joint mechanisms affecting human reproduction. **Keywords** Infertility, azoospermia, primary ovarian insufficiency, *ZSWIM7*

Mendelian Phenotypes Posters - Thursday

PB1740. A novel homozygous variant in *IBA57* causing multiple mitochondrial dysfunction syndrome type 3

Authors:

S. Lang, E. de Joya, P. Borjas-Mendoza; Univ. of Miami Miller Sch. of Med., Miami, FL

Abstract Body:

The multiple mitochondrial dysfunction syndromes (MMDS) are a rare cause of autosomal recessive leukodystrophy. Their molecular basis is rooted in impaired maturation of the cubane [4Fe-4S]-cluster containing proteins including complexes I-III of the respiratory chain and lipoic acid synthase, essential for critical modifications of several mitochondrial enzymes. Biallelic pathogenic variants in *IBA57* have been implicated in MMDS type 3, a phenotypically heterogeneous entity with presentations ranging from neonatal demise to benign neuropathy with survival into the sixth decade. In this report, we describe a novel homozygous variant in *IBA57* (NM_001010867.3): c.310G>T (p.Gly104Cys) causing MMDS type 3 in an infant.

A two-month-old female presented with one month history of progressive hypotonia, weakness, and upward gaze deviation. The patient was born full term from a consanguineous 27-year-old G1P0 mother and a 50-year-old father of Cuban origin. Prenatal history and perinatal period were unremarkable. Prior to the onset of symptoms, the patient achieved milestones appropriate for a one month old. Initial exam revealed significant head lag and upgaze deviation. Labs were notable for a lactic acidosis. MRI revealed widespread signal abnormalities involving subcortical and deep white matter tracts, optic nerves, as well as cystic signal abnormality within the medulla. Brain MR spectroscopy revealed a significant lactate peak. Multigene sequencing revealed a homozygous variant in *IBA57*, confirmed in trans on trio exome analysis. The infant showed worsening hypotonia and impending respiratory failure requiring intubation and ventilatory support by 3-month-old, ultimately expired.

This patient's findings of cystic leukodystrophy, lactate peak on MRS, progressive hypotonia, and developmental regression taken together with her homozygous missense variants in *IBA57* were strongly suggestive of MMDS 3. The variant, located within exon 1 of *IBA57*, is flanked by three pathogenic variants previously reported in MMDS 3 (OMIM ID: 615316.0007, 615316.0008, 615316.0010). In silico analysis supported a deleterious effect. Interestingly, the phenotypic outcomes of patients with any one of these four variants are highly variable, complicating the establishment of a clear genotype-phenotype correlation.

The MMDS present as a severe or fatal early onset encephalopathy. Clinical suspicion correlating with brain imaging and biochemical findings should prompt rapid confirmation through molecular analysis. This report describes novel homozygous variants in *IBA57* and expands the genotypic and phenotypic spectrum of MMDS 3.

Mendelian Phenotypes Posters - Wednesday

PB1741. A novel *LZTR1* variant in a Brazilian familial case with Noonan syndrome: comparative modeling and structural analysis

Authors:

N. Chaves Rabelo¹, D. Antunes², M. Gomes¹, I. de Oliveira Moraes¹, E. Caffarena³, J. Llerena⁴, S. Gonzalez¹; ¹Genomic Med. Lab./Reference Ctr. of Rare Diseases/Fernandes Figueira Inst./Fiocruz, Rio de Janeiro, Brazil, ²Functional Genomics and Bioinformatics/Oswaldo Cruz Inst./Fiocruz, Rio de Janeiro, Brazil, ³Computational Biophysics and Molecular Modeling Group, Scientific Computing Program/Fiocruz, Rio de Janeiro, Brazil, ⁴Med. genetics Dept./Reference Ctr. of Rare diseases/Fernandes Figueira Inst./Fiocruz, Rio de Janeiro, Brazil

Abstract Body:

Noonan syndrome (NS) is the most common genetic disorder in RASopathy group, that is characterized by short stature, congenital heart defects and facial features. Many genes involved in RAS/MAPK pathway activation have already been associated with NS, and the most frequently are: *PTPN11* (50%), *SOS1* (10-13%) and *LZTR1* (8%), among others. Pathogenic heterozygous and homozygous variants in the *LZTR1* gene have been associated with NS. The LZTR1 protein is composed by the Kelch and BTB domains that mediates protein-protein interactions with RAS family (HRAS, NRAS, and KRAS), acting as a negative regulator of the RAS-MAPK signaling pathway. A familial case of a father and daughter with suggestive features of RASopathy was investigated in a Reference Center of Rare Diseases in Brazil (RCRD/IFF/Fiocruz). Genomic DNA was obtained from the proband and her parents. The library construction was performed for the proband focused on 6,700 genes disease-associated (clinical exome sequencing), and the sequencing was performed in paired-end mode in the NextSeq 500 Platform. Data analysis was performed using a pipeline based on GATK Best Practices. Were considered variants in RASopathy-associated genes; non-synonymous or frameshift or in splicing sites; and, low-frequency variants (<0.1%). A missense not previously reported variant p.Pro225Leu in the *LZTR1* gene was found in heterozygosis, that has not yet been described in the population databases. Sanger sequencing revealed the same variant in her mildly affected father. This is a nonsynonymous substitution in a highly conserved region located in the Kelch domain of the LZTR1 protein and *in silico* predictors classified it as probably damaging. A symmetric structure of six blades characterizes this domain, and each of them is composed of four antiparallel β -strands, as previously described. Mutant and wild-type LZTR1 structural model comparison through Swiss-Model server showed differences in the loop that connects strand D (blade III) with strand A (blade IV), in addition to changes in blade III in the local secondary structure of the Kelch domain. Global changes are also observed, showing a decrease of approximately 1% in the β -sheet content (34.04% wild-type versus 32.97% mutant). These results reinforce the probable pathogenicity of the p.Pro225Leu variant, indicating the need for functional studies to assess its impact on RAS/MAPK pathway hyperactivation. Furthermore, the incorporation of NGS combined with computational analysis in the RCRD/IFF/Fiocruz have been contributing significantly to the clinical practice and the prognosis prediction of Rasopathy diseases.

Mendelian Phenotypes Posters - Thursday

PB1742. A novel *PUM1* heterozygous missense variant causing severe developmental delay and epilepsy from infancy with dysmorphic features.

Authors:

E. Sato¹, M. Fujimoto¹, T. Iwaki¹, Y. Nakamura¹, D. Ieda¹, A. Hattori^{1,2}, K. Kato^{1,3}, K. Narita⁴, H. Muramatsu⁴, Y. Okuno⁵, Y. Takahashi⁴, J. Natsume^{4,6}, S. Saitoh¹; ¹Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, ²Dept. of Pediatrics, Nagoya City Univ. East Med. Ctr., Nagoya, Japan, ³Sch. of Biochemistry, Faculty of Life Sci., BioMed. Sci. Building, Univ. of Bristol, Bristol, United Kingdom, ⁴Dept. of Pediatrics, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, ⁵Dept. of Virology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, ⁶Dept. of Dev.al Disability Med., Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan

Abstract Body:

[Background]PUM1 is an RNA-binding protein that acts as a transcriptional repressor of Ataxin1. Although specific missense variants in *PUM1* were known to cause adult-onset ataxia, recently it has been shown that complete loss of function in one *PUM1* allele causes severe early onset developmental delay, epilepsy and ataxia. However, there have been few reports of infantile onset *PUM1*-associated diseases, and the clinical features are yet to be elucidated. We herein describe an additional patient with a de novo heterozygous missense variant in *PUM1*. [Case report]The patient is a 27-year-old female born with a weight of 3230g at 39 weeks of an uneventful pregnancy. She could hold her neck at 5 months and sit at 9 months of age. She was diagnosed with West syndrome at 10 months of age. After that, she showed spasms, tonic seizures, myoclonic seizures, and other seizures despite multiple antiepileptic medications. Gradually she has shown severe developmental delay and does not speak any meaningful words. She shows some dysmorphic features, with hypertelorism, almond-shaped eyes, epicanthus, broad-nasal bridge, low-set ears, oligodontia with widely spaced teeth. Her head MRI at 17 years showed slight ventricular enlargement. We performed trio-based whole exome sequencing analysis and identified a de novo heterozygous variant in *PUM1* gene, NM_001020658: c.1760T>C, p.(V587A). This was a rare variant and was not registered in databases including gnomAD. CADD score was 30. Furthermore, her dysmorphic features were very similar to those in other reports. We considered that this newly identified variant caused severe developmental delay and epilepsy from infancy with dysmorphic features. [Conclusion]We identified a novel *PUM1* pathogenic variant in a patient with severe early onset developmental delay, and intractable epilepsy. Thus far only a recurrent variant of c.3439C>T (p.R1147W) has been reported to be associated with severe developmental delay and epilepsy. This case suggests that p. V587A, which is located outside RNA-binding Pumilio homology domain, has as much deleterious consequence as p.R1147W in PUM1 function.

Mendelian Phenotypes Posters - Wednesday

PB1743. A novel splice-site variant causes *MYH2*-associated myopathy in a large family.

Authors:

T. Cassini¹, M. Malicdan², E. Macnamara¹, T. Lehky¹, I. Horkayne-Szalaky³, Undiagnosed Diseases Network, Y. Huang¹, R. Jones³, R. godfrey⁴, L. Wolfe⁴, W. Gahl⁵, C. Toro²; ¹NIH, Bethesda, MD, ²NIH, Bethesda, MD, ³Defense Hlth.Agency, Silver Spring, MD, ⁴NIH - NHGRI, Bethesda, MD, ⁵NHGRI (NIH), Bethesda, MD

Abstract Body:

MYH2 encodes MyHCIIa, a myosin heavy chain found in fast type 2A fibers. Pathogenic variants in this gene have previously been implicated in dominant and recessive forms of myopathy typically manifesting as predominantly proximal muscle weakness and atrophy with external ophthalmoplegia. Onset is typically in adolescence or early adulthood and the course may be either slowly progressive or non-progressive. Additional features include facial weakness, tremors, and joint contractures at birth that sometimes resolve over time.

Three individuals evaluated by the Undiagnosed Diseases Program are part of a family in which four generations are affected by a slowly progressive, predominantly proximal myopathy in an autosomal dominant inheritance pattern. Affected individuals in this family lacked certain classic features of an *MYH2*-associated myopathy such as congenital contractures and ophthalmoplegia. Additional clinical evaluations included MRIs on all three individuals with muscle atrophy and fatty infiltration. A muscle biopsy was also performed on the proband and showed small type 2 myofibers involving some of the fascicles, and very tiny and markedly lost type 2 fibers in the adjacent fascicles accompanied by denervation atrophy and mild myopathic changes.

A novel variant, *MYH2* c.5673+1G>C, was detected in the proband. This variant was found to segregate with disease in five additional affected family members and was absent from population databases. Further studies confirmed that this variant affects splicing, resulting in novel transcripts. Reverse transcriptase PCR revealed the expected 402 base pair band and two additional bands in the proband and an affected relative as compared to an unaffected individual. Sanger sequencing of clones revealed the normal and three novel isoforms: an isoform with skipping of exon 39; an isoform with intronic retention of 83 nucleotides; and an isoform with partial deletion of exon 37 and 38. These data indicate that this family's *MYH2* variant is causative of their myopathy, adding to our understanding of the clinical and molecular characteristics of the disease.

Mendelian Phenotypes Posters - Thursday

PB1744. A novel XL gene for autism with psychiatric illness in adults

Authors:

H. Alakeel, W. Eyaid, A. AlZaben; Natl. Guard For Hlth.Affairs, Riyadh, Saudi Arabia

Abstract Body:

Mohammad; 7 years old male , known to have developmental delay, autism and ADHD. His WGS showed research finding, structural variant analysis demonestrated a hemizygous duplication (chrX:118541361-118569156 28kb) within SLC25A43, that is expected to affect the protien function. This variant was classified as 2P (likely affecting protein function).To date no OMIM phenotype has been associated to the pathogenic variants in the SLC25A43 gene. Quantative real time -PCR detected variable expression of SLC25A43 in almost all rat tissues examined. Expression was highest in brain coronal sections containing olfactory bulb and cerebral cortex, followed by adrenal gland and skeletal muscles. The segregation analysis showed that it is maternally inherited, the grandmother is not a carrier, but the grandfather is, and for that reason all his daughters are carriers. The grandfather and his elder brothers are reported to suffer from psychosis, maternal family history is as well significant for diverse mental manifestation (psychosis, ? Autism, and developmental delayes). We aim to furtherly study this gene function and it's pathogenicity to help this family. The young couple of Mohammad decided to stop reproducing till a definite diagnosis with prevention is made. He is there only child.

Mendelian Phenotypes Posters - Wednesday

PB1745*. A *Plod2* mutant mouse model of Bruck syndrome

Authors:

A. Kot¹, I. Duran², J. H. Martin¹, D. Wachtell¹, C. Chun¹, D. M. Hudson³, M. Weis³, D. R. Eyre³, J. Zieba¹, D. Krakow¹; ¹UCLA, Los Angeles, CA, ²Univ. of Malaga, Malaga, Spain, ³Univ. of Washington, Seattle, WA

Abstract Body:

Bruck syndrome is an autosomal recessive form of osteogenesis imperfecta (OI) caused by biallelic mutations in *PLOD2* or *FKBP10* and is characterized by joint contractures, bone fragility, short stature, and kyphoscoliosis. *PLOD2* encodes LH2, which hydroxylates type 1 collagen telopeptide lysines, an important step in proper collagen crosslink formation. *PLOD2* is alternatively spliced into short (LH2a) and long (LH2b) isoforms - LH2a is the dominant form early in development until LH2b predominates later. The role of *PLOD2* mutations in the pathogenesis of Bruck syndrome is not well understood and the global *Plod2* knockout mouse model is limited by early embryonic lethality. Therefore, we generated a novel *Plod2* mouse line to model a human mutation identified in two unrelated Bruck syndrome cases: *PLOD2* c.1559dupC, predicting a frameshift and loss of LH2b through nonsense mediated decay. We observed mice homozygous for the mutation in non-Mendelian ratios at E13.5 and E18.5 indicating variable, but early lethality. Homozygous mice were present at P0 but died immediately after birth. At E18.5, heterozygous mice were indistinguishable from WT. Homozygous mice displayed bilateral forelimb and hindlimb contractures, smaller body weights and crown-to-rump lengths, normal skeletal patterning but poorly mineralized calvaria, and loss of cervical spine curvature. Compared to WT, calvaria derived LH2b mRNA and protein were decreased in both heterozygous and homozygous mice. Type 1 collagen telopeptide lysine hydroxylation was reduced in homozygous but not in heterozygous bone. Von Kossa stained distal femurs showed homozygous mice had fewer and thinner trabeculae as well as smaller calcospherites. Femoral growth plate analysis revealed heterozygous and homozygous mice had decreased proliferative zone lengths indicating alterations in chondrocyte differentiation. Our model of *PLOD2* c.1559dupC survives longer than the existing global *Plod2* knockout model and demonstrates that loss of LH2b recapitulates important features of the Bruck syndrome phenotype including congenital contractures. The growth plate phenotype suggests *Plod2* plays a heretofore unknown role in chondrocyte development. This new model can be used to explore the role of LH2b in the developing skeleton and aid in understanding the molecular mechanisms underlying Bruck syndrome.

Mendelian Phenotypes Posters - Thursday

PB1746. A prospective, longitudinal observational natural history study of patients with NGLY1 Deficiency.

Authors:

S. Tong¹, P. Ventola^{2,3}, S. S. Dwight¹, W. F. Mueller¹, C. R. Stanclift¹, B. J. Beahm¹, M. Wilsey¹, K. J. Lee¹; ¹Grace Sci. Fndn., Menlo Park, CA, ²Yale Child Study Ctr., New Haven, CT, ³Cogstate, New Haven, CT

Abstract Body:

NGLY1 Deficiency is a debilitating, rare, autosomal recessive neurodevelopmental disorder with shortened life span and unmet need. It is caused by loss of *N*-glycanase 1 (NGLY1), a cytosolic enzyme that deglycosylates glycoproteins, and is characterized by global developmental delay, intellectual disability, hyperkinetic movement, elevated transaminases, (hypo)alacrima, and chronic polyneuropathy. A substrate biomarker, aspartylglucosamine (GNA), directly reflects loss of NGLY1 function. A prospective, longitudinal observational study was conducted to define NGLY1 Deficiency natural history. Participants (pts) (n=29; 15 on-site, 14 remote) with confirmed diagnosis of NGLY1 Deficiency were identified through the Grace Science Foundation and enrolled. The following were assessed: developmental status using validated, age-appropriate instruments; movement disorder; quality of life (QL); liver function tests; plasma GNA levels. On-site pt data are presented here; remote pt and additional longitudinal data will be presented. 15 on-site pts (11 unrelated, 2 sibling pairs; 7 girls, 8 boys), were followed for up to 3 yr (years). Mean age [range] was 12 [2.8-22] yr at entry and 6 [prenatal-14] yr at diagnosis. At baseline, caregivers were interviewed using the Vineland-3 (n= 14); adaptive skills were varied but generally low (ABC Standard Score mean 49 [23 - 88]). Age equivalents were globally delayed with mean [standard deviation (SD)]: Expressive 31months (m) [66m] or Receptive 38m [69m] Communication; Personal Care 22m [26m]; Fine 11m [7m] or Gross 9m [5m] Motor. On-site pts who completed the Mullen Scales of Early Learning (n=10) presented with marked developmental delay, mean DQ 11, range 1.4-18.8, with impairment across all domains. The age equivalent means [range] were: Visual Reception 16m, [1-27m]; Fine 11m [1-18m] or Gross 8m [1-16m] Motor; Receptive 12m [1-27m] or Expressive 7m [2-21m] Language. Caregivers for both onsite and remote pts reported poor QL: PedsQL (Total Score Mean [range] 50.48 [33.7-95.7]). Liver transaminase baseline mean values were elevated and improved slightly. Mean [SD] GNA was 115 ng/mL [28.4], range 57.8 - 189.0, which was 4.3-fold over related and 8.5-fold over unrelated unaffected controls. There was no significant correlation between GNA levels and age. Longitudinal GNA samples (n=10) showed ≤ 37 ng/mL change over 2-3 years (max - min concentration; mean change 23.0 ng/mL +/- 10.8ng/mL). In sum, pts with NGLY1 Deficiency have developmental delay, movement disorders, and inability for self-care. GNA biomarker levels are universally elevated and provide a potential tool for advancing a therapeutic intervention.

Mendelian Phenotypes Posters - Wednesday

PB1747. A recessive variant in *TFAM* causes mtDNA depletion associated with primary ovarian insufficiency, seizures, intellectual disability and hearing loss

Authors:

F. Ullah¹, W. Rauf², K. Khan³, S. Khan⁴, K. Bell⁵, V. Oliveira⁶, M. Tariq², S. Bakhshalizadeh⁵, P. Touraine⁷, A. Sinclair⁸, S. He⁹, E. Tucker¹⁰, S. Baig¹¹, E. Davis¹²; ¹Lurie children s Hosp. of Chicago, Chicago, IL, ²NIBGE Faislabad, Faislabad, Pakistan, ³Lurie Children s Hosp., Chicago, IL, ⁴Ann & Robert H. Lurie children's Hosp. of Chicago, Chicago, IL, ⁵Dept. of Pediatrics, Univ. of Melbourne, Melbourne, Australia, ⁶Dept. of Vet. Med., Faculty of Animal Sci. and Food Engineering, Univ. of São Paulo, Pirassununga, São Paulo, Brazil, ⁷Ctr. for Rare Endocrine and Gynecological Diseases, Sorbonne Université Pitié Salpêtrière Hosp., France, Paris, France, ⁸Murdoch Children s Res. Inst., Melbourne, Australia, ⁹BGI-Shenzhen, Shenzhen, China, ¹⁰MurdochChildren's Res. Inst., Royal Children's Hosp., Melbourne, Australia, ¹¹Pakistan Sci. Fndn., Islamabad, Pakistan, ¹²Lurie Children s Hosp. of Chicago, Chicago, IL

Abstract Body:

Mitochondrial disorders are individually rare but collectively common conditions that impact both pediatric and adult populations. A majority are caused by molecular defects in oxidative phosphorylation, failure of essential bioenergetic supply to mitochondria, and apoptosis, and therefore, it is not surprising that they are underscored by extensive genetic heterogeneity. Here, we present the genetic and functional analysis of four affected individuals from two independently analyzed consanguineous families of Pakistani origin with reproductive deficits accompanied by variable neurodevelopmental phenotypes and hearing loss. The females display primary ovarian insufficiency and incompletely penetrant seizures and hearing loss, while the male shows intellectual disability and abnormal sex hormone levels. We performed whole exome sequencing and homozygosity mapping and identified a recurrent recessive missense variant c.694C>T, p.Arg232Cys in *TFAM* that segregates with disease. TFAM (mitochondrial transcription factor A) is a component of the mitochondrial replisome machinery is involved in regulation of mitochondrial genome integrity and replication. There is a paucity of previously reported *in vivo* TFAM ablation data; *Tfam* knockout mice are lethal prior to weaning and one pathogenic p.Pro178Leu change has been shown to cause lethality during infancy in humans due to liver failure and intrauterine growth restriction. In patient derived fibroblasts from a p.Arg232Cys bearing case, we show significant depletion of mtDNA and significantly altered mitochondrial function and morphology. Moreover, we observed reduced nucleoid numbers with significant changes in nucleoid size or shape in fibroblasts from an affected individual compared to controls. Furthermore, we investigated the effect of *tfam* loss in zebrafish; homozygous knockout mutants recapitulate the mtDNA depletion and ovarian dysgenesis phenotypes observed in affected humans. Together, our genetic and functional data confirm that TFAM plays a pivotal role in gonad development, expands the repertoire of mitochondrial disease phenotypes, and exemplifies the increasingly prevalent phenomenon of variable phenotype penetrance and expressivity in Mendelian disorders.

Mendelian Phenotypes Posters - Thursday

PB1748*. A recurrent missense variant in *ITPR3* causes demyelinating Charcot-Marie-Tooth neuropathy.

Authors:

D. Beijer¹, M. F. Dohrn¹, A. Rebelo¹, C. Record², S. Feely³, M. Saporta¹, M. Reilly², S. Scherer⁴, Y-C. Lee⁵, M. Shy³, S. Zuchner¹; ¹Univ. of Miami, Miami, FL, ²UCL Queen Square Inst. of Neurology, London, United Kingdom, ³Univ. of Iowa, Iowa City, IA, ⁴Univ. of Pennsylvania, Philadelphia, PA, ⁵Taipei Veterans Gen. Hosp., Taipei, Taiwan

Abstract Body:

Charcot-Marie-Tooth disease is a rare neuromuscular disorder affecting the peripheral nervous system. The diagnostic yield in demyelinating CMT (CMT1) is typically ~80-95%, of which at least 60% is due to the *PMP22* gene duplication. The remainder of CMT1 is more genetically heterogeneous. We used whole exome sequencing (WES) and whole genome sequencing (WGS) data to investigate novel causal genes and mutations in a cohort of ~2,000 individuals with CMT disease submitted to the Genesis project. We identified a recurrent missense variant in *ITPR3*, a recently described CMT gene, in more than 16 individuals from seven different families. All families presented with slow nerve conduction velocities and an autosomal dominant or *de novo* inheritance, matching the diagnostic category of CMT1. Sanger sequencing confirmed the co-segregation of the CMT phenotype with the presence of the variant, including a four-generation family with multiple affected second-degree cousins, and a *de novo* inheritance in an isolated patient. *ITPR3* encodes IP₃R3 (inositol 1,4,5-trisphosphate receptor 3), which, like its paralogs *ITPR1* and *ITPR2*, is highly expressed in the nervous system. Based on protein modelling, this residue is located in the dimerization interface and could interfere with the dimerization process. We are currently testing this hypothesis using *in vitro* modelling. Here we show that a recurrent *ITPR3* missense mutation specifically causes a demyelinating Charcot-Marie-Tooth phenotype and could account for a relatively large proportion of unsolved CMT1 patients.

Mendelian Phenotypes Posters - Wednesday

PB1749. A single-center genetic analysis of oculopharyngeal myopathy

Authors:

T. Shimizu¹, H. Ishiura², M. Hara¹, S. Shibata³, A. Unuma^{1,4}, A. Kubota¹, K. Sakuishi^{5,6}, K. Inoue⁷, J. Goto^{8,6}, Y. Takahashi^{4,6}, J. Shimizu^{9,6}, S. Tsuji^{2,10}, T. Toda²; ¹The Univ. of Tokyo Hosp., Bunkyo-ku, Japan, ²The Univ. of Tokyo, Tokyo, Japan, ³Saitama Med. Ctr., Jichi Med. Univ., Tokyo, Japan, ⁴Natl. Ctr. of Neurology and Psychiatry, Kodaira, Tokyo, Japan, ⁵Teikyo Univ. Chiba Med. Ctr., Chiba, Japan, ⁶The Univ. of Tokyo Hosp., Bunkyo-ku, Tokyo, Japan, ⁷Fujieda Heise Mem. Hosp., Sizuoka, Japan, ⁸Intl. Univ. of Hlth.and Welfare Ichikawa Hosp., Ichikawa, Chiba, Japan, ⁹Tokyo Univ. of Technology, Tokyo, Japan, ¹⁰Inst. of Med. Genomics, Intl. Univ. of Hlth.and Welfare, Chiba, Japan

Abstract Body:

Background: Oculopharyngodistal myopathy (OPDM) is a distal myopathy characterized by ocular, pharyngeal, and facial muscle weakness with an autosomal dominant mode of inheritance. Recently, we have identified heterozygous CGG repeat expansions in *low-density lipoprotein receptor related protein 12 (LRP12)* that cause OPDM, and the disease was designated as OPDM type 1 (OPDM1). Thereafter, CGG repeat expansions in *NOTCH2NLC*, *GIPC1*, and *RILPL1* have been identified as causes of other types of OPDM. To further delineate molecular epidemiology, we performed genetic analysis of patients with OPDM. **Methods:** This study is a single-center case series of OPDM including ten patients from seven families. Every family had been shown to have rimmed vacuoles in the muscle biopsy and excluded to have GCN repeat expansions in *PABPN1*. Repeat-primed polymerase chain reaction and Southern blot analyses were performed to confirm the CGG repeat expansions in *LRP12*, *GIPC1*, and *NOTCH2NLC*. Whole genome sequencing (WGS) was performed in one patient. **Results:** Seven patients from five families and two patients from one family were identified as having CGG repeat expansions in *LRP12* (OPDM1) and *GIPC1* (OPDM2), respectively. In the remaining one patient, CGG repeat expansion was not found in these genes. In three OPDM1 patients, multiple expanded alleles, probably reflecting somatic instability, were found. In one patient without CGG repeat expansions in these genes, WGS failed to identify expanded CGG repeats. **Discussion:** In this single-center study of OPDM, OPDM1 is most common in Japan. In one patient, WGS did not reveal expanded CGG repeats, further suggesting pathomechanistic heterogeneity.

Mendelian Phenotypes Posters - Thursday

PB1750. A substantial proportion of high myopia is caused by mutations in *CACNA1F* (XL CSNB)

Authors:

L. Hoefsloot¹, P. A. T. Heutinck², M. van Tienhoven¹, A. A. H. Thiadens², A. E. G. Haarman², C. C. W. Klaver^{2,3,4,5}, V. Verhoeven^{1,2}; ¹Dept Clinical Genetics, Erasmus MC, Rotterdam, Netherlands, ²Dept Ophthalmology, Erasmus MC, Rotterdam, Netherlands, ³Dept Epidemiology, Erasmus MC, Rotterdam, Netherlands, ⁴Dept Ophthalmology, Radboud UMC, Nijmegen, Netherlands, ⁵Inst. of Molecular and Clinical Ophthalmology, Basel, Switzerland

Abstract Body:

CACNA1F is involved in the neurotransmission between the rod photoreceptors and the bipolar cells in the retina, and has been mostly associated with (in)complete congenital stationary night blindness (CSNB) (CSNB2A, OMIM 300071), but also with X-Linked (XL) Åland eye disease (OMIM 300600) and XL cone-rod dystrophy (CRD) (CORDX3, OMIM 300476). During the last six years, we have diagnostically screened >750 patients with various eye disorders (mainly inherited retinal dystrophy, age related macular degeneration, opticopathy, high myopia) using Whole Exome Sequencing with filtering for a panel of ~500 genes associated with vision disorders. In this cohort, 220 patients with high myopia were included in which we identified 11 males with mutations in *CACNA1F* (5%). Mean age of these patients was 19.7 years (SD 23.5). Early onset high myopia was the first symptom. At time of first investigation, the mean refractive error was -12.1 diopters (SD 3.36; range -6 to -24). Other reported symptoms at time of first investigation were night blindness (n=4) and decreased best corrected visual acuity (n=11), with a mean of 0.45 (SD 0.11). In young children, these symptoms can be difficult to notice. Eventually an electroretinogram (ERG) was conducted in 6 patients. This led to the diagnosis of CSNB (n=4) and CRD (n=1). At young age an ERG cannot always be made for practical reasons. Our data suggest that in young males with early onset high myopia *CACNA1F* should be investigated, and this should be part of the screening strategy.

Mendelian Phenotypes Posters - Wednesday

PB1751. A Zebrafish Model of Congenital Muscular Dystrophy Caused by POMT1 Loss of Function

Authors:

M. Manzini, B. F. Karas, K. R. Terez, B. M. Gural, N. Battula, K. Flannery; Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ

Abstract Body:

Dystroglycanopathies are a group of rare autosomal recessive congenital muscular dystrophies (CMDs) that present at birth and severely affect the brain, eyes, and muscles. While these disorders are highly genetically and clinically heterogeneous, the function of most known dystroglycanopathy genes converges on defects of glycosylation of the transmembrane glycoprotein alpha-dystroglycan (alpha-DG). Without proper alpha-DG glycosylation, transmembrane communication between the cell and extracellular matrix (ECM) is compromised in multiple tissues. In fact, most dystroglycanopathy genes are glycosyltransferases participating in the assembly of a specific glycan chain on alpha-DG involved in binding ECM components. The initial step in glycosylation is the addition of an O-linked mannose via Protein O-Mannosyltransferase 1 (POMT1) to alpha-DG and *POMT1* mutations in patients frequently manifest as the most severe form of CMD with early lethality, Walker-Warburg Syndrome (WWS). In mice, only conditional *Pomt1* knockouts (KOs) could be generated, since *Pomt1* loss of function leads to early embryonic lethality. In this study, we characterized a *potm1*-deficient zebrafish model that could provide new insight into the mechanisms of disease and treatment interventions. The zebrafish strain carries a premature stop codon leading to complete loss of POMT1 and loss of alpha-DG glycosylation. Our data show that loss of *potm1* leads to early lethality starting at 30 days post fertilization (dpf), small size, impaired mobility, muscle disease, and retinal defects. However, these phenotypes are delayed compared to dystroglycan (*dag1*) KO in fish. Since *potm1* is critical for early embryonic development and mRNA is provided by the mother in the embryo's yolk, we asked whether KO embryos and larvae obtained from *potm1* KO mothers would present more severe phenotypes. We found that zebrafish larvae from KO X heterozygote crosses begin to die at 10 dpf similar to *dag1* KO zebrafish and show severe mobility impairment starting at 5 dpf. Collectively, these data suggest the *potm1* mutations in the zebrafish lead to disease phenotypes consistent with the human presentation and could be used to test novel therapeutic approaches.

Mendelian Phenotypes Posters - Thursday

PB1752. Allele-specific inactivation of epidermolysis bullosa simplex mutations using CRISPR-Cas9 and spraying of the corrected cell suspension onto the wounds

Authors:

M. Bchetnia¹, J. Powell², C. Morin³, C. McCuaig², A. Dupérée³, L. Germain⁴, J-P. Tremblay⁵, C. Laprise¹; ¹Université du Québec à Chicoutimi (UQAC), Département des Sci.s fondamentales, Saguenay, QC, Canada, Chicoutimi, QC, Canada, ²Hôpital Ste-Justine, Montréal, QC, Canada, Montreal, QC, Canada, ³Ctr. intégré universitaire de santé et des services sociaux du Saguenay–Lac-St-Jean, Saguenay, QC, Canada, Chicoutimi, QC, Canada, ⁴Ctr. de recherche en organogénèse expérimentale de l'Université Laval/LOEX, Québec, QC, Canada, Québec, QC, Canada, ⁵Ctr. de recherche du Ctr. Hosp.ier universitaire (CHU) de Québec, Université Laval, QC, Canada, Québec, QC, Canada

Abstract Body:

Background: Epidermolysis bullosa simplex (EBS) is a rare mechanobullous disease caused by dominant-negative mutations in either keratin 5 (*KRT5*) or keratin 14 (*KRT14*) genes responsible for skin integrity via the building the keratin filament network in the epidermis. Until now, there is no cure for EBS and the care is primarily palliative. The discovery of the clustered regularly interspaced short palindromic repeat (CRISPR-Cas9) system, raised hope for the treatment of EBS and many other autosomal dominant diseases by specific disruption of the mutant allele. Moreover, the recent, clinically used, ReCell Spray-On Skin device seems to be a rapid autologous cell processing and delivery system where the corrected cell suspension could be sprayed onto the lesions in a non-invasive way. The regenerative nature of these skin cells is intended to promote rapid wound healing and the growth of healthy skin. **Objectives:** We aim to disrupt the mutant allele for the heterozygous EBS pathogenic variation c.449T>C (p.Leu150Pro) within *KRT5*. This mutation generates, naturally, a novel protospacer-adjacent motif (PAM) for the endonuclease *Streptococcus pyogenes Cas9* (*SpCas9*). The corrected cells will be delivered on the patient wounds by using the ReCell Spray-On Skin device. **Methods:** We designed a single guide RNA (sgRNA) that guides the Cas9 to introduce a DNA cleavage of the mutant allele in patient's keratinocytes. Transfected cells were subsequently single-cell cloned and analysed by deep sequencing at the DNA and RNA level. *KRT5* and *KRT14* expression were quantified in edited and non-edited cells and keratin intermediate filaments stability was assessed. **Results:** Single cell cloning showed stringent mutant allele specific knockout in some clones. An absence of synthesis of mutant transcript was further confirmed indicating permanent mutant allele-specific inactivation. Edited EBS patient keratinocytes have a lower amount of K5 and K14 compared to non-edited EBS cells but the protein product of the unaffected wild type allele should be sufficient for normal cellular function. No disturbance of cellular properties was observed. The cell suspension delivery protocol optimisation is in progress. **Conclusion:** This study is the first description of allele specific CRISPR-Cas9 gene inactivation at a novel PAM created by one EBS causing heterozygous pathogenic variation. The next step will be to use the ReCell Spray-On Skin device to deliver the corrected cells in the EBS lesions by a rapid and non-invasive method.

Mendelian Phenotypes Posters - Wednesday

PB1753. Alteration of mitochondrial proteostasis in Costello syndrome

Authors:

D. Lacombe^{1,2,3}, L. Dard^{4,3}, C. Hubert^{4,3}, N. Amoedo^{4,3}, R. Rossignol^{4,3}; ¹CHU BORDEAUX, BORDEAUX, France, ²BORDEAUX Univ., Bordeaux, France, ³INSERM U1211, Bordeaux, France, ⁴BORDEAUX Univ., BORDEAUX, France

Abstract Body:

Normal 0 21 false false false FR JA X-NONE /* Style Definitions */ table.MsoNormalTable {mso-style-name:"Tableau Normal"; mso-tstyle-rowband-size:0; mso-tstyle-colband-size:0; mso-style-noshow:yes; mso-style-priority:99; mso-style-parent:""; mso-padding-alt:0cm 5.4pt 0cm 5.4pt; mso-para-margin:0cm; mso-para-margin-bottom:.0001pt; mso-pagination:widow-orphan; font-size:12.0pt; font-family:"Times New Roman",serif; mso-bidi-font-family:"Times New Roman"; mso-bidi-theme-font:minor-bidi;} Germline mutations that activate genes in the canonical RAS/MAPK signaling pathway are responsible for rare human developmental disorders known as RASopathies. Here, we analyzed the molecular determinants of Costello syndrome (CS) using a mouse model expressing HRAS^{G12S}, patient skin fibroblasts, hiPSC-derived human cardiomyocytes, a HRAS^{G12V} zebrafish model and human fibroblasts expressing lentiviral constructs carrying HRAS^{G12S} or HRAS^{G12A} mutations. The findings revealed alteration of mitochondrial proteostasis and bioenergetics, and defective oxidative phosphorylation in the heart and skeletal muscle of Costello mice that were also found in the cell models of the disease. The underpinning mechanisms involved the miR-221*-dependent inhibition of AMPK α 2 expression and the concomitant alteration of LKB1 activation by mutant forms of HRAS, leading to alteration of mitochondrial turnover and bioenergetics. Pharmacological rescue of mitochondrial proteostasis restored organelle bioenergetics in HRAS^{G12A/S} cell models, reduced heart mass in CS mice and reduced the occurrence of developmental defects in the CS zebrafish model.

Mendelian Phenotypes Posters - Thursday

PB1754. An initiative for the diagnosis and study of rare and undiagnosed diseases in Mexico

Authors:

C. Gonzaga-Jauregui; Univ. Natl. Autónoma de México (UNAM), Juriquilla, Mexico

Abstract Body:

About 7000 rare diseases have been documented to date, the majority of which have a genetic cause or component and therefore can be molecularly diagnosed. The implementation of genomic sequencing of patients with suspected genetic disorders has revolutionized the investigation, diagnosis, treatment, and care of patients with rare diseases. Nevertheless, these technologies are mostly available to patients from higher income countries, exacerbating the health disparities suffered by patients living with rare and undiagnosed diseases in low- and middle-income countries where these technologies are difficult to access and too expensive.

Mexico has a population of 126 million people and it is the 17th economy in the world. However, the majority of the estimated 10 million people that live with a rare disease in the country do not have an accurate molecular diagnosis due to lack of access to genetic testing methods. Furthermore, an estimated 2 to 3 million of patients living with rare diseases are part of indigenous or underserved communities with poor access to healthcare services.

In 2022, we have launched a registry of patients with rare and low prevalence diseases in Mexico to obtain more accurate information on the number of patients, diseases, and disease prevalence in this population, and the challenges faced by these patients. The first analysis of the data collected from the registry reveals a great need for information and access to diagnostic options among patients and families living with rare diseases. The majority of patients registered so far are affected by neurodevelopmental or connective tissue disorders; other rare disorders have also been reported. Only 20% of registered patients have received a molecular diagnosis for their disorder. Concurrently, we have initiated a research program to investigate the molecular causes of diseases of rare and undiagnosed diseases in Mexico. We are performing genomic sequencing and analyses to identify the causes of rare and undiagnosed genetic disorders in this genomically underrepresented population. We present some initial findings from these efforts, including a novel candidate disease gene identified through the study of Mexican patients. We expect that these efforts will reveal novel disease associated variation present in Mexican genomes not previously reported in genomic databases, identify new candidate disease genes, and provide insights on medically relevant variation in the Mexican population that may serve for population screening and implementation of precision medicine in a middle-income country setting.

Mendelian Phenotypes Posters - Wednesday

PB1755. Angelman syndrome with mosaic paternal uniparental disomy caused by mitotic nondisjunction.

Authors:

M. Fujimoto¹, Y. Nakamura¹, T. Iwaki¹, E. Sato¹, D. Ieda¹, A. Hattori¹, A. Shiraki^{2,3}, S. Mizuno⁴, S. Saitoh¹; ¹Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, ²Dept. of Child Neurology, Toyota Municipal Child Dev. Ctr. Nozomi Clinic, Toyota, Japan, ³Dept. of Pediatrics, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, Nagoya, Japan, ⁴Dept. of Pediatrics, Central Hosp., Aichi Human Service Ctr., Kasugai, Aichi, Japan

Abstract Body:

Angelman syndrome (AS) is a severe neurodevelopmental disorder and caused by the functional absence of the maternal copy of the *ubiquitin-protein ligase E3A (UBE3A)* gene. Paternal uniparental disomy of chromosome 15 (UPD(15)pat) leads to absence of maternal *UBE3A* expression and is known as one of the molecular pathomechanisms of AS. Approximately 5% of AS is caused by UPD(15)pat, most of which are thought to be caused by monosomy rescue. However, few attention has been focused on how the UPD(15)pat would occur in each patient. Here, we demonstrated the mitotic nondisjunction mechanism that caused UPD(15)pat in a patient with AS. A 6-year-old boy presented neurodevelopmental disorder and distinctive physical findings that were in line with AS features. DNA methylation screening of 15q11-q13 showed a paternal band and a faint maternal band, suggestive of mosaic status. By trio-based microsatellite analysis using their blood, we confirmed the mosaic status, including a large proportion of UPD(15)pat cells and a small proportion of biparental-origin cells. Single nucleotide polymorphism (SNP) microarray revealed that the isodisomic region spanned the entire length of chromosome 15. Overall, these results suggest that the UPD(15)pat in the current patient did not occur by a pre-zygotic monosomy rescue mechanism but by post-zygotic mitotic nondisjunction mechanism. This report suggests that some AS patients with UPD(15)pat of entire isodisomy that thought to be caused by monosomy rescue may be caused by mitotic nondisjunction.

Mendelian Phenotypes Posters - Thursday

PB1756. *ARG1* variants in Arginase 1 Deficiency: genetic characterization of participants in the pegzilarginase clinical trials.

Authors:

M. McNutt¹, G. Bubb², L. Neuman², R. Zori³; ¹UT Southwestern Med. Ctr., Dallas, TX, ²Aeglea BioTherapeutics, Inc., Austin, TX, ³Univ. of Florida, Gainesville, FL

Abstract Body:

Arginase 1 Deficiency (ARG1-D; OMIM 207800) is a rare autosomal recessive metabolic disorder caused by homozygous or compound heterozygous mutations in *ARG1* (60813; chromosome 6q23) resulting in loss of arginase 1 activity and pathologic accumulation of arginine. ARG1-D is characterized by the hallmark of hyperargininemia and manifestations including developmental delay, intellectual disability, and prominent, progressive lower-limb spasticity. ARG1-D is distinct from other urea cycle disorders as symptomatic hyperammonemia is not uniformly present and not thought to be a key disease driver. The typical clinical course includes unremarkable infancy followed by development of progressive neurologic manifestations that begin in early childhood; however, for reasons currently unknown, there is a significant variability in age of onset, sequence and rate of progression, and severity, even among patients with the same *ARG1* variant(s). More than 60 *ARG1* variants have been reported to date, and since recent genetic prevalence estimates are higher than originally thought, additional variants may exist. The Phase 1/2 and Phase 3 clinical trials of pegzilarginase, which demonstrated significant reduction of plasma arginine to or below the guideline-recommended level of 200 $\mu\text{mol/L}$, provided the opportunity to investigate *ARG1* variants in 2 robustly characterized cohorts of patients with ARG1-D. Disease manifestations were heterogenous in nature and degree, as is typical of ARG1-D, and all patients had characteristic elevated plasma arginine levels despite dietary protein restriction. Genotyping data were available for 47 of 48 trial participants and identified 37 unique variants (homozygous, n=37 patients; compound heterozygous, n=10 patients). The most frequent variant was C.466-1G>C, identified in 9 patients (6 homozygous), followed by C.314_345delins20 (homozygous, n=3, heterozygous, n=1), C.61C>T (homozygous, n=2; heterozygous, n=2), and C23T>A (n=3, all homozygous). A total of 18 variants (in 29 patients) have been described in the literature suggesting the pegzilarginase study patients are representative of the broader ARG1-D population. The remaining 19 variants, including C.314_345delins20, appear to be novel. Consistent with the literature, no clear genotype/phenotype association was evident in this large cohort. Future research into the impact of *ARG1* variants, both known and novel, is warranted and may increase understanding of this heterogenous and debilitating disease. Importantly, the arginine-lowering effect of pegzilarginase was consistent irrespective of patients' genetic variant(s) and clinical presentation.

Mendelian Phenotypes Posters - Wednesday

PB1757. *ARHGAP32*, encoding the Rho GTPase Activating Protein 32, is a novel candidate gene involving in autosomal dominant neurodevelopmental disorder spectrum

Authors:

Y. Cao^{1,2,3}, A. Kwan¹, Y. ZHENG¹, M. Chau^{1,3}, Z. Dong^{1,3}, S. Chong², Y. Kwok¹, P. Liu⁴, Y. Yang⁴, R. Choy^{1,3}, H. Dai⁴; ¹Dept. of Obstetrics and Gynaecology, Prince of Wales Hosp., The Chinese Univ. of Hong Kong, Hong Kong, China, ²Dept. of Paediatrics, Prince of Wales Hosp., The Chinese Univ. of Hong Kong, Hong Kong, China, ³Hong Kong Hub of Paediatric Excellence, The Chinese Univ. of Hong Kong, Hong Kong, China, ⁴Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX

Abstract Body:

Rho GTPase-activating proteins (GAPs) regulate the Rho GTPases which play essential roles in regulating neuronal morphology and function. Currently more than 70 types of Rho GTPase Activating Protein have been discovered. Recent evidence suggests that dysfunction of Rho GTPase signaling contributes substantially to the pathogenesis of neurodevelopmental and neuropsychiatric disorders. Studies in large cohorts of subjects with autistic spectrum disorders or intellectual disability have identified de novo variants in *ARHGAP32* gene, suggesting that this gene is involved in the development of such disorders. The *ARHGAP32* isoform including PX, SH3 domain deficient mice mimic autistic feature of patients while its ASD-like behavioral problems are ameliorated by enhancing inhibitory synaptic transmission with a GABAAR agonist. In this study we would further provide evidence supporting *ARHGAP32* gene-disease validity. Total 8 patients were collected in this study including four unrelated patients harboring a heterozygous de novo variant in *ARHGAP32* gene through worldwide collaboration and four previously published patients from large cohort studies. Except a prenatal case, all case showed neurodevelopmental or neuropsychiatric disorders commonly presenting autism spectrum disorders(43%), ADHD(43%), moderate intellectual disabilities(29%), brain image abnormalities(43%). One postnatal case presented macrocephaly while the prenatal case had head circumference higher than 97th percentile. Its pLi Score as 1 and LOEUF as 0.25 indicated its intolerance to the loss of function. We demonstrated that 4 de novo truncating variants in our patients are in the latter half of *ARHGAP32* Protein. However, we did not observe a clear genotype-phenotype correlation between the types and locations of these 8 variants and variability of clinical features. We suggest *ARHGAP32* gene is involved in development of the autism spectrum disorder with variable additional neurodevelopmental abnormalities.

Mendelian Phenotypes Posters - Thursday

PB1758. ATP2C1 as a candidate for Interstitial Cystitis/Bladder Pain Syndrome

Authors:

E. Estrella¹, S. Rockowitz¹, M. Thorne², P. Smith¹, J. Petit¹, V. Zehnder¹, R. Yu¹, S. Bauer¹, C. Berde¹, P. Agrawal³, A. Beggs¹, A. Gharavi⁴, L. Kunkel¹, C. Brownstein³; ¹Boston Children's Hosp., Boston, MA, ²Biogen, Cambridge, MA, ³Boston Children's Hosp., Boston, MA, ⁴Columbia Univ, New York, NY

Abstract Body:

Interstitial Cystitis/Bladder Pain Syndrome (IC/BPS) is a chronic pain disorder causing symptoms of urinary frequency, urgency and bladder discomfort or pain. Although this condition affects a large population, little is known about its etiology. We performed genetic analyses of whole exome sequencing on a total of 109 individuals with IC/BPS. One family had a previously reported *SIX5* mutation (p.A158T), consistent with Branchiootorenal syndrome 2 (BOR2). A likely pathogenic heterozygous variant in *ATP2A2* (p.E79K) was identified in another proband, indicating possible Darier-White disease. Two private heterozygous variants were identified in *ATP2C1* (p.E786D (VUS/Likely Pathogenic) and p.T330S (Likely Pathogenic)), indicative of Hailey-Hailey Disease. Both Darier-White and Hailey-Hailey have specific therapies. SKAT analysis found a trend towards increased burden of rare *ATP2C1* variants in the IC/BPS cases vs a control cohort (p=0.03, OR=6.76), though did not survive Bonferroni correction. Our data suggest that some individuals with IC/BPS may have unrecognized Mendelian syndromes. Comprehensive phenotyping and genotyping aids in understanding the range of diagnoses in our population-based IC/BPS cohort. Conversely, *ATP2C1*, *ATP2A2*, and *SIX5* could be candidate genes for IC/BPS. Further evaluation with larger numbers is needed. Genetically screening individuals with IC/BPS may be useful in diagnosing and treating this painful disorder due to its heterogenous nature.

Mendelian Phenotypes Posters - Wednesday

PB1759. Atypical 260 kb deletion on distal 22q11.22 involving *TOP3B* shows the significance of this gene in autism spectrum disorder

Authors:

D. Evans¹, Y. Qiao², K. Calli³, S. Martel⁴, M. Hrynchak⁵, S. Lewis⁶; ¹Univ. of British Columbia, Vancouver BC, BC, Canada, ²Univ. of British Columbia, Vancouver, BC, Canada, ³Univ British Columbia, Vancouver, BC, Canada, ⁴Med. Genetics, Univ. of British Columbia, Vancouver BC, BC, Canada, ⁵Royal Columbian Hosp, New Westminster, BC, Canada, ⁶BC Children's & Women's Hlth.Ctr, The Univ. of British Columbia, Vancouver, BC, Canada

Abstract Body:

In Canada the prevalence of autism spectrum disorder (ASD) is 1 in 66 children. ASD is a heterogeneous group of neurodevelopmental disorders defined by deficits in social communication and interaction, as well as restrictive, repetitive patterns of behavior. There are more than 40 recurrent CNVs that have been implicated in ASD susceptibility. The 22q11.21 deletion syndromes result in an array of phenotypes such as velocardiofacial syndrome (OMIM 192430) and DiGeorge syndrome (OMIM 188400). Interestingly, distal deletions at the recurrent 22q11.2 locus (OMIM# 611867) comprise a distinct disorder separate from DiGeorge and velocardiofacial syndrome, which has been the topic of several candidate gene studies. Notably, *TOP3B* and *IGLV2-14* were previously described as putative candidate genes within a 22q11.22 critical interval. Here, we investigate and characterize a 13-year-old male with psychometrically confirmed ASD and other co-morbidities such as intellectual disability, ADHD, specific learning disabilities and growth delays in the absence of any craniofacial dysmorphisms. We employed whole genome sequencing of the parent-offspring trio and did not identify any candidate pathogenic variants. Our CNV analysis identified a 260Kb deletion at 22q11.2 in the proband: arr[Hg38]22q11.21-q11.22(21960001-22220000)X1. This deletion was confirmed using FISH and involves only the *TOP3B* gene. Therefore, our study narrows the critical interval to *TOP3B*, extensively characterizes the phenotype in this patient with ASD, and demonstrates *TOP3B* is a strong candidate gene for ASD.

Mendelian Phenotypes Posters - Thursday

PB1760. Atypical molecular findings in patients with capillary malformations.

Authors:

C. Montano¹, E. Wohler², S. Yeom³, B. Cohen², C. Weiss², T. Garg², A. Hammill⁴, A. Comi³, N. Sobreira²; ¹NHGRI/NIH, Bethesda, MD, ²Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ³Kennedy Krieger Res. Inst., Baltimore, MD, ⁴Cincinnati Children's, Cincinnati, OH

Abstract Body:

Capillary malformations (CMs), commonly known as port-wine stains, are partially blanchable erythematous patches. Sturge-Weber syndrome (SWS), the most well-known CM, is characterized by facial port-wine stains involving the upper quadrant of the face with ipsilateral ocular and leptomeningeal anomalies. CMs occurring as part of SWS or in isolation are sporadic events frequently associated with somatic gain-of-function variants in *GNAQ*, leading to constitutive activation of the RAS/MAPK pathway. CMs also can occur in combination with other vascular anomalies and limb overgrowth, such as in Klippel-Trenaunay syndrome (KTS) and in Parkes Weber Syndrome (associated with arteriovenous malformations and loss of function variants in *RASA1*). Hemangiomas, vascular tumors that mimic CMs, can occur in PHACES syndrome (posterior fossa or other CNS malformations, facial hemangioma, arterial anomalies, cardiac defects, eye anomalies, and sternal defects). The molecular basis for KTS and PHACES syndrome is not known. Here, we describe the phenotypic features and the germline and somatic exome sequencing findings of four patients with CMs: two with clinical diagnosis of SWS, one with a clinical diagnosis of KTS and one with a clinical diagnosis of PWS. Of the two patients with clinical diagnosis of SWS, one had a 580kb germline deletion on chromosome 2 that included *COL3A1*, *COL5A2* and *SLC40A1*; the other had a paternally inherited pathogenic variant in *KIT*. The patient with a clinical diagnosis of KTS had a somatic pathogenic variant in *PIK3CA* and was diagnosed with megalencephaly-capillary malformation (MCAP). The patient with a clinical diagnosis of PWS had a somatic pathogenic variant in *KRAS*. Understanding the genetic drivers of vascular anomalies is critical for classification, management, and the development of targeted therapeutics. Our molecular findings have implications for prognosis, drug trial eligibility, and clinical care. Based on our experience, germline exome sequencing, followed by somatic exome sequencing if germline testing is negative, is indicated for molecular evaluation of these patients to ensure the identification of germline and somatic single nucleotide variants, indels and copy number variations.

Mendelian Phenotypes Posters - Thursday

PB1761. Bi-allelic ACBD6 variants lead to a neurodevelopmental syndrome with progressive complex movement disorders

Authors:

R. Maroofian; The UCL Queen Square Inst. of Neurology, london, United Kingdom

Abstract Body:

The acyl-CoA-binding domain-containing protein 6 (ACBD6) is ubiquitously expressed, plays a role in the acylation of lipids and proteins, and regulates the N-myristoylation of proteins via N-myristoyltransferase enzymes (NMTs). However, its precise function in cells is still unclear, as is the consequence of ACBD6 defects on human pathophysiology. Utilizing exome sequencing and extensive international data sharing efforts, we describe 35 affected individuals from 22 unrelated families with bi-allelic pathogenic variants in ACBD6. The affected individuals, with ages ranging from 1 to 50 years old, present with moderate-to-severe global developmental delay/intellectual disability combined with complex and progressive movement disorders (mainly dystonia phenotype), epilepsy, behavioral abnormalities, facial dysmorphism, and midline brain malformations. Unlike ACBD5, ACBD6 did not show a peroxisomal localisation and ACBD6-deficiency was not associated with altered peroxisomal parameters in patient fibroblasts. No significant difference in N-myristoylation of the 57 detected co-translationally N-myristoylated proteins between healthy and patient-derived fibroblasts was observed. Functional studies in a zebrafish knockout generated by CRISPR/Cas9 recapitulated many clinical phenotypes reported in affected individuals. The present study provides evidence that defective ACBD6 causes a distinct neurodevelopmental syndrome accompanied by complex and progressive cognitive and movement disorders.

Mendelian Phenotypes Posters - Wednesday

PB1762*. Biallelic *DMD* variants in a mildly-affected female without a Duchenne or Becker phenotype

Authors:

X-R. Yang¹, C. Hahn², R. Lamont¹, S. Ashtiani¹, M. Innes¹; ¹Dept. of Med. Genetics, Univ. of Calgary, Calgary, AB, Canada, ²Dept. of Clinical NeuroSci.s, Univ. of Calgary, Calgary, AB, Canada

Abstract Body:

Very few females with biallelic *DMD* variants have been reported in the literature to date. While hemizygous males tend to manifest with Duchenne or Becker muscular dystrophy (DMD or BMD), female carriers are most often asymptomatic, or exhibit mild symptoms secondary to skewed X-inactivation. Of the few reported cases of biallelic females, their clinical presentations seem consistent with those seen in males with DMD or BMD (Quan et al. 1997, Fujii et al. 2009, Soltanzadeh et al. 2010, Takeshita et al. 2017). However, here we report a biallelic female with unexpectedly mild, late-onset symptoms.

We report a 74 y.o. female who had two sons with a clinical diagnosis of DMD. Both passed away in their teenage years prior to the introduction of molecular testing. Once molecular testing became available in the early 2000s, our proband underwent MLPA testing for deletions and duplications in *DMD*, confirming carrier status for an in-frame duplication of exons 10-41 of the *DMD* gene as the likely explanation for her two sons with DMD. She had no clinical symptoms at this time. Her healthy daughter subsequently had carrier testing and was found not to have inherited the duplication. However, that daughter went on to have an affected son with DMD. Further testing revealed a c.5475del, p.(Glu1826Asnfs*22) pathogenic frameshift variant in *DMD* - present in the affected grandson, his mother, and our proband.

Our proband went on to develop slowly progressive proximal > distal weakness in her 60s. Her CK levels ranged from 200-1400 U/L. A muscle biopsy revealed myopathic features and loss of dystrophin staining in scattered fibers (approx. 1%). Recent updated genetic testing confirmed the presence of both *DMD* variants in our proband, with variant allele frequencies providing strong evidence for the variants being present in trans. Chromosomal microarray showed no other evidence of X chromosome copy number variants. Currently, in her mid-70s, her symptoms remain milder than would be expected for even BMD.

In this female with a pathogenic *DMD* frameshift variant in trans with a multi-exon duplication, the question remains as to why she did not manifest symptoms until her early 60s, and even now why she is not more severely affected. Although her exact duplication has never been reported, multi-exon deletions and duplications are the most common mutational mechanism, and overlapping and nested duplications have been reported in affected individuals (Kesari et al. 2008). Long-read sequencing may perhaps provide insight, as it is possible her duplication may be inserted elsewhere in the genome, leaving an intact copy of *DMD*. Alternate theories to account for her mild phenotype will also be discussed.

Mendelian Phenotypes Posters - Thursday

PB1763. Biallelic founder mutation in *PDE2A* causes paroxysmal dyskinesia with Intellectual disability in Pakistani families

Authors:

A. Fatima¹, H. Yousaf², S. Baig³, I. zafar⁴, M. Toft⁵; ¹Aga Khan Univ., Karachi, Karachi, Pakistan, ²Natl. Inst. for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan, Faisalabad, Pakistan, ³Dept. of Biological and BioMed. Sci., The Aga Khan Univ., Karachi, Pakistan, Karachi, Pakistan, ⁴Dept. of Neurology, Oslo Univ. Hosp., Domus Medica 4, L-264, Gaustadalleen 34, 0372 Oslo, Norway., Oslo, Norway, ⁵Inst. of Clinical Med., Univ. of Oslo, Oslo, Norway., Oslo, Norway

Abstract Body:

Intellectual developmental disorder with paroxysmal dyskinesia or seizures (IDDPADS, OMIM# 619150) is an ultra-rare childhood-onset movement disorder manifesting paroxysmal dyskinesia slowly progressive global developmental delay, impaired cognitive development and seizures. We investigated three unrelated Pakistani families with six affected individuals born to cousin marriages. Belonging to diverse age groups (13 years - 60 years), patients presented with severe developmental delay, cognitive abnormalities, speech impairment and seizures with onset in early years of age. Whole exome sequencing (WES) was used to detect potentially pathogenic variants and identify shared regions of homozygosity. We identified a rare homozygous missense variant c.1490T>C (p. Phe497Ser) in Phosphodiesterase2A (*PDE2A*) gene NM_002599.5. c.1490T>C variant segregates with the phenotype in all three families. To assess the possibility of a founder mutation, we performed homozygosity mapping which revealed a shared 5.5 Mb homozygous region at Chr11: q13.4-q13.5 among all three families. *PDE2A* is involved in hydrolysis of second messengers cAMP and cGMP. It is activated by binding of cGMP to GAF-B domain of the protein. The function of *PDE2A* is very critical for temporal regulation of these second messengers consequently modulating critical cellular processes like proliferation, neuronal function, apoptosis and differentiation¹. So far, only five variants and five patients are reported with disrupted *PDE2A*¹⁻³. Our findings add important details to the clinical and genetic spectrum of *PDE2A*, and highlight c.1490T>C (p. Phe497Ser) as a founder mutation in Pakistani families.

Mendelian Phenotypes Posters - Wednesday

PB1764*. Bi-allelic LETM1 variants perturb mitochondrial ion homeostasis leading to a clinical spectrum with predominant nervous system involvement

Authors:

R. Kaiyrzhanov, LETM1 study group; South Kazakhstan Med. Academy, Shymkent, Kazakhstan

Abstract Body:

Background/Objectives: The Leucine zipper-EF-hand containing transmembrane protein 1 (*LETMI*) gene encodes an inner mitochondrial membrane protein with an osmoregulatory function controlling mitochondrial volume and ion homeostasis [1]. *LETMI* is deleted in Wolf-Hirschhorn syndrome, which results from de novo monoallelic deletion of chromosome 4p16.3 [1]. Here we describe the first association of bi-allelic *LETMI* variants with human disease. Methods: Exome sequencing and international gene-matching efforts were used to identify affected families with bi-allelic pathogenic variants in *LETMI*. Biochemical and morphological studies on mitochondrial K⁺ activities, proteins, and shape in patient-derived fibroblasts, muscles, and in *S. cerevisiae* as an important model organism for mitochondrial osmotic regulation were performed. Results: Eighteen affected individuals from eleven unrelated families harboring ultra-rare bi-allelic *LETMI* variants and clinical presentations highly suggestive of mitochondrial disease were identified. These manifested as a spectrum of predominantly infantile-onset (14/18, 78%) and variably progressive neurological, metabolic, and dysmorphic symptoms, and multiple organ dysfunction associated with neurodegeneration. The common features included respiratory chain complex deficiencies (100%), global developmental delay (94%), optic atrophy (83%), sensorineural hearing loss (78%), and cerebellar ataxia (78%) followed by epilepsy (67%), spasticity (53%), and myopathy (50%). Our experiments showed defective mitochondrial K⁺ efflux, swollen mitochondrial matrix structures, and loss of important mitochondrial oxidative phosphorylation protein components. Conclusion: Our findings highlight the implication of perturbed mitochondrial osmoregulation caused by *LETMI* variants in neurological and mitochondrial pathologies.

References: Austin, S., and Nowikovsky, K. (2019). LETM1: Essential for Mitochondrial Biology and Cation Homeostasis? Trends in Biochemical Sciences. 44, 648-658

Mendelian Phenotypes Posters - Thursday

PB1765. Biallelic Loss of Function Mutations in *PYGM* Cause Hereditary Macular Dystrophy

Authors:

R. Hussein, A. Tayyib, K. Ahmed, E. Tavares, E. Heon, A. Vincent; The Hosp. for Sick Children, Toronto, ON, Canada

Abstract Body:

BACKGROUND: Hereditary Macular Dystrophies (HMDs) are a genetically and phenotypically heterogeneous group of disorders that lead to irreversible vision loss. Biallelic mutations in the *PYGM* gene, encoding glycogen phosphorylase, cause McArdle disease, with a few cases documented to have HMD as an association. **PURPOSE** To identify the disease-causing variants in an autosomal recessive family with HMD. **METHODS:** The proband tested negative for ~280 known retinal dystrophy genes. Hence, additional family members were recruited (n = 7; two affected) to a research study. Both affected individuals underwent eye exams and paired-end genome sequencing. Candidate variants likely to cause disease were filtered based on customized pipelines. Variants were prioritized if they were shared between the affected, were rare (population frequency < 0.5%), had at least two variants per gene, and had high pathogenicity scores in two predictive algorithms. Variants of interest were segregated in all family members using Sanger sequencing. Immunohistochemistry (IHC) of human retinal sections was conducted using an anti-*PYGM* antibody. **RESULTS:** Both affected individuals showed progressive distance vision loss and patchy macular and peripheral retinal atrophy. Genome filtering revealed a known pathogenic loss of function single nucleotide variant in *PYGM* (NM_005609 c.148C>T; p.Arg50*) in a homozygous state in both individuals. Segregation analysis revealed the variant was also present in a sibling who had not undergone eye exam. Although this variant in *PYGM* is known to cause McArdle disease, the affected individuals in the study were never diagnosed with any systemic features and hence, re-phenotyping was performed. Affected individuals had symptoms and signs of McArdle disease on re-phenotyping, including early muscle pain, exercise intolerance, rhabdomyolysis, and elevated serum creatine kinase. IHC results showed *PYGM* presence in retinal nuclear layers, outer plexiform layer, ganglion cell layer, and nerve fiber layer. **CONCLUSIONS:** This study has identified the cause underlying HMD in the pedigree to be consequent upon biallelic *PYGM* variants and supports the inclusion of *PYGM* in clinical HMD panels. IHC demonstrated *PYGM* presence in the cytoplasm of multiple retinal layers. Together, these findings suggest a role for *PYGM* in retinal glucose metabolism and this mechanism should be further explored. Despite the relatively common prevalence of McArdle disease, the association of HMD is sparse. As such, evaluating McArdle patients for retinopathy may be useful to further delineate this connection.

Mendelian Phenotypes Posters - Wednesday

PB1766. Bi-allelic loss of function variant in the *NRCAM* gene is associated with hereditary polyneuropathy phenotype

Authors:

Z. Elahi^{1,2}, M. Soveyzi¹, M. Goleyjani Moghadam¹, S. Nafissi³, Y. Nilipour⁴, A. Kariminejad², H. Najmabadi^{1,2}, Z. Fattahi^{1,2}; ¹Genetics Res. Ctr., Univ. of Social Welfare and Rehabilitation Sci.s, Tehran, Iran, Islamic Republic of, ²Kariminejad-Najmabadi Pathology & Genetics Ctr., Tehran, Iran, Islamic Republic of, ³Dept. of Neurology, Iranian Ctr. of Neurological Res., Tehran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, ⁴Pediatric Pathology Res. Ctr., Res. Inst. for Children's Hlth., Mofid Children Hosp., Sch. of Med., Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of

Abstract Body:

Recently, the role of Neuronal Cell Adhesion Molecule (NRCAM) in a neurodevelopmental disorder (NDD) has been defined. The phenotype is mainly recognized by varying severity of global developmental delay/intellectual disability, hypotonia, spasticity, and peripheral neuropathy. The affected individuals carry homozygous pathogenic variants located in different domains of the NRCAM protein, mainly in immunoglobulin-like (Ig-like) and fibronectin type III (FN-III) domains. The varying severity of the symptoms ranges from motor neuropathy with secondary myopathic involvement without developmental delay due to homozygous missense variant in Ig-like 1 domain into the most severe forms represented by severe global developmental delay (GDD) along with the neurological and neuromuscular features due to homozygous loss of function (LOF) variants.

Here, as the second report, we present a 19-year-old male, who manifests polyneuropathy, progressive muscle weakness in proximal limbs, and lordosis with no other NDD features. Whole-exome sequencing led to prioritizing a homozygous LOF variant defined as c.73C>T (p.Gln25X) in the *NRCAM* gene (NM_001037132.4). The variant was inherited from the consanguine parents and was located in an 8.82 Mbp ROH interval which is among the most considerable ROH intervals detected in this patient.

We expand the clinical and molecular spectrum of the *NRCAM*-related disorder as a requisite for a more accurate genotype-phenotype correlation, ultimately delineating the *NRCAM*-related conditions. Although in the first *NRCAM* study, patients with LOF variants showed more severe clinical features, the obtained clinical data, together with the molecular findings in our patient, shows that the type of the pathogenic variant in the *NRCAM* gene does not necessarily determine the severity of the phenotype. Accordingly, there could be modifier effects like other genes that compensate for the absence of NRCAM protein, leading to a single but heterogeneous neurodevelopmental disorder. However, the disorder might be categorized as different separate syndromes caused by the *NRCAM* deficiency in future, similar to the other members of the L1 subgroup of immunoglobulin (Ig)-CAMs proteins such as LICAM- and NFASC-related diseases.

Mendelian Phenotypes Posters - Thursday

PB1767. Biallelic loss-of-function variants in mitochondrial phospholipase *PNPLA8* decrease in the number of basal radial glial cells and lead to microcephaly with simplified gyral pattern

Authors:

Y. Nakamura¹, I. S. Shimada², M. Fujimoto¹, E. Sato¹, A. Miyauchi³, F. Miya⁴, T. Tsunoda⁵, Y. Okubo⁶, K. Haginoya⁶, E. Koshimizu⁷, S. Miyatake^{7,8}, N. Matsumoto⁷, Y. Arioka⁹, N. Ozaki⁹, M. Honda¹⁰, S. Oki¹⁰, Y. Taketomi¹¹, M. Murakami¹¹, Y. Kato², S. Saitoh¹; ¹Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, ²Dept. of Cell Biology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, ³Dept. of Pediatrics, Jichi Med. Univ., Tochigi, Japan, ⁴Ctr. for Med. Genetics, Keio Univ. Sch. of Med., Tokyo, Japan, ⁵Lab. for Med. Sci. Mathematics, Dept. of Biological Sci., Graduate Sch. of Sci., The Univ. of Tokyo, Tokyo, Japan, ⁶Miyagi Children's Hosp., Miyagi, Japan, ⁷Dept. of Human Genetics, Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, ⁸Clinical Genetics Dept., Yokohama City Univ. Hosp., Yokohama, Japan, ⁹Dept. of Psychiatry, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, ¹⁰Dept. of Drug Discovery Med., Kyoto Univ. Graduate Sch. of Med., Kyoto, Japan, ¹¹Lab. of Microenvironmental and Metabolic Hlth.Sci., Ctr. for Disease Biology and Integrative Med., Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan

Abstract Body:

Evolutionary expansion and gyrification of the cerebral cortex are both linked to an abundance of basal radial glial cells (bRGCs) because of their highly proliferative potential leading to a pronounced increase in the number of neurons. However, genes involved in bRGC development are not fully understood. We herein report two patients with severe neurodevelopmental disability exhibiting congenital microcephaly with simplified gyral pattern, who carried biallelic loss-of-function variants in *PNPLA8*. *PNPLA8*, encoding mitochondrial phospholipase, is essential for maintaining mitochondrial shape and function through tailoring mitochondrial membrane lipid metabolism and composition, without known link to developing human cortex. To gain insight into brain development of *PNPLA8*-deficient patients, we generated cerebral organoids using *PNPLA8* gene-edited human induced pluripotent stem cells (hiPSCs). The number of bRGCs and upper-layer neurons were significantly reduced in *PNPLA8* knockout (KO) cerebral organoids compared to wild type cerebral organoids. These data suggest that loss of *PNPLA8* function results in microcephaly and simplified gyrus through the reduced number of bRGCs. To further investigate how the bRGCs are reduced, we *in vitro* generated neural stem cells (NSCs) induced from *PNPLA8* gene-edited hiPSCs. Mitochondria were hyperfused and enlarged with obscure cristae shape in *PNPLA8* KO hiPSC-derived NSCs. Given many links between mitochondria and cell fate decision, loss of *PNPLA8* function could affect mitochondria in NSCs, thereby dysregulate bRGC generation. Overall, our data reframe our understanding of the pathophysiology of *PNPLA8*-related disorder, and more broadly, underscore the importance of *PNPLA8*-dependent mitochondrial quality control for NSC fate decisions in gyrencephalic species.

Mendelian Phenotypes Posters - Wednesday

PB1768. Bi-allelic loss-of-function variants in *PPFIBP1* cause a severe neurodevelopmental disorder with microcephaly, epilepsy and periventricular calcifications

Authors:

K. Platzer¹, E. Rosenhahn¹, T. J. O'Brien², M. S. Zaki³, I. Sorge⁴, D. Wieczorek⁵, K. Rostasy⁶, A. Vitobello⁷, S. NAMBOT⁸, F. Alkuraya⁹, M. O. Hashem¹⁰, A. Alhashem¹¹, B. Tabarki¹¹, A. S. Alamri¹², A. H. Al Safar¹², D. K. Bubshait¹², N. F. Alahmady¹³, J. G. Gleeson¹⁴, M. S. Abdel-Hamid¹⁵, N. Lesko¹⁶, S. Ygberg¹⁷, S. P. Correia¹⁸, A. Wredenberg¹⁹, S. Alavi²⁰, S. M. Seyedhassani²¹, M. Ebrahimi Nasab²¹, H. Hussien²², T. Omar²², I. Harzallah²³, R. Touraine²⁴, H. Tajsharghi²⁵, H. Morsy²⁶, H. Houlden²⁷, M. Shahrooei²⁸, M. Ghavideldarestani²⁸, G. Abdel-Salam²⁹, A. Torella³⁰, M. Zanobio³⁰, G. Terrone³¹, N. Brunetti-Pierri³², A. Omrani³³, J. Hentschel¹, J. R. Lemke¹, R. Abou Jamra¹, H. Sticht³⁴, A. E. X. Brown³⁵, R. Maroofian²⁷; ¹Inst. of Human Genetics, Univ. of Leipzig Med. Ctr., Leipzig, Germany, ²MRC London Inst. of Med. Sci., London, United Kingdom, ³Natl. Res. Ctr., Cairo, Egypt, ⁴Dept. of Pediatric Radiology, Univ. Hosp. Leipzig, Leipzig, Germany, ⁵Inst. of Human Genetics, Med. Faculty and Univ. Hosp. Düsseldorf, Düsseldorf, Germany, ⁶Dept. of Pediatric Neurology, Children's and Adolescents' Hosp. Datteln, Witten/Herdecke Univ., Witten, Germany, ⁷CHU Dijon Bourgogne, Dijon, France, ⁸Genetic center, CHU, Dijon, France, ⁹King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia, ¹⁰Dept. of Translational Genomics, Ctr. for Genomic Med., King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, ¹¹Dept. of Pediatrics, Prince Sultan Military Med. City, Riyadh, Saudi Arabia, ¹²Dept. of Pediatrics, Imam Abdulrahman bin Faisal Univ., Dammam, Saudi Arabia, ¹³Biology Dept., Imam Abdulrahman bin Faisal Univ., Damm, Saudi Arabia, ¹⁴Dept. of NeuroSci.s, Univ. of California, San Diego, La Jolla, CA, ¹⁵Natl. Res. Ctr., Giza, Egypt, ¹⁶Dept. of Med. Biochemistry and Biophysics, Karolinska Inst.t, Stockholm, Sweden, ¹⁷Neuropediatric Unit, Dept. of Women's and Children's Hlth., Karolinska Univ. Hosp., Stockholm, Sweden, ¹⁸Ctr. for Inherited Metabolic Diseases, Karolinska Univ. Hosp., Stockholm, Sweden, ¹⁹Dept. of Med. Biochemistry and Biophysics, Karolinska Inst., Stockholm, Sweden, ²⁰Dept. of Cell and Molecular Biology and Microbiol., Faculty of Biological Sci. and Technology, Univ. of Isfahan, Isfahan, Iran, Islamic Republic of, ²¹Dr. Seyedhassani Med. Genetic Ctr., Yazd, Iran, Islamic Republic of, ²²Alexandria Univ. Children's Hosp., Faculty of Med., Alexandria Univ., Alexandria, Egypt, ²³Clinical, Chromosomal and Molecular Genetics Dept., Univ. Hosp. Ctr., Saint-Etienne, France, ²⁴CHU de Saint Etienne, Saint Etienne, France, ²⁵Sch. of Hlth.Sci., Translational Med., Univ. of Skövde, Skövde, Sweden, ²⁶Queen Square Inst. of Neurology, London, United Kingdom, ²⁷UCL Queen Square Inst. of Neurology, Univ. Coll. London, London, United Kingdom, ²⁸Specialized Immunology Lab. of Dr. Shahrooei, Sina Med. Complex, Ahvaz, Iran, Islamic Republic of, ²⁹Clinical Genetics Dept., Human Genetics and Genome Res. Inst., Natl. Res. Ctr., Cairo, Egypt, ³⁰Dept. of Precision Med., Univ. of Campania 'Luigi Vanvitelli', Naples, Italy, ³¹Child Neurology Unit, Dept. of Translational Med. Sci., Univ. of Naples Federico II, Naples, Italy, ³²Telethon Inst Genetics & Med., Naples, Italy, ³³Div. of clinical studies, The Persian Gulf Nuclear Med. Res. Ctr., Bushehr Univ. of Med. Sci., Bushehr, Iran, Islamic Republic of, ³⁴Inst. of Biochemistry, Friedrich-Alexander-Univ. Erlangen-Nürnberg, Erlangen, Germany, ³⁵MRC London Inst. of Med. Sci., London W12 0NN, UK Faculty of Med., Inst. of Clinical Sci., Imperial Coll. London, London, United Kingdom

Abstract Body:

PPFIBP1 encodes for the liprin-β1 protein which has been shown to play a role in neuronal outgrowth and synapse formation in *Drosophila melanogaster*. By exome and genome sequencing, we detected nine ultra-rare homozygous loss-of-function variants in 16 individuals from 12 unrelated families. The individuals presented with moderate to profound developmental delay, often refractory early-onset epilepsy and progressive microcephaly. Further common clinical findings included muscular hyper- and

hypotonia, spasticity, failure to thrive and short stature, feeding difficulties, impaired vision, and congenital heart defects. Neuroimaging revealed abnormalities of brain morphology with leukoencephalopathy, ventriculomegaly, cortical abnormalities, and intracranial periventricular calcifications as major features. In a fetus with intracranial calcifications, we identified a rare homozygous missense variant that by structural analysis was predicted to disturb the topology of the SAM-domain region that is essential for protein-protein interaction. For further insight in the effects of *PPFIBP1* loss-of-function, we performed automated behavioural phenotyping of a *Caenorhabditis elegans* *PPFIBP1/hlb-1* knockout model which revealed defects in spontaneous and light-induced behaviour and confirmed resistance to the acetylcholinesterase inhibitor aldicarb suggesting a defect in the neuronal presynaptic zone. In conclusion, we establish bi-allelic loss-of-function variants in *PPFIBP1* as a cause of an autosomal recessive severe neurodevelopmental disorder with early-onset epilepsy, microcephaly and periventricular calcifications.

Mendelian Phenotypes Posters - Thursday

PB1769. Biallelic nonsynonymous variants in *LTV1* are associated with LIPHAK syndrome, a novel poikiloderma-like disorder

Authors:

J. Han^{1,2}, G. Ryan³, A. Guy⁴, L. Liu⁴, M. Quinodoz^{1,2,5}, I. Helbling⁶, J. Lai-Cheong⁷, Genomics England Research Consortium, J. Barwell^{5,8}, M. Folcher^{1,2}, J. McGrath^{9,10}, C. Moss^{11,12}, C. Rivolta^{1,2,5}; ¹Inst. of Molecular and Clinical Ophthalmology Basel (IOB), Basel, Switzerland, ²Dept. of Ophthalmology, Univ. of Basel, Basel, Switzerland, ³West Midlands Regional Genetics Lab., Central and South Genomic Lab. Hub, Birmingham, United Kingdom, ⁴Viapath, The Natl. Diagnostic Epidermolysis Bullosa Lab., London, United Kingdom, ⁵Dept. of Genetics and Genome Biology, Univ. of Leicester, Leicester, United Kingdom, ⁶Dept. of Dermatology, Univ. Hosp. of Leicester NHS Trust, Leicester, United Kingdom, ⁷Frimley Park NHS Fdn. Trust, Camberley, United Kingdom, ⁸Dept. of Clinical Genetics, Univ. Hosp. of Leicester NHS Trust, Leicester, United Kingdom, ⁹NIHR BioMed. Res. Ctr., Guy's and St Thomas' NHS Fdn. Trust and King's Coll. London, London, United Kingdom, ¹⁰St John's Inst. of Dermatology, King's Coll. London (Guy's Campus), London, United Kingdom, ¹¹Dept. of Paediatric Dermatology, Birmingham Women's and Children's Hosp. NHS FT, Birmingham, United Kingdom, ¹²Coll. of Med. and Dental Sci., Univ. of Birmingham, Birmingham, United Kingdom

Abstract Body:

As part of the 100,000 Genomes Project, we investigated the genetic origin of a previously undescribed recessive dermatological condition that we named LIPHAK (*LTV1*-associated Inflammatory Poikiloderma with Hair abnormalities and Acral Keratoses) in four affected individuals from two UK families of Pakistani and Indian origins, respectively. Patients' skin conditions did not entirely fit the term 'poikiloderma', as they displayed limited telangiectasia and atrophy, as well as a pattern of discrete hyper- and hypopigmented macules. By combining *in silico* high-resolution homozygosity mapping with filtered WGS data, we identified a single variant that had characteristics compatible of those of a Mendelian mutation: the NM 032860.5:c.503A>G, p.(Asn168Ser) change in *LTV1*, a gene encoding a protein involved in the biogenesis of the 40S subunit of the ribosome. This DNA change was present homozygously in all four patients and heterozygously in their parents. Despite potentially leading to a simple missense, *in silico* analysis predicted that the c.503A>G transition might interfere with the canonical splicing pattern of *LTV1*, and in particular it could potentially create a new donor site for exon 5. We tested this hypothesis by producing a recombinant exon-intron-exon construct bearing this variant, as well as its wild-type counterpart. Following transfection of HEK293T cells, *LTV1* minigenes with the mutation indeed resulted in an aberrant and shorter transcript, carrying the deletion of 37 bp at the 3' side of exon 5, leading in turn to a frameshift and the creation of the premature termination codon. RNA analysis and immunofluorescence microscopy of a patient's skin biopsy showed the same splicing pattern, which resulted in an overall reduced expression of *LTV1* protein. Although the precise role of *LTV1* mutations in LIPHAK is still unknown, our data suggest that this novel dermatological condition could be classified as a ribosomopathy.

Mendelian Phenotypes Posters - Wednesday

PB1770. Biallelic pathogenic variants in the endosomal transport regulator *SNAPIN* cause a rare prenatal neuroanatomical syndrome

Authors:

H. Yousaf^{1,2}, M. Koning³, K. Gilmore⁴, M. Hoffer³, G. Kellaris¹, P. Adama van Scheltema⁵, M. Haak⁵, C. Peeters-Scholte⁶, D. Heron⁷, B. Keren⁷, K. Akouri⁸, S. Lanone⁹, M. Dagouassat⁹, B. Yalcin¹⁰, N. Vora⁴, G. Santen³, E. Davis¹; ¹Stanley Manne Children's Res. Inst., Ann & Robert H Lurie Children's Hosp. of Chicago, Chicago, IL, ²Human Molecular Genetics Lab., Natl. Inst. for Biotechnology and Genetic Engineering Coll. (NIBGE-C), Pakistan Inst. of Engineering and Applied Sci. (PIEAS), Islamabad, Faisalabad, Pakistan, ³Dept. of Clinical Genetics, LUMC, Leiden, Netherlands, ⁴Dept. of Obstetrics and Gynecology, Div. of Maternal-Fetal Med., The Univ. of North Carolina Sch. of Med., Chapel Hill, NC, ⁵Dept. of Obstetrics, LUMC, Leiden, Netherlands, ⁶Dept. of Neurology, LUMC, Leiden, Netherlands, ⁷Dept. of Genetics, Pitié-Salpêtrière Hosp., Paris, France, ⁸Dept. of Med. Genetics, Hamad Med. Corp., Doha, Qatar, ⁹Univ Paris Est Creteil, INSERM, IMRB, F-94010, Creteil, France, ¹⁰Inserm Unit 1231, Univ. of Bourgogne Franche-Comté, Dijon, France

Abstract Body:

Fetal brain anomalies cast a considerable healthcare burden with ~1/1,000 affected newborns every year in the US. To identify underlying etiology, prenatal ultrasound and/or magnetic resonance imaging combined with trio whole exome sequencing (WES) has emerged as a comprehensive diagnostic paradigm with a reported diagnostic rate up to ~35%. As part of an ongoing fetus-parent trio WES research study, we identified a fetus presenting with brain anomalies that harbors biallelic nonsense variants in *SNAPIN*, encoding SNARE-associated protein. *SNAPIN* is a ubiquitously expressed component of the autophagy-lysosomal pathway that catalyzes retrograde axonal transport, late endocytic transport, synaptic transmission, and synaptic vesicle fusion and release. Previously reported homozygous *Snapin* knockout mice are embryonic (E14.5) and perinatal lethal and show complex brain deficits. Using GeneMatcher, we identified 5 additional cases from three unrelated families with biallelic rare variants in *SNAPIN*. Prenatal imaging revealed ventriculomegaly (6/6), cerebellar hypoplasia (5/6), corpus callosum agenesis (3/6), abnormal head size (4/6), and lissencephaly (3/6). We also noticed additional anatomical features with variable incidence; talipes equinovarus (5/6) and abnormal craniofacial morphology (3/6) were the most prevalent. None of the cases survive beyond the perinatal period either due to intrauterine demise (1 case), elective termination (3 cases), or neonatal death (2 cases). To investigate further the role of *SNAPIN* in development of anterior structures, we targeted the single ortholog in zebrafish. We generated F0 mutants using CRISPR/Cas9 and performed transient knockdown using morpholinos. Both *snapin* loss-of-function models recapitulated human-relevant disease phenotypes including cerebellar hypoplasia, reduced intertectal neuron count, diminished optic tectum size, and aberrant facial patterning in comparison to wild type. Our data suggest that biallelic variants in *SNAPIN* are a likely cause of a hitherto unreported severe neuroanatomical syndrome and add to the growing list of lysosome-autophagy effectors that are critical for human brain development.

Mendelian Phenotypes Posters - Thursday

PB1771*. Biallelic *RAD51C* loss-of-function variants drive perizygotic SNV/indel hypermutator phenotype in a subject with Fanconi anemia complementation group O.

Authors:

R. Zemet Lazar¹, H. Du¹, T. Gambin², J. Lupski¹, P. Liu^{1,3}, P. Stankiewicz¹; ¹Baylor Coll. of Med., Houston, TX, ²) Inst. of Computer Sci., Warsaw Univ. of Technology, Warsaw, Poland, ³Baylor Genetics, Houston, TX

Abstract Body:

The mutation rate of single nucleotide variants (SNVs) in humans has been estimated at $1.0\text{-}1.8 \times 10^{-8}$ variants per base per generation, giving rise to 60-70 *de novo* variants per genome, including approximately two affecting the coding sequence. Locus-specific mutation rates for CNVs are approximately 100 to 10,000 times greater, i.e., $\sim 10^{-6}$ to 10^{-4} per generation, resulting in 0-1 *de novo* CNV per genome. Interestingly, a rare phenomenon of hypermutation with unusually large numbers of *de novo* mutations (DNMs) has been described for multiple *de novo* CNVs (PMID 28235197) and SNVs (PMID 35545669). We have reported a family with a prenatal diagnosis of multiple congenital anomalies in whom exome sequencing (ES) study revealed compound heterozygous *RAD51C* variants, and the fetus was diagnosed with Fanconi anemia (FA), complementation group O (FANCO). FA proteins have been classified into three groups based on their primary function in interstrand crosslinks (ICL) repair: components of the FA core complex, components of the FA ID2 complex, and the group of repair factors. Belonging to the third category, *RAD51C* is essential for homologous recombination-mediated repair of ICL and double-strand DNA breaks. Unexpectedly, in addition to *RAD51C* variants, the trio-ES analyses also detected eight apparent *de novo* mosaic variants with variant allele fraction (VAF) ranging between 13 and 35% (PMID 29278735). Subsequent trio whole-genome sequencing also revealed an unusual rise in the number of apparent *de novo* SNVs (n=467) and indels (n=34) with no evidence of an increased CNVs rate. Heterozygous variants with VAF ranging between 36% and 64% were considered nonmosaic, likely representing germline mutations (n=82). The VAFs of mosaic variants showed multinomial distribution with most of the mosaic variants' VAFs oscillating around 25%, indicating they most likely occurred during DNA replication in the zygote. Of note, the mutational pattern analyses of the mosaic variants revealed multiple known single base substitution signatures (SBS1, SBS5), including one associated with homologous recombination-based DNA damage repair error (SBS3). In conclusion, our data demonstrate that biallelic *RAD51C* variants and defects in the DNA damage repair process result in a hypermutator phenotype at least in the prenatal period, potentially with the accumulation of mutations in the post-zygotic period. This phenomenon might contribute to the hematological manifestations and the predisposition to tumors in patients with FA. We propose that other FA groups should be investigated for DNMs.

Mendelian Phenotypes Posters - Wednesday

PB1772. Biallelic variants in *CSPG4* cause a novel neurodevelopmental disorder with intellectual disability, global developmental delay, and facial anomalies.

Authors:

S. Da'as^{1,2}, W. Aamer¹, R. Kaiyrzhanov³, S. Efthymiou⁴, A. Thanassoulas⁵, W. Hasan¹, N. Al-Naama¹, R. TEMANNI⁶, S. Tomei⁷, I. Poggiolini¹, A. Almaraghi¹, D. Abdelrahman¹, J. Al-Rayahi¹, N. E. Shahbeck⁸, R. Al Saleh¹, A. Akil¹, I. Alkhawaja⁹, U. Guliyeva¹⁰, S. Gulieva¹⁰, K. Salayev¹¹, I. Hajjiyeva¹¹, SYNaPS Study Group, Queen Square Genomics, N. Rostomashvili¹², A. Verloes¹³, A. Vitobello¹⁴, C. Thauvin-Robinet¹⁵, M. Nomikos⁵, T. Ben-Omran¹, M. F. Elsaid¹, H. Houlden³, K. Fakhro^{16,17}; ¹Sidra Med., Doha, Qatar, ²Hamad Bin Khalifa Univ., Doha, Qatar, ³UCL, London, United Kingdom, ⁴UCL Inst. of Neurology, London, United Kingdom, ⁵Qatar Univ., Doha, Qatar, ⁶SIDRA Med., Doha, Doha, Qatar, ⁷Sidra Med. and Res. Ctr., Doha, Qatar, ⁸Hamad Med. City, Doha, Qatar, ⁹Elbashir Hosp., Amman, Jordan, ¹⁰MediClub Hosp., Baku, Azerbaijan, ¹¹Azerbaijan Med. Univ., Baku, Azerbaijan, ¹²Hôpital Robert-Debré, Paris, France, ¹³Robert DEBRE Univ. Hosp., Paris, France, ¹⁴CHU Dijon Bourgogne, Dijon, France, ¹⁵FHU-TRANSLAD, Dijon, Dijon, France, ¹⁶Sidra Med., Doha, Qatar, Qatar, ¹⁷Weill Cornell Med., Doha, Qatar

Abstract Body:

Background: Chondroitin sulfate proteoglycan 4 (CSPG4) is an extracellular matrix (ECM) component involved in cell proliferation and differentiation, angiogenesis, and neuronal development. **Method:** Clinical and genetic data from affected individuals were collected through the recruitment of patients with neurological disorders and by analyzing rare variants of candidate genes across families from four global sites. We investigated the impact of novel *CSPG4* missense variants utilizing *in silico* tools and *in vivo* functional characterization in the zebrafish model. The impact on neurological and craniofacial development was analyzed in CRISPR/Cas9-mediated zebrafish models. **Results:** Here, we report four recessive *CSPG4* (NM_001897) missense variants [three homozygous: c.A1370G (p.Asp457Gly), c.2627G>A (p.Arg876His) and c.3247C>A (p.Gln1083Lys), and one compound heterozygote: c.A1370G and c.5156A>G (p.Asp457Gly and p.Gln1719Arg)] identified by next-generation sequencing of four unrelated families. All subjects share a novel neurodevelopmental syndrome characterized by severe intellectual disability, global developmental delay, delayed ability to walk, speech and language delay, distinctive facial features, along with varying degrees of neurological impairment, including hypotonia, cerebellar hypoplasia, and / or seizures. All *CSPG4* variants were predicted to be damaging using *in silico* tools, and three-dimensional molecular modeling suggested significant alterations in protein stability, compromising inter- and intra-molecular interactions. Functional studies in zebrafish using CRISPR gene-editing showed that *CSPG4* variants affected notochord development, motor neuron axonal growth, cerebellar structure along with disorganized head scaffold with anomalous skeletal and cartilage components. RNA sequencing of crispant and control fish revealed significant perturbation of ECM regulating pathways and genes previously described to cause mental retardation, dwarfism, and facial anomalies in humans. **Conclusions:** Our study links rare, damaging variants in the ECM gene *CSPG4* to a novel recessive Mendelian neurodevelopmental disorder in humans and supports the role of CSPG4 in early development.

Mendelian Phenotypes Posters - Thursday

PB1773. Biallelic variants in *FICD* leading to inactivation of BiP cause motor neuron disease.

Authors:

A. Rebelo¹, A. Ruiz¹, M. Dohrn¹, M. Wayand², A. Farooq¹, M. Danzi³, D. Beijer¹, B. Aaron¹, J. Vandrovцова⁴, H. Houlden⁴, L. Matalonga⁵, I. abreu¹, G. Rouleau⁶, E. Mehrdad⁷, L. Van der Vondel⁸, Z. Gan-Or⁹, J. Baetz⁸, R. Schule¹⁰, S. Zuchner¹; ¹Univ. of Miami, Miami, FL, ²Univ. of Tübingen, Tübingen, Germany, ³Univ. of Miami, Dakota Dunes, SD, ⁴UCL Queen Square Inst. of Neurology, London, United Kingdom, ⁵CNAG-CRG, Barcelona, Spain, ⁶Montreal Neurological Inst.-Hosp., Montreal, QC, Canada, ⁷McGill Univ., Montreal, QC, Canada, ⁸Univ. of Antwerp, Antwerp, Belgium, ⁹Montreal Neurological Inst., McGill Univ., Montreal, QC, Canada, ¹⁰Ctr. for Neurology and Hertie Inst. for Clinical Brain Res., Tübingen, Germany

Abstract Body:

The chaperone protein BiP, binding immunoglobulin protein, is the master regulator of the unfolded protein response in the endoplasmic reticulum. BiP activity is regulated by the post-translational modification AMPylation, exclusively provided by FICD (FIC domain protein adenylyltransferase) in humans. FICD is a bifunctional enzyme that inactivates BiP by AMPylation during low levels of unfolded proteins, while it deAMPylates BiP when the levels of unfolded proteins rise. Here we identify biallelic variants in *FICD* causing a neurodegenerative disease of upper and lower motor neurons. Affected individuals in four unrelated families harbor a recurring missense variant, Arg374His, critically positioned in the catalytic sequence motif of the enzyme and important for ATP binding. The mutated residue abolishes intramolecular interaction with the regulatory residue Glu234, which is essential to inhibit AMPylation and to promote de-AMPylation by FICD. Consequently, fibroblasts from patients with *FICD* variants have a defective deAMPylase activity, resulting in abnormally increased levels of AMPylated and thus inactivated BiP. Loss of BiP activity in patients likely results in a chronic impairment of the protein quality control system in the ER. These findings will guide the development of therapeutic strategies for motor neuron and related diseases linked to proteotoxic stress.

Mendelian Phenotypes Posters - Wednesday

PB1774. Bibliometric analysis of Alport Syndrome - the impact of next generation sequencing on diagnosis.

Authors:

E. Salia¹, C. L. Simpson¹, R. Wadie², S. Salia³; ¹Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, ²Korle Bu Teaching Hosp., Accra, Ghana, ³Cape Coast Teaching Hosp., Cape Coast, Ghana

Abstract Body:

New Generation Sequencing (NGS) technology has revolutionized the field of genomics. Since the advent of NGS, there has been an improvement in the diagnosing of genetic conditions and the identification of novel mutations associated with some genetic conditions. Alport Syndrome is a rare, inherited disorder characterized by kidney disease, and ocular and hearing abnormalities. It has primarily been linked to mutations in genes belonging to the collagen family, notably: *COL4A3*, *COL4A4*, and *COL4A5* variants. However, some literature has also suggested that novel mutations are implicated. In this study, we analyze the global scientific research output relating to Alport Syndrome. Again, we look at the impact of NGS on the diagnosis of Alport Syndrome and the identification of novel mutations since its commencement at the start of the 21st century. We also look at the number of peer-reviewed articles on Alport Syndrome that emerged from Africa among other indices in this analysis. In addition, we will explore the number of persons diagnosed with Alport Syndrome who had overt abnormalities involving all the 3 organ systems. A standard bibliometric approach is underway. The team first defined the search criteria, keywords, and a definite time period and chose the relevant databases. Currently, the team is refining the research criteria, will export the result, and perform analysis of the results generated. Two independent reviewers are involved in the screening of selected articles based on robust inclusion and exclusion criteria, with a third reviewer acting as a tiebreaker. We will further create graphics to visualize the network of collaborators. In its conclusion, this study will provide a clearer picture of the research done in the field of Alport Syndrome, identify key contributors and collaborations, and highlight the influence of the NGS on the discovery of pathogenic variants. It will also identify gaps in the body of knowledge available and serve as a guide toward targeted future research in such rare conditions.

Mendelian Phenotypes Posters - Thursday

PB1775. Biallelic *ATOH1* Variant in Siblings With Pontocerebellar Hypoplasia, Developmental Delay, and Hearing Loss

Authors:

K. Writzl¹, T. Višnjari¹, A. Maver¹, O. Maloku², G. Bergant¹, H. Jaklič¹, D. Neubauer³, F. Fogolari², N. Pečarič Meglič⁴, B. Peterlin¹; ¹Clinical Inst. of Genomic Med. UMCL, Ljubljana, Slovenia, ²Dept. of Mathematics, Informatics and Physics, Univ. of Udine, Udine, Italy, ³Div. of Paediatrics (D.N.), Dept. of Child, Adolescent & Dev.al Neurology, Univ. Med. Ctr. Ljubljana,, Ljubljana, Slovenia, ⁴Clinical Inst. of Radiology, Univ. Med. Ctr. Ljubljana,, Ljubljana, Slovenia

Abstract Body:

Background and Objectives:To report on the novel association of biallelic variant in atonal basic helix-loop-helix transcription factor 1 (*ATOH1*) gene and pontocerebellar hypoplasia (PCH), severe global developmental delay, intellectual disability, and hearing loss in a family with 2 affected siblings.**Methods:**A detailed clinical assessment and exome sequencing of peripheral blood sample were performed. Segregation analysis with Sanger sequencing and structural modeling of the variant was performed to support the pathogenicity of the variant.**Results:** A homozygous missense variant (NM_005172.1:c.481C>G) in the *ATOH1* gene was identified in the proband and his affected sister. The segregation analysis subsequently confirmed its segregation with an apparently recessive PCH in this family. *ATOH1* encodes for the atonal basic helix-loop-helix (bHLH) transcription factor 1, a core transcription factor in the developing cerebellum, brainstem, and dorsal spinal cord, and in the ear. The identified variant results in the p.(Arg161Gly) amino acid substitution in the evolutionarily conserved DNA-binding bHLH domain of the ATOH1 protein. Biallelic missense variants in this domain were previously reported to result in disordered cerebellar development and hearing loss in animal models. In silico homology modeling revealed that p.Arg161Gly in ATOH1 protein probably disrupts a salt bridge with DNA backbone phosphate and increases the flexibility of the bHLH helix-both of which together affect the binding capability of the bHLH domain to the DNA.**Discussion:**Based on the sequencing results and evidence from structural modeling of the identified variant, as well as with previous reports of *ATOH1* gene disruption, we conclude that *ATOH1* may represent a novel candidate gene associated with the phenotype of PCH, global developmental delay, and hearing loss in humans.

Mendelian Phenotypes Posters - Wednesday

PB1776. Biotin-thiamine responsive basal ganglia disease: A retrospective review of the clinical, radiological and molecular findings of cases in Kuwait with novel variants

Authors:

M. Aburezq¹, A. Alahmad², R. Alsafi³, A. Al-Tawari⁴, D. Ramadan⁴, M. Shafik¹, O. Abdelaty⁵, N. Makhseed⁶, R. Elshafie², M. Ayed⁷, A. Hayat⁸, F. Dashti⁹, D. Marafi^{10,3}, B. Albash², L. Bastaki², H. Alsharhan^{1,10,2}; ¹Dept. of Pediatrics, Farwaniya Hosp., Ministry of Hlth., Farwaniya Governorate, Kuwait, ²Kuwait Med. Genetics Ctr., Ministry of Hlth., Sulaibikhat, Kuwait, ³Dept. of Pediatrics, Adan Hosp., Ministry of Hlth., Hadiya, Kuwait, ⁴Dept. of Pediatrics, Al-Sabah Hosp., Ministry of Hlth., Shuwaikh, Kuwait, ⁵Dept. of Tadiology, Farwaniya Hosp., Ministry of Hlth., Farwaniya Governorate, Kuwait, ⁶Dept. of Pediatrics, Jahra Hosp., Ministry of Hlth., Hadiya, Kuwait, ⁷Dept. of Neonatology, Farwaniya Hosp., Ministry of Hlth., Farwaniya Governorate, Kuwait, ⁸Dept. of Radiology, Adan Hosp., Ministry of Hlth., Hadiya, Kuwait, ⁹Dept. of Radiology, Ibn Sina Hosp., Ministry of Hlth., Shuwaikh, Kuwait, ¹⁰Dept. of Pediatrics, Hlth.Sci. Ctr., Faculty of Med., Jabriya, Kuwait

Abstract Body:

Background: Biotin-thiamine-responsive basal ganglia disease (BTRBGD) is a rare autosomal recessive neurometabolic disorder caused by biallelic pathogenic *SLC19A3* variants. It is characterized by subacute encephalopathy that is usually triggered by febrile illness and associated with confusion, convulsions, or other neurological manifestations.

Methods: A retrospective analysis conducted on the data registry in Kuwait Medical Genetics Center for all cases clinically and radiographically diagnosed with BTRBGD and genetically confirmed to have deleterious biallelic variants in *SLC19A3*.

Results: Eighteen individuals from 12 families were diagnosed with BTRBGD in Kuwait including 10 (55%) males and 8 (44%) females aged 2-36 years. All were Kuwaiti nationals except for two individuals. Average age of diagnosis ranged from 2 to 3 years with four cases having a significantly delayed diagnosis at 32, 20, 6 and 4 years of age. Most cases (89%) presented initially with dystonia, subacute encephalopathy, confusion, convulsions, dysarthria, or dysphagia that has resolved completely over 2 weeks in most cases (66%) but progressed in only few (22%) to severe cogwheel rigidity, dystonia and quadriparesis. Two individuals were diagnosed pre-symptomatically and were started on treatment. Neuroradiological findings of most cases (83%) revealed bilateral central necrosis of the head of the caudate with partial or complete involvement of the putamen. The molecular diagnosis was confirmed in all cases, which revealed a previously reported Saudi founder homozygous pathogenic *SLC19A3* variant, c.1264A>G p.(Thr422Ala) in 16 cases and two novel homozygous missense variants, c.952G>A, p.(Ala318Thr) in a Kuwaiti individual and c.175T>C, p.(Trp59Arg) in a Jordanian individual. All individuals are still alive at age 2-32 years on high doses of biotin and thiamine with no neurological sequelae, apart from the oldest four cases who have residual neurological deficits in form of persistent dystonic movements and quadriparesis.

Conclusion: We present 18 unreported individuals from Kuwait with BTRBGD and describe two novel *SLC19A3* variants, in addition to the previously reported Saudi founder variant. BTRBGD is a rare treatable condition that requires high level of suspicion to achieve early diagnosis and prompt management. Individuals presenting with unexplained encephalopathy and characteristic brain MRI findings of bilateral signal alteration of caudate nucleus and putamen should be managed immediately with a trial of high dose of biotin and thiamine supplements as disease progression and prognosis is greatly affected by the timing of treatment initiation.

Mendelian Phenotypes Posters - Thursday

PB1777. Brachyolmia with amylogenesis imperfecta: Identification of a Novel LTBP3 gene Mutation in several Druze Arab patients from the north of Israel

Authors:

Y. Hadid; The Bnai Zion Med. Ctr., Haifa, Israel

Abstract Body:

Background: Short stature is a common finding among the general population. A few cases of the combination of short stature and dental congenital anomalies have been reported. Recently, we examined several patients from related families sharing the same clinical phenotype which includes: short stature and congenital dental abnormalities. **Purpose:** We describe seven patients (5 males 2 females) 8-60 years old from four consanguineous Arab Druze related families from the north of Israel. All patients presented with: short stature, short trunk (brachyolmia), and teeth abnormalities (defect in enamel formation and mineralization, oligodontia, abnormal shape and color, retarded eruption and avulsion). **Methods:** the clinical characterization of patients: medical information retrieved from medical history interviewing patients and parents, from medical records and by patient's physical examination. In order to identify the causing gene and mutation, DNA samples were extracted from sixteen patients and healthy family members from the four families. The analysis was done by using CMA- chromosomal microarray analysis with the **Homozygosity mapping and candidate gene approach**. We used the SNP chromosomal microarrays (CMA) 750K CytoScan kit from Affymetrix. We look for the candidate genes within the homozygote mapped regions and then sequenced the selected gene by Sanger technology. **Results:** CMA analysis in 3 patients and 2 healthy members from the four families showed no CNV-copy number variations alterations. The analysis revealed only one homozygote region of 20 Mb in chromosome 11, shared between all affected patients. This region spans from 11p11.2- 11q13.3 and is shared in all affected patients. Revision of the physical map revealed a chromosomal region that contains 301 OMIM genes. This region harbors the LTBP3 gene (LATENT TRANSFORMING GROWTH FACTOR-BETA-BINDING PROTEIN-3) which is responsible for brachyolmia with amylogenesis imperfecta (OMIM 602090). Thus, it was the first candidate gene to be sequenced. We sequenced all 29 LTBP3 exons all were normal but a splice mutation, c.1346-1G>A chr11:65319629, in exon 8 was identified. All tested patients were homozygous for this alteration and as expected, the available parents were heterozygous. This is a novel mutation since it segregated very well in patients and healthy family members. **Conclusions:** Here we describe the full clinical picture of brachyolmia with amylogenesis imperfecta patients and the identification of their LTBP3 gene novel mutation.

Mendelian Phenotypes Posters - Thursday

PB1780. Characterization of associated non-classical phenotypes in patients with deletion in WAGR region identified by chromosomal microarray: new insights and literature review

Authors:

S. Oliveira, V. Souza, G. Cunha, B. Versiani, C. de Oliveira, M. da Silva Rosa, **C. Mendes-Júnior**, P. Moretti, J. Mazzeu, A. Pic-Taylor; Univ.e de Brasilia, Brasilia, Brazil

Abstract Body:

WAGR syndrome (Wilms' tumor, aniridia, genitourinary changes, and intellectual disability) is a contiguous gene deletion syndrome characterized by the joint deletion of PAX6 and WT1 genes, located in the short arm of chromosome 11. However, most deletions include other genes, leading to multiple associated phenotypes. Therefore, understanding how genes deleted together can contribute to other clinical phenotypes is still considered a challenge. In order to establish genotype-phenotype correlation in patients with interstitial deletions of the short arm of chromosome 11, we selected 17 patients with deletions identified by chromosomal microarray analysis: 4 new subjects and 13 subjects previously described in the literature with detailed clinical data. Through the analysis of deleted regions and the phenotypic changes, it was possible to suggest the contribution of specific genes to several nonclassical phenotypes, contributing to the accuracy of clinical characterization of the syndrome and emphasizing the broad phenotypic spectrum found in the patients. This study reports the first patient with a PAX6 partial deletion who does not present any eye anomaly thus opening a new set of questions about the functional activity of PAX6.

Mendelian Phenotypes Posters - Wednesday

PB1781. Characterization of *CFTR* Variants Across Ancestries

Authors:

J. Ideozu, M. Liu, B. Riley-Gillis, A. Singh, J. Waring, A. Vasanthakumar; AbbVie, North Chicago, IL

Abstract Body:

Introduction: Cystic fibrosis (CF) is an autosomal recessive disease caused by pathogenic variants in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. Small molecule therapies targeting the basic *CFTR* defects have shown clinical benefits, but therapeutic responses are largely dependent on the patients' *CFTR* genotype. Although CF is most common among Caucasians, CF-causing variants tend to be ancestry-specific. Therefore, characterizing the broad spectrum of pathogenic and non-pathogenic *CFTR* variants across ancestries is critical for revolutionizing molecular diagnoses of CF. **Methods:** We interrogated 454,857 whole exome sequences available in UK Biobank (UKBB) to characterize the ancestral diversity of *CFTR* variants. We used the ancestry assignment performed by the pan-UKBB to investigate the distribution of *CFTR* variants in African, American, Central South Asian, East Asian, European, and Middle Eastern populations. We then performed an overlap analysis to decipher ancestry-specific *CFTR* variants. CF-causing variants were annotated using the CFTR2 database, and then their distribution across the six ancestries was recorded. **Results:** Overall, we detected more than 3000 variants in the *CFTR* gene. Many of these variants have never been reported in CF. The highest number of *CFTR* variants were detected in Europeans [n=3,192] while the American group had the least number of *CFTR* variants [n=1507]. We found several variants specific to each ancestry, with Europeans having most of the unique variants [n=2118], while the American group had the least number of unique *CFTR* variants [n=19]. F508del was the most prevalent CF-causing variant found in all ancestries, except in East Asia, where V520F was the most prevalent. **Conclusion:** Systematic characterization of *CFTR* variants across six ancestral populations in the UKBB identified several unreported *CFTR* variants, with most of the unique variants found in Europeans. The identification of several unique variants in other ancestries warrants the need for further studies to characterize their possible phenotype relevance.

Mendelian Phenotypes Posters - Thursday

PB1782. Characterization of transthyretin *TTR* missense variants for associations with hereditary transthyretin-mediated (hATTR) amyloidosis and hallmark symptoms in 470,000 UK Biobank whole exome sequences

Authors:

M. Plekan, R. Hoffing, L. Ward, A. Deaton, C. Willis, P. LoGerfo, L. Krohn, A. Holleman, P. Nioi; Alnylam Pharmaceuticals, Cambridge, MA

Abstract Body:

Over 100 distinct missense variants in the *TTR* gene are known to cause hereditary transthyretin-mediated (hATTR) amyloidosis, a progressively debilitating, life-threatening disease. We aimed to characterize missense *TTR* variants in a large cohort of 470,000 exomes and to test their association with hATTR amyloidosis and hallmark symptoms of hATTR amyloidosis. Missense variants in *TTR* were classified as pathogenic, benign, conflicting interpretation, uncertain significance, or novel based on ClinVar annotations, with “novel” indicating the variant had not been previously assessed for pathogenicity. For these variants, the prevalence of an hATTR amyloidosis diagnosis and common hATTR amyloidosis symptoms (polyneuropathy, carpal tunnel syndrome, heart failure and cardiomyopathy) were characterized. Of 84 missense *TTR* variants identified in the European ancestry population (N=363,973), carriers of eight different missense variants had a diagnosis of amyloidosis. Of these eight variants, six are rare (minor allele frequency [MAF] < 0.1%), including known pathogenic variants Glu89Lys (1/1 carriers with a diagnosis), Ser23Asn (1/1 carriers with a diagnosis), Thr60Ala (4/19 carriers with a diagnosis), and Val122Ile (9/358 carriers with a diagnosis), and two variants presenting with conflicting interpretation of pathogenicity, Arg104His (1/60 carriers with a diagnosis) and ENST00000610404.5:p.Glu119Lys (6/235 carriers with a diagnosis). Carriers of two more common (MAF > 0.1%) benign variants Gly6Ser (66/67,191 carriers with a diagnosis) and Thr119Met (1/3,207 carriers with a diagnosis) were also diagnosed with amyloidosis. However, we did not detect significant associations of these benign variants with hATTR amyloidosis or any hallmark symptoms. Despite low penetrance of an hATTR amyloidosis diagnosis among missense *TTR* variant carriers, 58/416 carriers of pathogenic *TTR* variants have been diagnosed with common symptoms of hATTR amyloidosis but no formal diagnosis of amyloidosis, suggesting an underdiagnosis of disease. There is no evidence to suggest pathogenicity in novel *TTR* variants after assessing diagnoses of these variants for hATTR amyloidosis.

Mendelian Phenotypes Posters - Wednesday

PB1783. Characterizing human disease-causing mutations: A zebrafish-based approach

Authors:

J. Burgess¹, R. Noche², X. Cui¹, R. Simonian¹, E. Pannia¹, J. Dowling¹; ¹The Hosp. for Sick Children, Toronto, ON, Canada, ²Yale Zebrafish Res. Core, Dept. of Comparative Med., Yale Sch. of Med., New Haven, CT

Abstract Body:

Next-generation sequencing has allowed for rapid advances in the identification of candidate disease-causing mutations. However, functional testing and validation of these mutations has lagged, particularly for rare mutations found only in single families or small populations. At the Hospital for Sick Children in Toronto (SickKids), we have established a Zebrafish Genetics and Disease Models Core Facility to allow for testing of candidate disease genes in zebrafish, and to facilitate precision medicine initiatives. Zebrafish have quickly emerged as a powerful model organism to validate gene function due to several key features. Zebrafish exhibit high genetic conservation, with 84% of human disease-associated genes having a clear zebrafish counterpart. As a vertebrate, zebrafish also have many organs and tissues that are similar to humans. Zebrafish offer several distinct advantages over mouse models, including large clutch sizes, ex-utero development, transparent bodies through early development which facilitates imaging, and the ability to perform large-scale drug screens in 96-well plates using zebrafish larvae. Our Zebrafish Core Facility provides services to efficiently generate mutations in zebrafish. We make use of CRISPR-Cas9 to generate small insertion/deletion mutations in a gene of interest. We can also perform targeted gene knock-ins to introduce specific mutations identified in patients. Once mutant zebrafish lines are established, we can perform phenotypic analysis to validate zebrafish disease models, and conduct drug screens to help identify novel therapeutic agents. To date, we have developed zebrafish models for a diverse set of human diseases including inflammatory bowel disease, pediatric cancer, cardiac arrhythmia, and childhood muscle disease. In addition, we can use *Tol2*-mediated transgenesis for overexpression studies. We have worked with over 20 labs and generated insertion-deletions in over 30 genes using CRISPR-Cas9, including at least 3 large deletions (several kb), as well as over 9 targeted mutation knock-ins based on candidate human disease genes. Here we present results of our high-throughput mutation generation efforts and summarize our pipeline approach.

Mendelian Phenotypes Posters - Thursday

PB1784. *Chd8* mutation in mouse impacts cerebellar function across anatomical, genomic, electrophysiological and behavioral axes

Authors:

C. Canales, A. D'Ambra, S. Jung Jung, N. Seban, N. Chu, S. Lozano, K. Cichewicz, E. Smith, R. Ortiz, J. Zhu, M. Corea, D. Rahbarian, D. Fioravante, A. Nord; Univ. of California Davis, Davis, CA

Abstract Body:

The role of the cerebellum (CB) in proprioception and motor control is well-established, however recent studies implicate this structure in high order cognitive functions that are tightly related to Intellectual Disability and Autism Spectrum Disorder ASD. Nevertheless, its role of non-motor contributions such as cognitive processing and affective regulation, though often overlooked in ASD research, is becoming an intriguing focal point for CB research. De novo mutations in the chromatin-remodeling factor *CHD8* are strongly associated with ASD and more generally with neurodevelopmental disorders (NDDs). Using a germline *Chd8*^{5bp-del} haploinsufficient mouse model, we applied neuroanatomical, genomic, and electrophysiological assays to test for deficits in cerebellar development at postnatal day (PND) 12. Additionally, we evaluated putative *Chd8* function in Deep Cerebellar Nuclei (DCN) signaling by assessing anxiety-like behavior, motor-coordination, associative learning and social behavior in CB-specific, viral-induced Cre mediated null *Chd8* adult mouse. We analyzed transcriptional landscape in the CB at single cell resolution at PND12 via snRNA-sequencing as well as neuroanatomical and electrophysiological properties via immunohistochemistry (IHC) and Purkinje neuron (PN) patch clamp. Ex vivo PN electrophysiology showed sexually dimorphic phenotypes indicating differences in inhibitory synaptic transmission. Adult *Chd8* cre-mediated mutation resulted in aberrant sex-specific associative learning and social behavior, with no effects in anxiety-like behavior or motor coordination. Differentially expressed genes (DEGs) indicated perturbation to metabolic and synaptic pathways within specific cerebellar cell populations. Ongoing work aims to replicate findings and validate impacted cell types and functional signaling consequences of these effects. Altogether, our findings are foundation for future studies towards the ultimate goal of understanding how *Chd8* mutation impacts specific cerebellar cell-types and circuits, and how aberrant CB signaling is associated with ASD-relevant neuropathology and social and cognitive deficits.

Mendelian Phenotypes Posters - Wednesday

PB1785. *CHP2* is a genetic modifier of risk of chronic *Pseudomonas aeruginosa* airway infection in cystic fibrosis.

Authors:

K. Buckingham¹, W. Gordon¹, A. Faino², A. Stilp³, E. Blue^{4,5}, Y-H. Zhou^{6,7}, M. Rosenfeld^{8,9}, P. Qu², J. Collaco¹⁰, R. Pace¹¹, G. Cutting¹², M. Knowles¹¹, R. Gibson^{8,9}, M. Bamshad^{1,5,8,13}, The Cystic Fibrosis Genome Project; ¹Dept. of Pediatrics, Div. of Genetic Med., Univ. of Washington, Seattle, WA, ²Children's Core for Biostatistics, Epidemiology and Analytics in Res., Seattle Children's Res. Inst., Seattle, WA, ³Dept. of Biostatistics, Univ. of Washington, Seattle, WA, ⁴Dept. of Med., Div. of Med. Genetics, Univ. of Washington, Seattle, WA, ⁵Brotman Baty Inst. for Precision Med., Seattle, WA, ⁶Bioinformatics Res. Ctr., North Carolina State Univ., Raleigh, NC, ⁷Dept. of Biological Sci., North Carolina State Univ., Raleigh, NC, ⁸Ctr. for Clinical and Translational Res., Seattle Children's Res. Inst., Seattle, WA, ⁹Dept. of Pediatrics, Univ. of Washington, Seattle, WA, ¹⁰Eudowood Div. of Pediatric Respiratory Sci., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ¹¹Marsico Lung Inst./UNC CF Res. Ctr., Sch. of Med., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ¹²McKusick-Nathans Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ¹³Dept. of Genome Sci., Univ. of Washington, Seattle, WA

Abstract Body:

Chronic *Pseudomonas aeruginosa* (Pa) airway infection in persons with cystic fibrosis (PwCF) is common, and is associated with substantial morbidities (e.g., lung function decline) and mortality. Utilizing whole genome sequencing and phenotypic data from 2,358 participants in the Cystic Fibrosis Genome Project (CFGP) who had their first Pa culture by age 3, we conducted a time-to-event genome-wide association study (GWAS) using a definition of chronic Pa that is applicable to both annual and encounter-based Pa culture data. Specifically, we performed a survival analysis of chronic Pa with an endpoint defined as the second Pa positive year within a sliding window of three years (i.e., the "2 of 3" chronic Pa definition). A Cox regression, adjusting for study site, sex, and stratifying on birth cohort to account for treatment eras was performed. Martingale residuals from this regression were normalized and used as the GWAS outcome. Relatedness estimates and principal components (PCs) were calculated using PC-Relate and PC-AiR. A GWAS of the normalized Martingale residuals from the survival analyses was performed, adjusting for sex, study site, and PCs 1 and 2 as covariates, birth cohort as a grouping covariate, and the genetic relatedness matrix as a random effect. Additive effect score tests were conducted for high quality variants with minor allele frequency (MAF) >1% and missingness <2%. We identified a region of association on chromosome 16 containing a lead single nucleotide variant (SNV), rs194810 (MAF=0.43), that meets the threshold for genome-wide significance ($p=2.2E-8$). Genomic inflation was absent ($\lambda=1.01$). This SNV is located near *CHP2*, which encodes calcineurin like EF-hand protein 2. The protein product of *CHP2* is a calcium-binding protein that contributes to the regulation of cell pH by controlling the exchange of sodium and hydrogen ions across the plasma membrane. Another SNV nearby *CHP2*, rs194788, has been previously associated with lower lung function in PwCF and higher risk for *Mycobacterium avium* complex airway disease in persons without CF, and in our cohort, showed strong association ($p=5.5E-8$) with time to chronic Pa. Sensitivity analyses repeating the GWAS while adjusting for lung function revealed little change in the strength of association for either rs194810 ($p=8.6E-8$) or rs194788 ($p=7.4E-7$). Collectively, these results suggest that *CHP2* is a novel risk modifier of time to chronic Pa airway infection in PwCF and that it plays multiple roles in lung function and disease in persons with and without cystic fibrosis.

Mendelian Phenotypes Posters - Wednesday

PB1787. Clinical and molecular characteristics of idiopathic midaortic syndrome in pediatric patients - preliminary results.

Authors:

M. Pelc¹, E. Ciara¹, L. Obrycki¹, J. Antoniewicz¹, P. Halat-Wolska¹, D. Piekutowska-Abramczuk¹, B. Chalupczynska¹, M. Gawlik¹, D. Siestrzykowska¹, K. Chrzanowska¹, R. Ploski², M. Litwin¹; ¹The Children's Mem. Hlth.Inst., Warsaw, Poland, ²Warsaw Med. Univ., Warsaw, Poland

Abstract Body:

Introduction: Midaortic syndrome (MAS) is a rare malformation causing segmental or diffuse stenosis of the thoracic or abdominal aorta, which leads to severe, refractory hypertension. It can be associated with stenoses of other arteries, including visceral, renal, carotid and intracranial, and result in an extensive vascular disease. Congenital MAS may occur in the course of genetic syndromes such as neurofibromatosis type 1, Williams, Alagille, or in moyamoya angiopathy. However, most cases are considered idiopathic/non-syndromic (nMAS), presumably with monogenic background. The molecular basis and clinical symptoms of nMAS are not well characterized.

Methods: Whole-exome sequencing was performed in 10 unrelated Polish children diagnosed with nMAS. The primary analysis focused on over 200 genes potentially associated with the etiology of vasculopathy, including 39 candidates for MAS.

Results: Five of ten patients had isolated middle aortic stenosis (one case of atresia), while in the other five children extra-aortic locations involved renal and visceral arteries. One patient also presented with stenoses of carotid arteries and in two cases aortic aneurysm was found. In 3 of 10 patients a new single nucleotide variant of potential clinical significance was identified in a candidate gene for nMAS: *ELN* (n=2, a missense change and a frameshift/splice-site duplication) or *TBX1* (n=1, missense). Two of three patients with candidate gene defect had extra-aortic stenoses (*ELN* or *TBX1*-positive) and the third (with *ELN* missense variant) only isolated aorta narrowing. Candidate genes were proposed for clinical evaluation in another 3 probands: *ABCC6*, *ABLI*, *CNOT3*, *MYLK2*, *SKI*.

Conclusions: Idiopathic MAS is an anatomically variable vasculopathy, which may be problematic to classify as non-syndromic or diffuse/syndromic depending on the extent of multivessel involvement and associated clinical findings. It is still unclear if the location, extent and the course of vascular disease depends on the genetic background of nMAS. Our preliminary results suggest that postulated gene candidates: *ELN* and *TBX1* may be involved in the etiology of nMAS. Also, we broaden the molecular spectrum of MAS. Nevertheless, unraveling the molecular background and genotype-phenotype correlations in nMAS requires studies in larger groups of patients. It is important for genetic counseling, risk stratification and individualized therapy for each patient, as severe vascular occlusions associated with nMAS require multiple invasive interventions in all children, as well as long-term monitoring.

Grants: Supported by CMHI S162/2018, MEiN 7071/IB/SN/2020, 7088/II-KDM/SN/2020.

Mendelian Phenotypes Posters - Thursday

PB1788. Clinical diversity and molecular mechanism of VPS35L-associated Ritscher-Schinzel syndrome.

Authors:

S. Otsuji¹, Y. Nishio¹, M. Tsujita², R. Marlene³, C. HUBER LEQUESNE³, C. Antón-Plágaro⁴, S. Mizuno⁵, Y. Kawano⁶, S. Miyatake^{7,8}, M. Simon⁹, E. Van Binsbergen⁹, R. van Jaarsveld⁹, N. Matsumoto⁷, V. Cormier-Daire³, P. J Cullen⁴, S. Saitoh¹, K. Kato⁴; ¹Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, ²Dept. of Biochemistry, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, ³Dept. of Genetics, Hôpital Necker Enfants Malades (AP-HP), Paris, France, ⁴Sch. of Biochemistry, Univ. of Bristol, Bristol, United Kingdom, ⁵Dept. of Pediatrics, Aichi Dev.al Disability Ctr., Kasugai, Aichi, Japan, ⁶Dept. of Pediatrics, Toyota Mem. Hosp., Toyota, Japan, ⁷Human Genetics, Yokohama City Univ., Graduate Sch. of Med., Yokohama, Japan, ⁸Clinical Genetics Dept., Yokohama City Univ. Hosp., Yokohama, Japan, ⁹Dept. of Genetics, Univ. Med. Ctr. Utrecht, Utrecht, Netherlands

Abstract Body:

Purpose. The Retriever subunit *VPS35L* is the third responsible gene for Ritscher-Schinzel syndrome (RSS) after *WASHC5* and *CCDC22*. To date, only one pair of siblings have been reported and their condition was significantly more severe than typical RSS. This study aimed to understand the clinical spectrum and underlying molecular mechanism in *VPS35L*-associated RSS. **Methods.** We report three new patients with biallelic *VPS35L* variants. Biochemical and cellular analyses were performed to elucidate disease etiology. **Results.** In addition to typical features of RSS, we confirmed hypercholesterolemia, hypogammaglobulinemia and intestinal lymphangiectasia as novel complications of *VPS35L*-associated RSS. The latter two complications as well as proteinuria, which was observed in three of four patients with *VPS35L* variants, have not been reported in patients with *CCDC22* and *WASHC5* variants. One patient showed a severe phenotype and the other two were milder. Cells established from patients with the milder phenotypes showed relatively higher *VPS35L* protein expression compared to those from patient with more severe form. Cellular analysis found *VPS35L* ablation decreased the cell surface level of LRP1 and LDLR, resulting in reduced LDL cellular uptake. **Conclusion.** *VPS35L*-associated RSS is a distinct clinical entity with diverse phenotype and severity, with a possible molecular mechanism of hypercholesterolemia. These findings provide new insight into the essential and distinctive role of Retriever in human development.

Mendelian Phenotypes Posters - Thursday

PB1789. Clinical usefulness of a custom Next Generation Sequencing gene panel in providing molecular diagnosis to patients with a broad spectrum of Neurodevelopmental Disorders

Authors:

M. Giordano¹, C. Puricelli², D. Vurchio², S. Ronzani³, S. Favini³, A. Maruzzi², A. Spano³, F. SIRCHIA⁴, G. Mandrile⁵, A. Pelle⁶, A. Zonta⁷, I. Rabbone¹, M. Viri³, S. Mellone³; ¹Univ. of Eastern Piedmont and Maggiore Hosp. della carità, Novara, Italy, ²Univ. of Eastern Piedmont, Novara, Italy, ³Maggiore Hosp. della Carità, Novara, Italy, ⁴Univ. of Pavia, Cuneo, Italy, ⁵Unit of Med. Genetics, Univ. of Turin, Torino, Italy, ⁶Unit of Med. Genetics, Univ. of Turin, Torino, Italy, ⁷Unit of Med. Genetics, Univ. of Turin, Novara, Italy

Abstract Body:

Neurodevelopmental disorders (NDDs) comprise a clinically and genetically heterogeneous group of conditions that affect 2-5% of children and represents a public health challenge due to complexity of the etiology. Only few patients with unexplained syndromic and nonsyndromic NDDs receive a diagnosis through first-tier genetic tests as array-CGH and the search for FMR1 CGG expansion. The aim of this study was to evaluate the clinical performance of a targeted next-generation sequencing (NGS) gene panel as a second-tier test in a group of undiagnosed patients with NDDs. A cohort of 450 individuals (268 males and 182 females) with a broad spectrum of NDDs (ID, ASD, motor and/or language disorders, and dysmorphic features potentially due to a syndromic condition) was analyzed with a NGS panel including 221 genes found to be altered in NDDs. Epilepsy and ASD, as well as other clinical features, were present both as a comorbidity in subjects with ID and as conditions in subjects with normal cognitive function. A manual prioritization procedure based on expert knowledge related to the disease phenotype and gene functions detected Pathogenic (N=84) and Likely Pathogenic variants (N=11) in 95 patients corresponding to 21% of the whole cohort. A de novo origin was present in 61 (64%) of the available trios. The diagnostic yield was significantly higher in females than in males (29.4% vs 15.3%; p=0.0019) in particular in ASD (36.8% vs 7.6%; p=0.0026) and Epilepsy (38.9% vs 14.4% p= 0.001). The most involved genes were SLC2A1, SCN1A, ANKRD11, ATP1A2, CACNA1A, FOXP1, and GNAS altered in more than two patients accounting for the 19.7% of the diagnosis. SLC2A1, causing GLUT-1 deficiency, was mutated in 5 patients with both comorbid and isolated epilepsy. Early detection of this alterations is clinically important, since subjects may benefit from ketogenic diet, which may mitigate symptoms and even prevent their progression. Our findings showed that this NGS panel represents a powerful and affordable clinical tool, significantly increasing the diagnostic yield in patients with different form of NDDs in a cost- and time-effective manner without the need of large investments in data storage and bioinformatic analysis.

Mendelian Phenotypes Posters - Wednesday

PB1790. Combined brain abnormalities in patients with beta-tubulin mutations.

Authors:

H. Bae¹, G-H. Kim², I. Choi³, H. Do⁴, S. Kim⁴, J. Kim⁴, B. Lee^{1,2}; ¹Dept. of Pediatrics, Asan Med. Ctr. Children's Hosp., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, ²Med. Genetics Ctr., Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, ³Dept. of Genetic Counseling, Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, ⁴Asan Inst. for Life Sci., Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of

Abstract Body:

Background/Objectives: Tubulins are the structural subunit proteins that form the microtubules. There are 7 beta-tubulin genes with products known, out of a total 26 human tubulin genes (*TUBB*, *TUBB2A*, *TUBB2B*, *TUBB3*, *TUBB4A*, *TUBB4B* and *TUBB6*). Mutations of these genes result in an extremely rare neuronal migration disorder known as tubulinopathy. **Methods:** The genetic and clinical characteristics of 10 unrelated Asian patients with causative beta-tubulin gene mutations were described. **Results:** The median age at diagnosis was 2.4 years (range, 1 month to 9 years). Delayed development was noted in all 10 patients. In addition, hypotonia, ataxia, seizure, and tremor were observed in 10, 5, 3, and 1 patients, respectively. Brain abnormalities were revealed as anomaly of corpus callosum (8pts), enlargement of ventricle (7 pts), and basal ganglia abnormalities (7 pts). In addition, there were complete corpus callosal agenesis (1 pts), thin corpus callosum (7 pts), thinning of white matter (7 pts), hypoplasia of pons (6 pts), cerebellar atrophy/hypoplasia (6 pts), periventricular abnormal intensities (3 pts), and perisylvian polymicrogyria (2 pts). Of note, combined brain abnormalities were observed; three (2 pts), and four or more abnormalities (7 pts). A total of 8 different variants were identified (two *TUBB2A*, one *TUBB2B*, three *TUBB3*, and two *TUBB4A* variants). All variants were missense variants and classified as pathogenic (4 variants) or likely-pathogenic (4 variants) according to the ACMG guidelines. One *TUBB2A* variant was previously unreported. All patients were confirmed as a de novo variant by trio Sanger sequencing. **Conclusions:** The tubulinopathy, even if it is extremely rare, should be suspected in patients with delayed development and abnormal combined brain abnormalities such thin corpus callosum, basal ganglia and ventriculomegaly.

Mendelian Phenotypes Posters - Thursday

PB1791. Comparing phenotypes across five developmental and epileptic encephalopathies (DEEs) through evaluation of 2490 patient data years.

Authors:

E. Mallory, M. Zanna, G. Beek, E. McNaughton, A. Lacoste, E. Brimble; Invitae, San Francisco, CA

Abstract Body:

Developmental and epileptic encephalopathies (DEEs) are severe early-onset epilepsy syndromes with poor prognosis. Characterizing onset and evolution of phenotypes can provide valuable insights, but remains challenging given their rarity and heterogeneity. To address this, we leverage a novel real-world data platform to evaluate relative severity and phenotypic variability in five DEEs. Invitae's Ciitizen is a patient-facing platform that transforms medical records into structured datasets that describe phenotypes and interventions. In this work, we compared diagnoses and volume of data across five DEEs. To quantify relative seizure burden, we developed a composite seizure severity score. Data were analyzed using nonparametric tests. We evaluated data from 338 individuals with DEEs caused by variants in *FOXG1* (n=85), *SCN2A* (n=46), *SCN8A* (n=72), *STXBPI* (n=70), and *SYNGAPI* (n=65). Mean age of participants was 8.5 ± 5.6 years; there were no significant differences in age or sex between DEE cohorts. Mean number of medical record pages reviewed per patient ranged from 791.9 ± 673.4 (*SYNGAPI*) to 2034.9 ± 2696.1 (*SCN8A*); individuals with an *SCN8A* variant had significantly more pages than those with variants in *FOXG1* ($P=0.0273$) and *STXBPI* ($P=0.0001$). This volume of data represents an average of 7.4 ± 5.0 years of observed data per patient, totaling 2490 years. We reviewed 9565 extracted clinical concepts to evaluate relative enrichment between cohorts in the first years of life. We observed significant enrichment for "microcephaly" and "strabismus" in *FOXG1*; "apnea" and "respiratory distress" in *SCN8A*; "infantile spasms" and "ataxia" in *STXBPI*; and "gastroesophageal reflux disease" in *SYNGAPI*. Through ages 0-4 years, the five DEEs shared 12.8% of clinical concepts, with *FOXG1* and *SCN8A* showing the highest number of unique concepts (12.0% and 12.8%, respectively). Mean lifetime seizure severity scores were lowest in *STXBPI* and highest in *SCN8A*. Scores for *SCN8A* were significantly greater than *FOXG1* ($p<0.0001$), *SCN2A* ($P=0.0167$), *STXBPI* ($p<0.0001$), and *SYNGAPI* ($p<0.001$). Seizure severity scores were also significantly different across cohorts grouped by age, primarily in years 0-12. *SCN8A* and *STXBPI* were both associated with early and severe onset, with scores at 0 years significantly greater than subsequent years. In contrast, *SYNGAPI* was associated with childhood onset, and scores in childhood were significantly greater than those in infancy. This study demonstrates the utility of real-world data in describing clinical phenotypes in rare disease populations, highlighting age-dependent phenotypic distinctions between five genetic DEEs.

Mendelian Phenotypes Posters - Wednesday

PB1792. Compound heterozygosity for variants in *DCAF1* in fraternal twins with neurodevelopmental disorder

Authors:

J. Pappas, R. Rabin; NYU Grossman Sch. of Med., New York, NY

Abstract Body:

We present fraternal twins with autism spectrum disorder (ASD), intellectual disability (ID) and compound heterozygosity for variants in the *DCAF1*. *DCAF1* codes for the CRL4 substrate receptor protein DDB1/CUL4-associated factor 1 involved in many cellular processes (Schabla NM et al, 2019). The DDB1/CUL4 E3 ubiquitin ligase complex is involved in methylation-dependent ubiquitination (Lee et al., 2012). Several variants in E3 ligase genes have been associated with neurological conditions (George AJ et al, 2018). De novo variant in *DCAF1* has been reported in possible association with ASD and ID in a case report (Clothier JL et al, 2022). Our patients are boy-girl twins born after 38-week gestation via cesarean section to 38 year old G2P1 without perinatal complications. Both children had delayed speech development and some signs of autism in their second year. Our evaluation at 6 revealed that the girl had ID functioning at 3-4 year old level. The boy was less delayed but had attention deficit and hyperactivity. Both children had frequent behavioral outbursts. Both parents had typical development. Chromosome microarray and DNA test for fragile-X were reported negative. Whole exome sequencing with samples from both parents and both twin children revealed three variants of uncertain clinical significance (VUS). Maternally inherited VUS c.680 G>C p.(R227P) was reported in *SH2B1* in both children. This variant is likely noncontributory to the developmental disorder of the twins because it is present in the unaffected mother and the twins do not present with severe early-onset obesity and insulin resistance as affected individuals with *SH2B1* related disorder (Li et al., 2016). Both twins were compound heterozygotes for VUSs in *DCAF1*. The paternally inherited variant p.(Ala270Thr) (GCA>ACA): c.808 G>A in exon 8 of the *DCAF1* (NM_014703.2) has been observed 0.0699% in the population cohort gnomAD and it is unlikely to change the protein structure/function by in silico analysis (PROVEAN). The maternally inherited variant p.(Leu209Pro) (CTG>CCG): c.626 T>C in exon 8 of the *DCAF1* (NM_014703.2) has not been observed in gnomAD and it is predicted to change the protein structure/function by in silico analysis (PROVEAN). The compound heterozygosity for *DCAF1* variants in our patients suggests possible association of *DCAF1* with neurodevelopmental disorders. Our patients share ASD, ID and frequent behavioral outbursts with the case with de novo variant in *DCAF1* (Clothier JL et al, 2022). Functional studies to establish pathogenicity of the *DCAF1* variants and additional case reports are needed to validate this possible association of *DCAF1* with neurodevelopmental disorders.

Mendelian Phenotypes Posters - Thursday

PB1793. *Compound heterozygous IFT81 variants causing brachydactyly with cone-shaped epiphyses and generalized metaphyseal dysplasia*

Authors:

E. Carter; Hosp. for Special Surgery, New York, NY

Abstract Body:

Skeletal dysplasias are a heterogeneous group of >450 disorders impacting the size and shape of the skeleton. The ciliopathies with major skeletal involvement comprise a single group within the current skeletal dysplasia nosology. IFT81 is core protein in the intraflagellar transport complex B (IFT-B), which is involved in anterograde transport in the cilium. Most patients reported with skeletal dysplasia due to recessive variants in *IFT81* have a short-rib thoracic dysplasia phenotype (OMIM 617895). We report 2 unrelated individuals with recessive *IFT81* gene variants and a milder skeletal dysplasia phenotype of brachydactyly with cone-shaped epiphyses, generalized metaphyseal dysplasia, and progressive growth delay. Patient 1 is a now 13-year-old girl with a history of progressive short stature and metaphyseal dysplasia who was referred to our service for work-up and genetic analysis when she was 6 years old. Radiographic review detected generalized metaphyseal dysplasia, and brachydactyly with cone-shaped epiphyses. Initially her height plotted at the 10th percentile, decreasing to -3 SD with time (mid-parental height 70"/177.8cm). Prior consultations had ruled out hypochondroplasia (*FGFR3* negative). Panel testing detected compound heterozygous pathogenic *IFT81* variants: p.Leu29Phe (paternally-inherited) and splice acceptor c.249-1G>A (maternally-inherited). Patient 2 is a now 27-year-old man referred at age 23 years for diagnosis by the spine surgeon treating him for symptomatic spinal stenosis. Radiographic review when he was 8 years old detected generalized metaphyseal dysplasia and brachydactyly with cone-shaped epiphyses. Early childhood height measurements plotted ~50th percentile decreasing with age. His final adult height is 4'4"/133cm (mid-parental height 5'4.5"/63.7cm). Prior consultations had ruled out cartilage hair hypoplasia (*RMRP* negative). Panel testing detected compound heterozygous pathogenic *IFT81* variants: p.Arg512* (maternally-inherited) and p.Leu29Phe. Paternal DNA was unavailable for confirmatory testing. We present clinical and X-ray characteristics of 2 patients with a milder skeletal dysplasia phenotype without polydactyly or any visceral or congenital anomalies due to recessive *IFT81* variants. Interestingly, both variants detected in patient 1 have been reported before in a child with asphyxiating thoracic dystrophy who died at 19 months of age due to progressive respiratory compromise. Functional studies of the cilia and IFT components in our patients would be useful for further characterization of the pathogenesis of their diagnosis.

Mendelian Phenotypes Posters - Wednesday

PB1794. Compound heterozygous variants in *UFMI* gene in a case of hypomyelinating leukodystrophy

Authors:

B. Kang¹, M. Wagner¹, E. Ames²; ¹Univ. of Michigan, Ann Arbor, MI, ²Univ MICHIGAN Hlth.System, Ann Arbor, MI

Abstract Body:

UFMI-related hypomyelinating leukodystrophy is a severe, early-onset, autosomal recessive neurodevelopmental disorder characterized by developmental delay, hypotonia, absent speech, and hypomyelination. Some of the patients also have additional features such as seizures, hearing loss, or blindness. To date, only 20 patients have been reported and 16 of them are of Roma descent and the other four of them are from two consanguineous Sudanese families. Here we present the case of a seven-month-old Caucasian male with a hypomyelination by MRI, developmental delay, hypotonia, delayed visual maturation, feeding difficulty, absence of speech, abnormal breathing, mild pulmonary valve stenosis, and an abnormal electroencephalogram without definitive epileptiform abnormalities. The patient's parents are healthy and denied consanguinity. There is no significant family history. A chromosomal microarray was normal and two heterozygous variants of unknown significance in the *UFMI* gene (c.215G>A (p.Gly72Glu) and c.277C>T (p.Arg93Trp), NM_001286704.1) were identified by leukoencephalopathy sequencing and deletion/duplication panel testing of 697 genes via Invitae. Subsequent parental testing confirmed these variants to be in *trans*. *UFMI* is located on chromosome 13 and encodes ubiquitin-fold modifier 1 (UFM1) protein. UFM1 protein is a ubiquitin-like protein that is conjugated to target proteins by E1-like activating enzyme UBE1DC1 encoded by the *UBA5* gene and E2-like conjugating enzyme UFC1 in a manner analogous to ubiquitylation. Loss-of-function is a known disease mechanism. The paternally inherited variant (p.Gly72Glu) has not been reported in the general population (gnomAD database), in the published literature, or variant database associated with leukodystrophy (ClinVar and HGMD). The maternally inherited variant (p.Arg93Trp) has been reported in the general population (13 heterozygous alleles in the gnomAD database) and also reported in one affected individual of Roma descent with *UFMI*-related hypomyelinating leukodystrophy. Both parental variants result in alteration of highly conserved amino acids across species. *In silico* tools (SIFT, MutationTaster, PolyPhen2, MCAP, and PredictSNP2) predict a deleterious effect for both variants. To our knowledge, it is the very first *UFMI*-related hypomyelinating leukodystrophy case with compound heterozygous missense variants. Research on the molecular basis of the disease and the various signaling pathways involved will help improve the diagnosis and management of patients with hypomyelinating leukodystrophy, as well as develop more precise genetic counseling and screening tests.

Mendelian Phenotypes Posters - Thursday

PB1795. Confirmation of association of TGFBI p.Ser591Phe mutation with variant lattice corneal dystrophy

Authors:

D. Chung¹, C. Choo¹, K. Ledwitch², A. Kassels¹, J. Meiler², A. Aldave¹; ¹Stein Eye Inst., Univ. of California, Los Angeles, Los Angeles, CA, ²Ctr. for Structural Biology, Vanderbilt Univ., Nashville, TN

Abstract Body:

Purpose: To provide the initial confirmation of the c.1772C>T (p.Ser591Phe) mutation in the transforming growth factor-beta-induced (*TGFBI*) gene as being associated with variant lattice corneal dystrophy (LCD).

Methods: Ophthalmologic examination of the proband was performed with slit lamp biomicroscopy. Saliva was collected as a source of DNA for screening all 17 exons of *TGFBI*, after which three family members were selectively screened for identified presumed pathogenic variants. Rosetta-based structure prediction was used to calculate changes in TGFBI protein (TGFBIp) stability secondary to the c.1772C>T (p.Ser591Phe) missense mutation.

Results: Slit lamp examination of the 38-year-old proband revealed a clear cornea in the right eye and unilateral, discrete and branching lattice lines in the anterior and mid-stroma of the central cornea in the left eye. Screening of *TGFBI* in the proband revealed a heterozygous missense mutation in exon 13 (c.1772C>T (p.Ser591Phe)) that was also identified in her affected mother but not in her unaffected brother or maternal grandmother. Calculated energy changes in Rosetta (ΔG) for the TGFBIp variant p.Ser591Phe was 23.5, indicating a thermodynamic destabilization resulting from energetic frustration.

Conclusions: The p.Ser591Phe mutation in TGFBIp is associated with a unilateral variant of LCD. Rosetta-predicted stability changes indicate that this mutation is destabilizing, which is consistent with the observations of decreased protein stability in other LCD-associated mutations.

Mendelian Phenotypes Posters - Wednesday

PB1796. Congenital Anomalies Cluster: Linking clinicians and model system experts.

Authors:

S. Twigg¹, D. J. Henderson², S. Wells³, N. D. E. Greene⁴, P. Mill⁵, K. J. Liu⁶; ¹Oxford Univ., Oxford, United Kingdom, ²Newcastle Univ., Newcastle, United Kingdom, ³MRC Natl. Mouse Genetic Network, Mary Lyon Ctr. at MRC Harwell, Harwell, United Kingdom, ⁴Univ. Coll. London, London, United Kingdom, ⁵MRC Human Genetics Unit, The Univ. of Edinburgh, Edinburgh, United Kingdom, ⁶King's Coll. London, London, United Kingdom

Abstract Body:

As part of the new UKRI MRC National Mouse Genetics Network we have launched a research cluster, *Congenital Anomalies: patient-led functional genomics*, to accelerate basic research into the genetic causes and pathomechanisms underlying early life anomalies. Approximately 1 in 20 babies are born with severe anatomical malformations, such as cleft lip/palate, spina bifida and heart anomalies. Each year this equates to 8 million affected newborns, with lifelong impacts for patients and families. With advances in sequencing technology and the advent of large genome sequencing programmes, the identification of possible disease-causing variants in these patients is rapidly increasing. However, it remains a challenge to determine the causal variant, and to establish mechanistic links between genetic changes and anatomical anomalies. Many of the genes implicated in congenital anomalies play multiple roles in different tissues before and after birth. These genes are difficult to study in humans, even in 'disease-in-a-dish' models. Our goal is to make precisely-engineered mouse models of patient variants, which will help to replicate complex interactions disrupted during early life, across multiple organ systems with a focus on craniofacial, neural, cardiac and kidney anomalies. We also aim to improve live monitoring of that critical perinatal window in our disease models, which currently remains technically challenging thus obscuring roles for these candidates in those first few hours of life. Importantly, we will align our deep phenotyping of these mouse models with human pathology to better understand disease progression with age but also to serve as platforms for developing much needed therapeutic interventions. This is a community-led initiative looking to bring together human geneticists, rare disease specialists, pathologists and developmental biologists. Here we discuss how we will: 1) prioritise clinically relevant variants submitted to our pipeline; 2) apply deep phenotyping to investigate multi-system abnormalities; 3) analyse phenotypic pleiotropy; 4) integrate with the wider Mouse Genetic Network and other stakeholders; and 5) improve discussions on genetic cause and effect together with clinical geneticists, medical teams, patients and families. In summary, with our pipeline, we hope to provide a route for establishing and investigating pathogenicity, and for improved understanding of penetrance, phenotypic pleiotropy and environmental influence in congenital anomalies.

Mendelian Phenotypes Posters - Thursday

PB1797. Co-occurring anomalies in cases with achondroplasia

Authors:

C. Stoll; Faculté de Médecine, Strasbourg, France

Abstract Body:

Co-occurring congenital anomalies may be observed in cases with achondroplasia. The prevalence reported in the literature and the types of co-occurring congenital anomalies are variable between the reported studies. The aim of this study was to establish the prevalence and to describe the co-occurring anomalies in cases with achondroplasia. This study included 25 cases ascertained from our registry of congenital anomalies including all terminations of pregnancy, stillbirths and live births between 1979 and 2007 in 387,067 births (the prevalence of achondroplasia was 6.4 per 100,000 births), and 223 cases ascertained from the French organization, Association des Personnes de Petite Taille (APPT) built on the model of LPA (Little People of America, Inc.). Out of these 248 cases of achondroplasia 37 (14.9 %) had associated anomalies including 4 (1.6 %) cases with chromosomal abnormalities (2 trisomies 21, one 22 q11.2 deletion, and one 47, XXX), 2 (0.8%) cases with recognizable non-chromosomal conditions (one Moebius syndrome and one Pierre Robin sequence) and 31(12.5%) cases with MCA (multiple congenital anomalies). The 31 patients with MCA had 54 anomalies. Anomalies in the central nervous system (25.9%), the urogenital system (20.3%), the cardiovascular system (16.6%), the musculoskeletal system (12.8%), the eye (9.2%), and the orofacial system (7.4%) were the most common MCA. The overall prevalence of associated anomalies shows that the cases with achondroplasia need a careful screening for other congenital anomalies.

Mendelian Phenotypes Posters - Wednesday

PB1798*. C-X3-C motif chemokine receptor 1 (*CX3CR1*) gene variants associated with lesion burden in Cerebral Cavernous Malformation

Authors:

S. Weinsheimer¹, J. Nelson², B. Gongol³, M. C. Mabray⁴, J. M. Zabramski⁵, A. Akers⁶, B. L. Hart⁴, L. Morrison⁷, A. A. Abl⁸, N. U. Ko⁹, C. Tsang², C. E. McCulloch¹⁰, H. Kim¹¹, M. A. Lopez-Ramirez¹²; ¹Dept. of Anesthesia and Perioperative Care, Ctr. for Cerebrovascular Res., Inst. for Human Genetics, Univ. of California San Francisco, San Francisco, CA, ²Dept. of Anesthesia and Perioperative Care, Ctr. for Cerebrovascular Res., Univ. of California San Francisco, San Francisco, CA, ³Inst. for Integrative Genome Biology, Univ. of California Riverside, Riverside, CA, ⁴Dept. of Radiology, Univ. of New Mexico, Albuquerque, NM, ⁵Dept. of Neurosurgery, Barrow Neurological Inst., Phoenix, AZ, ⁶Angioma Alliance, Charlottesville, VA, ⁷Dept. of Neurology, Univ. of New Mexico, Albuquerque, NM, ⁸Dept. of Neurological Surgery, Univ. of California San Francisco, San Francisco, CA, ⁹Dept. of Neurology, Univ. of California San Francisco, San Francisco, CA, ¹⁰Dept. of Epidemiology and Biostatistics, Univ. of California San Francisco, San Francisco, CA, ¹¹Dept. of Anesthesia and Perioperative Care, Ctr. for Cerebrovascular Res., Inst. for Human Genetics, Epidemiology and Biostatistics, Univ. of California San Francisco, San Francisco, CA, ¹²Dept. of Med. and Pharmacology, Univ. of California San Diego, San Diego, CA

Abstract Body:

Background: Familial cerebral cavernous malformation (FCCM) is an autosomal dominant disorder caused by mutations in *KRIT1*, *CCM2* and *PDCD10* genes. Clinical heterogeneity of FCCM suggests a role for genetic modifiers of lesion burden and neuroinflammation. CCM mouse transcriptomics studies from major cell types involved in CCM disease, such as astrocytes and CCM endothelium, have identified a role for hypoxia-related genes. However, it is unknown whether these genes contribute to disease severity phenotypes, including intracranial hemorrhage (ICH), total and large lesion count in FCCM. We hypothesized that common genetic variants in hypoxia-related genes are associated with ICH and lesion burden in FCCM. **Methods:** FCCM cases (mean age 39±21y, 62% female, 93% White) enrolled in the Brain Vascular Malformation Consortium were included (n=338). Total and large lesions (≥5mm) were counted on MRI; clinical history of ICH at enrollment was assessed by medical records. Individuals were genotyped using the Affymetrix Axiom Genome-Wide LAT1 Human Array. We tested the association of 5,909 single nucleotide polymorphisms (SNPs) mapping within 5kb of 309 hypoxia-related genes identified from CCM mouse cell-type-specific transcriptomes with minor allele frequency (MAF) ≥1% using multivariable logistic regression for ICH and multivariable linear regression for log transformed total and large lesion count, adjusting for age, sex, and 3 principal components to adjust for population stratification. Significance was based on Bonferroni adjustment for multiple comparisons (P=0.05/5,909 SNPs=8.5E-06). **Results:** Two missense SNPs in *CX3CR1*, T2080M (rs3732378 G>A, MAF = 0.15) and V249I (rs3732379 C>T, MAF 0.23), were significantly associated with reduced CCM lesion count (Beta_{T2080M} = -0.5, 95% Confidence Interval = -0.71 to -0.29, P = 3.7E-06; Beta_{V249I} = -0.4, 95% Confidence Interval = -0.60 to -0.24, P = 7.0E-06). Thus, individuals with the T2080M variant have a 39% reduction in lesion count. These functional SNPs affect CX3CR1 protein activity and are associated with several neuroinflammatory disorders. No SNPs were associated with ICH or large lesion size. **Conclusions:** Two missense SNPs in the hypoxia-related *CX3CR1* gene are associated with total lesion count in FCCM. Further work is needed to confirm these findings in a replication cohort. These data in addition to functional studies in animal models support a role for *CX3CR1* inflammatory genetic modifiers in specific cell types that influence lesion burden, perhaps by altering the CX3CR1-CX3CL1 signaling, and provide an example of how genetic markers may be used for patient stratification for CCM severity.

Mendelian Phenotypes Posters - Thursday

PB1799. De novo (deep-)intronic splicing mutations in patients with neurodevelopmental disorders

Authors:

B. van der Sanden^{1,2}, G. Schobers^{1,2}, J. Corominas Galbany^{1,3}, G. Khazeeva^{1,3}, R. Derks¹, D. Koolen^{1,3}, M. Sinnema⁴, M. Nelen¹, H. Brunner^{1,2,4,5}, C. Gilissen^{1,3}, A. Hoischen^{1,3}, L. Vissers^{1,2}; ¹Dept. of Human Genetics, Radboud Univ. Med. center, Nijmegen, Netherlands, ²Donders Inst. for Brain, Cognition and Behaviour, Nijmegen, Netherlands, ³Radboud Inst. for Molecular Life Sci., Nijmegen, Netherlands, ⁴Dept. of Clinical Genetics, Maastricht Univ. Med. Ctr., Maastricht, Netherlands, ⁵GROW Sch. for Oncology and Dev.al Biology, Maastricht, Netherlands

Abstract Body:

Multiple studies have shown that de novo mutations (DNMs) are a major cause of unexplained neurodevelopmental disorders (NDD). However, these studies are often limited to coding variant interpretation. Whole genome sequencing (WGS) allows for comprehensive variant calling of all types of variation not limited to the coding sequence. Others previously showed that non-coding DNMs in conserved fetal brain-active regulatory elements may explain up to 0.15% of NDD. We hypothesized that intronic splice-altering variants could also have a pathogenic effect. In this study, we systematically assessed (deep-)intronic splice-altering DNMs and their potential pathogenic effect for NDDs. We performed WGS for a cohort of 156 exome-negative NDD patient-parent-trios. De novo single nucleotide variants (SNVs) were called using two algorithms, a standard in-house GATK-based algorithm and DeNovoCNN, a deep-learning based algorithm. DNMs called by both methods were marked as high confidence DNM and considered for this study. DNMs called by only one of the two methods were only considered when surpassing high quality thresholds. Splicing effects for all DNMs were predicted using SpliceAI, and variants with a SpliceAI delta score ≥ 0.02 were included in this study. In total, we identified 26,354 DNMs in 156 patients, resulting an average of 169 variants per patient. Of these, 12,083 were called by both algorithms and 14,271 by only one. On average, 63 DNMs affected a protein-coding gene, as defined by the start of 5'UTR, all intronic and coding sequence, up to the last base of the 3'UTR. Of these 70, only two were located in the gene's coding sequence. Filtering the variants located in a gene for predicted splice-altering effects based on SpliceAI prediction scores yielded a total of 633 DNMs (2.4%), of which 494 were located in intronic sequence, 94 in coding sequence, and the remaining 45 in either of the UTRs. In total, 68 of the intronic splice DNMs (0.26%) were located in one of 1,596 previously NDD associated genes. However, there was no enrichment of DNMs in the introns of NDD genes compared to the introns of other genes ($p = 0.326$, chi-squared test). We find that intronic splice DNMs in NDD associated genes occur at a rate of 0.44 events per genome. Using the recommended SpliceAI delta score threshold of ≥ 0.5 for pathogenicity, provides a potential diagnosis in 2/156 patients (1.3%). Pending functional confirmation of these variants in our patients, this suggests that de novo splice-altering mutations might be contributing to NDD to a larger degree than non-coding DNMs in conserved fetal brain-active regulatory elements.

Mendelian Phenotypes Posters - Wednesday

PB1800. *De novo* missense variants in *RPS5* cause a newly characterized ribosomopathy.

Authors:

F. Ladha¹, M. Balasubramanian², T. Juul³, L. Hansen³, C. Fagerberg³, Q. Hao³, Undiagnosed Diseases Network, J. Rosenfeld¹, L. Burrage¹, C. Murali¹; ¹Baylor Coll. of Med., Houston, TX, ²Univ. of Sheffield, Sheffield, United Kingdom, ³Odense Univ. Hosp., Odense, Denmark

Abstract Body:

RPS5 encodes one of eighty proteins that make up the mammalian ribosome. Ribosomopathies constitute a diverse spectrum of disorders. We detail three individuals with *de novo* missense variants in *RPS5* who present with dysmorphic features, developmental delay, hypotonia, poor growth, visual and ocular anomalies, and immunodeficiency. Patient 1 is a 12-year-old female with prenatal intrauterine growth restriction (IUGR). She had mild global developmental delay with hypotonia, immunodeficiency, myopia, relative short stature, learning disability, and dysmorphic features. Her variant in *RPS5* is NM_001009.3:c.593G>A (p.Arg198His). Patient 2 is a 21-month-old male with a prenatal history of small placenta. He has global developmental delay with hypotonia, short stature, poor growth, renal hypoplasia, and ocular anomalies with dysmorphic features. His *RPS5* variant is the same as Patient 1's. Patient 3 is an 8-year-old female with a prenatal history of IUGR. She has global developmental delay with hypotonia, intellectual disability, renal anomalies, bilateral sensorineural hearing loss, dysmorphic features, and hypogammaglobulinemia. Her *RPS5* variant is NM_001009.3: c.554C>T (p.Ser185Leu). Both variants are novel and located in exon 6, the final exon of *RPS5*. Both affect residues that are highly conserved through zebrafish. The variant in Patients 1 and 2 is located in the final turn domain of the protein, and Patient 3's variant is located in the final helix domain. The predicted constraint metric for missense variants in *RPS5* shows a Z-score of 2.79, indicating high intolerance to amino acid changes. Our findings suggest that *de novo* missense variants in *RPS5* may cause a newly recognized ribosomopathy. Further studies are needed to provide insight on the functional effects of these variants and whether exon 6 is a hotspot for variation associated with this novel ribosomopathy phenotype.

Mendelian Phenotypes Posters - Thursday

PB1801. *De novo* missense variants in *SEPHS1* Cause a Novel Neurodevelopmental Disorder with Developmental Delay, Hypotonia, Muscle Weakness, Speech Delay, and Growth delay

Authors:

S. Mullegama¹, K. Kiernan², E. Torti¹, J. Alaimo³, E. Pavlovsky⁴, N. Tilton⁴, A. Sekula⁵, H. Gao¹, M. Griffin⁴, R. Hickey⁶, D. Grange⁷, R. Slauch⁸, K. Engleman⁹, S. Sawyer¹⁰, E. Rush¹¹, E. Muller¹², Care4Rare Canada Consortium, H. Hare¹³, C. Phornphutkul¹⁴, M. Osmond¹⁵, L. Massingham¹⁶, K. Dipple¹⁷, UDN, B. Wolf¹⁸, L. Mehta¹⁹, B. Palculict²⁰, V. Fettig²¹, D. Miller¹⁷, A. Telegraf²², I. Glass²³, D. Carere²⁴, I. Wentzensen²⁵, M. Simonovic²⁶, M. Morrow¹, K. Monaghan²⁷, J. Juusola²⁸; ¹Gene DX, Gaithersburg, MD, ²UIC, Chicago, IL, ³Children s Mercy Hosp., Kansas City, MO, ⁴Coll. of Osteopathic Med., Sam Houston State Univ., Conroe, TX, ⁵SHSU-COM, Montgomery, TX, ⁶Lurie childrens, Chicago, IL, ⁷St Louis Child Hosp, Wash Univ, St Louis, MO, ⁸wust, St. Louis, MO, ⁹Children's Mercy Hosp, Kansas City, MO, ¹⁰Children's Hosp. of Eastern Ontario Res. Inst., Ottawa, ON, Canada, ¹¹CMH, Kansas City, MO, ¹²stanford childrens, San Francisco, CA, ¹³Children s Hosp. of Eastern Ontario, Ottawa, ON, Canada, Ottawa, ON, Canada, ¹⁴Rhode Island Hosp, Providence, RI, ¹⁵Children s Hosp. of Eastern Ontario, Ottawa, ON, Canada, ¹⁶8 Barberry La, N Easton, MA, ¹⁷Univ. of Washington, Seattle, WA, ¹⁸Lurie Childrens, Chicaog, IL, ¹⁹Columbia Univ. Med. Ctr., New York, NY, ²⁰GeneDx, Gaithersburg, MD, ²¹ICAHN, New York City, NY, ²²Gene Dx, Gaithersburg, MD, ²³Univ of Washington, Seattle, WA, ²⁴GeneDx, Delaware, ON, Canada, ²⁵GeneDx/Sema4, Bethesda, MD, ²⁶NIH, DC, DC, ²⁷GeneDx, Lake Orion, MI, ²⁸GeneDx, Clarksburg, MD

Abstract Body:

Selenophosphate synthetase (SEPHS) plays an essential role in selenium metabolism. SEPHS is an ATPase that synthesizes selenophosphate from ATP and selenide. As the primary selenium donor in the selenocysteine biosynthetic pathway, SEPHS is necessary for the efficient production of selenoproteins. Two mammalian SEPHS paralogues, *SEPHS1* and *SEPHS2*, share high sequence identity and structural homology with SEPHS. These two genes have yet to be implicated in human disease. It is conceivable that variants in *SEPHS1* (MIM 600902) or *SEPHS2* (MIM 606218) may be linked to a genetic disorder. Here, we report nine individuals with heterozygous missense variants in *SEPHS1*, all residing in the C-terminal domain in or near the AIR synthase-related domain. Eight of these individuals had a recurrent variant at amino acid position 371 of SEPHS1. Eight of these variants were known to be *de novo*. This cohort demonstrated a variable neurodevelopmental disorder characterized by developmental delay, hypotonia, muscle weakness, speech delay, and growth delay. Structural modeling and biochemical assays were utilized to understand the effect of these variants on SEPHS1 protein function. We found that a variant at residue W352 results in local structural changes of the C-terminal region of SEPHS1 that decrease the overall thermal stability of the enzyme. In contrast, variants of a solvent-exposed residue R371 do not impact enzyme stability and folding but could modulate direct protein-protein interactions of SEPHS1 with cellular factors involved in promoting cell proliferation and development. Our study supports that SEPHS1 plays a critical role during human development and provides a basis for further investigation into the molecular mechanisms employed by SEPHS1. In summary, variants in *SEPHS1* are associated with a novel neurodevelopment disorder.

Mendelian Phenotypes Posters - Wednesday

PB1802. De novo non-synonymous *CTR9* variants are associated with motor delay and macrocephaly

Authors:

H. Suzuki¹, K. Aoki², K. Kurosawa³, K. Imagawa⁴, T. Ohto⁴, M. Yamada¹, T. Takenouchi¹, K. Kosaki¹, T. Ishitani²; ¹Keio Univ., Tokyo, Japan, ²Osaka Univ., Osaka, Japan, ³Kanagawa Children s Med Ctr, Yokohama, Japan, ⁴Univ. of Tsukuba, Tsukuba, Japan

Abstract Body:

[Background]*CTR9* is one of five genes that form the PAF1 complex, which binds to RNA polymerase II and plays critical roles in transcriptional elongation and transcription-coupled histone modifications including histone H3K4me3 and H3K36me3. We identified two de novo missense variants in *CTR9* gene (NM_14633.3) in two unrelated patients with macrocephaly, motor delay, and intellectual disability. To confirm the pathological significance of *CTR9*, we performed functional analysis using zebrafish.

[Patients]Patient 1 was a 4-year-old girl. At 4 years of age, her height was 107 cm (+0.7 SD), her weight was 17.0 kg (+0.1 SD), and her head circumference was 53.3 cm (+2.1 SD). She walked alone at 36 months and spoke meaningful words at 42 months. Her developmental quotient was 40 at the age of five years. An exome analysis revealed that the patient had a *de novo* heterozygous variant in the *CTR9* gene: c.74C>G, p.(Pro25Arg). Patient 2 was a 6-year-old girl. At 4 years of age, her height was 100.8 cm (-0.2 SD), her weight was 18.0 kg (+0.9 SD), and her head circumference was 54.5 cm (+3.8 SD). She walked alone at 19 months and spoke meaningful words at 18 months. An exome analysis revealed that the patient had a *de novo* heterozygous variant in the *CTR9* gene: c.45G>A, p.(Glu15Asp).

[Method]We generated a *ctr9*-knockout zebrafish using the triple-target CRISPR method. The rescue experiments were performed on *ctr9*-knockout zebrafish with human wild-type and mutant *CTR9* mRNA inserted. Overexpression studies were also performed on the wild type and each variant.

[Result]The *ctr9*-knockout zebrafish showed macrocephaly and poor movement. Those with wild-type mRNA inserted in rescue experiments returned to normal, whereas the co-injection of the two hCTR9 variants did not exhibit such rescue effects. Overexpression studies showed no phenotypic change in the wild type, but the mutant was able to reproduce macrocephaly. The results suggest that the mutants detected have a dominant-negative effect.

[Discussion]The mechanistic link between the aberrant function of *CTR9* and macrocephaly associated with intellectual disability is unknown. We speculated that a reduction in lysine-36 trimethylation of histone H3 (H3K36me3) may be responsible, since H3K36me3 levels were also downregulated in *ctr9*-KO zebrafish [Conclusion] We concluded that the two missense variants in *CTR9* (p.(Glu15Asp) and p.(Pro25Arg)) cause a new syndrome involving macrocephaly, motor delay, and intellectual disability through the loss of the normal function of *CTR9* and the inhibition of the normal intrinsic *CTR9* function of the contralateral allele.

Mendelian Phenotypes Posters - Thursday

PB1803. *De novo* truncating *ADNP* variants as a recurrent cause of neurodevelopmental disorders

Authors:

M. Gusic, S. Eck, O. Wachter, J. Philippou-Massier, K. Hörtnagel; MVZ Martinsried GmbH / Med.ver Genetics, Martinsried, Germany

Abstract Body:

Neurodevelopmental disorders (NDDs) are defined as a heterogenous group of conditions affecting the brain and resulting in impaired cognition, behaviour, language and motor function, as well as additional (non)-neurological symptoms¹. Increased implementation of large-scale sequencing methods in the last decade has revealed a significant enrichment of *de novo* variants in individuals with NDDs^{2,3}. By performing parent-offspring trio exome sequencing as a diagnostic approach in a genetic diagnostics center, we report three unrelated individuals carrying *de novo* pathogenic variants in the *ADNP* gene, encoding for a transcription factor involved in the SWI/SNF remodeling complex⁴. The reported individuals presented global developmental delay and intellectual disability (ID) in common, in addition to other various craniofacial and muscular symptoms, reflecting the typical manifestations of the *ADNP*-associated Helsmoortel-Van der Aa syndrome⁵. All three variants (the most prevalent one and two novel) were truncating, located in the fifth (the last) exon of the gene and thus predicted to escape the nonsense-mediated decay. In addition to the three case reports, we provide an overview of clinical and genetic spectrum of the *ADNP*-related disorder. By doing so, we highlight the *ADNP* gene as one of the most frequent causes of neurodevelopmental disorders as well as the clinical utility of trio sequencing for the rapid discovery of pathogenic variants.

Literature:

1. Moreno-De-Luca et al. Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence (2013) *Lancet Neurol* 12(4):406-14.
2. Short et al. *De novo* mutations in regulatory elements in neurodevelopmental disorders (2018) *Nature* 555(7698):611-616.
3. Brunet et al. *De novo* variants in neurodevelopmental disorders-experiences from a tertiary care center (2021) *Clin Genet* 100(1):14-28.
4. Helsmoortel et al. Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis (2014) *Nature Genet* 46(4):380-384.
5. Van Dijck et al. Clinical Presentation of a Complex Neurodevelopmental Disorder Caused by Mutations in *ADNP* (2019) *Biol Psychiatry* 85(4):287-297.

Mendelian Phenotypes Posters - Wednesday

PB1804. Deciphering a Diagnostic Odyssey of Atypical Free Sialic Acid Storage Disorder Associated with Tissue Specific Mosaic Variant in *SLC17A5*

Authors:

M. Shinawi¹, D. Wegner¹, A. Paul¹, J. Sheehan¹, S. Jaiprakash¹, S. Marco¹, J. Cooper¹, W. Buchser², S. Sukipolvi¹, K. Sisco¹, M. Schroeder³, K. Bowling¹, T. Turner⁴, R. Mitra¹, D. Grange⁴, D. Baldrige⁴, T. Wang⁵, U. Undiagnosed Diseases Network¹, J. Wambach⁶, P. Dickson¹; ¹Washington Univ. Sch. of Med., Saint Louis, MO, ²Washington Univ. Sch. of Med., St Louis, MO, ³Washington Univ. Sch. of Med., Webster Groves, MO, ⁴Washington Univ. Sch. of Med., St. Louis, MO, ⁵Washington Univ. in St. Louis, St. Louis, MO, ⁶Washington Univ, St. Louis, MO

Abstract Body:

Background: Free sialic acid storage disorder (FSASD) is a rare autosomal recessive lysosomal storage disease caused by bi-allelic pathogenic variants in *SLC17A5*, which encodes sialin. Patients with FSASD exhibit a spectrum of phenotypes and severity that correlates with their genotype and free sialic acid levels. Because of its nonspecific symptoms and lack of routine biochemical markers, FSASD has a diagnostic lag and is likely underdiagnosed. **Methods:** We performed a multidisciplinary clinical assessment of a patient with suspected lysosomal storage disease of uncharacterized molecular etiology. We conducted comprehensive molecular studies, including deep sequencing and RNASeq, biochemical analyses and in vitro functional studies. **Results:** The proband is now a 15-year-old girl with mild to moderate intellectual disability, coarse facial features, multi-valvular heart disease, thoracolumbar scoliosis post-surgical repair, abnormally shaped ribs, surgically repaired pes cavus and tendo-Achilles tightness, restricted joint movements, periarticular subcutaneous nodules, gingival hypertrophy, and faint corneal whirling. Electron micrographs showed prominent cytoplasmic vacuolation in diverse cell types in skin and muscle. Clinical exome identified a maternally-inherited frameshift variant, c.533delC;p.Thr178Asnfs*34, in *SLC17A5*. Free sialic acid levels in urine were 860-1434 (nl:123-508). These findings prompted us to search for a second pathogenic variant in *SLC17A5*. Whole genome sequencing, exome and genome data reanalysis, and targeted sequencing and dosage studies of *SLC17A5* were negative. RNASeq in fibroblasts revealed exon 3 skipping in ~50% of reads on the paternal allele, which was not detected in blood or in her parents. Western blot using protein extracted from proband's fibroblast showed 2 bands corresponding to a predominant wild type protein (47 KD) and a smaller size protein (40 KD). Targeted capture/deep sequencing of fibroblast DNA revealed a 185 bp deletion in ~15% of reads, encompassing the 3' end of exon 3, including the donor splice site. **Conclusions:** We report an unusual case of FSASD with tissue specific mosaicism for a copy number variant. Significant residual sialin function in different tissues may explain the mild elevation of sialic acid and atypical clinical course and findings. Our work illustrates the difficulty in diagnosing FSASD and highlights the need to pursue advanced molecular technologies when there is a high index of suspicion of a specific phenotype with incomplete molecular diagnosis.

Mendelian Phenotypes Posters - Thursday

PB1805. Deciphering the phenotypic variability of the most common deafness-causative *GJB2* p.V37I variant in Asia

Authors:

Y-T. Chiang¹, P-H. Lin², M-Y. Lo³, H-L. Chen², C-Y. Lee¹, C-Y. Tsai⁴, Y-H. Lin⁴, P-L. Chen⁵, J. Hsu¹, C-C. Wu⁶; ¹Natl. Taiwan Univ., Coll. of Med., Taipei City, Taiwan, ²Natl. Taiwan Univ. Hosp., Dept. of Otolaryngology, Taipei City, Taiwan, ³Natl. Taiwan Univ. Hosp., Dept. of Otolaryngology, Taipei city, Taiwan, ⁴Natl. Taiwan Univ., Coll. of Med., Taipei, Taiwan, ⁵Natl. Taiwan Univ. Hosp., Taipei, Taiwan, ⁶Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Abstract Body:

The *GJB2* p.V37I (rs72474224) variant is the most prevalent genetic cause of sensorineural hearing impairment (SNHI) worldwide: approximately 500 million Asian individuals are homozygous for *GJB2* p.V37I. The phenotypes associated with p.V37I homozygosity are widely variable, with most patients presenting slight to moderate SNHI. However, severe to profound SNHI are occasionally observed in 10-20% of the patients (e.g., 10%, 57/561 in our Taiwanese cohort). To investigate the genetic factors contributing to the phenotypic variabilities, we performed targeted sequencing for *GJB2* and other deafness genes, including gap junction genes *GJB4*, *GJA1*, *GJB1*, *GJB3*, and *GJB6* in 121 p.V37I homozygotes. We identified at least one additional pathogenic variant in other deafness genes (*TBC1D24*, *COL4A3*, *SIX1*, *GRHL2*, *CDH23*, *NARS2*) in six patients who presented profound SNHI. We randomly selected p.V37I homozygotes with severe-to-profound SNHI (n=36), p.V37I homozygotes with mild-to-moderate SNHI (n=79), and subjects from the Taiwan Biobank (TWB, n=120) for case-control association analyses. We found the severe-to-profound group had a higher frequency of a *CRYL1* synonymous variant (rs14236), which is located in the upstream cis-regulatory element of *GJB2*, compared to the mild-to-moderate ($p < 0.02$; OR = 2.69; 95% CI, 1.06-7.23) and TWB group ($p < 8.34 \times 10^{-7}$; OR = 5.16; 95% CI, 2.58-10.35) respectively. Although digenic inheritance of *GJB2* and other gap junction genes had been documented previously, we did not detect any significant gap junction variants across the three groups. Our results uncover that additional pathogenic variants in other deafness genes and the *CRYL1* rs14236 variant may contribute to more severe hearing phenotypes in *GJB2* p.V37I homozygotes, underscoring the importance of the comprehensive genomic survey for these patients in clinical diagnostic practice.

Mendelian Phenotypes Posters - Thursday

PB1806. Deep resequencing of the 1q22 locus in non-lobar intracerebral hemorrhage

Authors:

L. Parodi^{1,2,3}, M. E. Comeau⁴, G. J. Falcone⁵, D. Woo⁶, J. Rosand^{1,2,3}, C. D. Langefeld⁴, C. D. Anderson^{7,1,2,3}; ¹Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ²McCance Ctr. for Brain Hlth., Massachusetts Gen. Hosp., Boston, MA, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Dept. of Biostatistics and Data Sci., Div. of Publ. Hlth.Sci., Wake Forest Sch. of Med., Winston-Salem, NC, ⁵Div. of Neurocritical Care and Emergency Neurology, Dept. of Neurology, Yale Sch. of Med., New Haven, CT, ⁶Dept. of Neurology, Univ. of Cincinnati, Cincinnati, OH, ⁷Dept. of Neurology, Brigham and Women's Hosp., Boston, MA

Abstract Body:

Background: Genome-wide association studies have identified *Iq22* as a susceptibility locus for non-lobar intracerebral hemorrhages (ICH), arising following the rupture of small vessels in deep structures of the cerebral hemisphere, brainstem and cerebellum. We performed targeted high-depth sequencing of *Iq22* in ICH cases and controls to further characterize this locus and prioritize potential causal mechanisms.

Methods: 95,000 base pairs spanning *Iq22*, including *SEMA4A*, *SLC25A44* and *PMF1/PMF1-BGLAP* were sequenced in 1,055 ICH cases (534 lobar and 521 non-lobar) and 1,078 controls. Because *Iq22* is a susceptibility locus for non-lobar ICH alone, non-lobar ICH patients were compared with lobar ICH patients and ICH-free controls pooled together, with sensitivity analyses comparing non-lobar ICH with only ICH-free controls. Analyses included: 1) Firth regression to assess the role of common variants, 2) z-score based RIFT analysis to assess the role of rare variants, 3) chromatin interaction using ChIA-PET data from ENCODE and Hi-C data (analyzed using Juicebox), and 4) multivariable Mendelian randomization of expression data from the eQTLGen Consortium to assess whether alteration in gene-specific expression relative to regionally co-expressed genes could be causally related to non-lobar ICH risk at *Iq22*.

Results: Common and rare variant analyses both prioritized variants in *SEMA4A* 3'-UTR and *PMF1* intronic regions, overlapping with active promoter and enhancer regions based on ENCODE annotation. ChIA-PET data analysis highlighted the presence of long-range interactions between the two *SEMA4A*-promoter and *PMF1*-enhancer regions previously prioritized. Hi-C data analysis clarified that the *Iq22* locus is spatially organized in a single chromatin loop and that the harbored genes belong to the same Topologically Associating Domain (TAD), with potentially shared expression regulation. Mendelian randomization analyses revealed that only overexpression of *PMF1*, controlling for co-expression of other genes across the TAD, could be causally related to the higher risk of non-lobar ICH observed at *Iq22* ($p = 0.004$). *PMF1* codes for Polyamine Modulated Factor-1, a regulator of polyamine catabolism.

Conclusion: Single variant and gene-based analyses of targeted sequencing data combined with orthogonal methods revealed potential causal mechanisms for the established associations between variants at *Iq22* and non-lobar ICH. Based on our findings, we hypothesize the increased promoter-enhancer interactions leading to *PMF1* overexpression, potentially causing the dysregulation of polyamine catabolism, as a possible underlying causal mechanism.

Mendelian Phenotypes Posters - Wednesday

PB1807. Defining the clinical heterogeneity of *GARS1*-related Charcot-Marie-Tooth disease

Authors:

S. Marte, A. Antonellis; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Aminoacyl-tRNA synthetases (ARSs) are essential enzymes required for charging tRNA molecules to cognate amino acids in the cytoplasm and mitochondria. Although ARSs are essential and ubiquitously expressed, loss-of-function (LOF) missense and small, in-frame deletion mutations in five dimeric ARS enzymes have been associated with dominant peripheral neuropathy (also known as Charcot-Marie-Tooth disease [CMT]). CMT is a genetically and clinically heterogeneous inherited peripheral neuropathy characterized by the progressive loss of motor and sensory function. Mutations in glycyl-tRNA synthetase (*GARS1*) have been associated with distinct clinical phenotypes where individuals present with either later-onset CMT or early-onset infantile spinal muscular atrophy. While protein translation and the integrated stress response have been implicated in *GARS1*-related CMT, the mechanism by which mutations in *GARS1* lead to distinct clinical phenotypes is unclear. Since all five implicated ARSs function as dimers and since the majority of CMT-associated ARS variants cause a loss-of-function effect, we propose the possibility of a dominant-negative mechanism. To test dominant-negative effects of pathogenic glycyl-tRNA synthetase (*GARS1*) variants, we will develop a humanized yeast model and test *GARS1* mutations for the ability to repress a wild-type copy of *GARS1*. To better understand the distinct clinical phenotypes, we will assess the dominant toxicity of a series of pathogenic *GARS1* alleles to determine if toxicity in our yeast model correlates with disease severity. Finally, we will identify pathways that, when manipulated, improve *GARS1* function by performing experimental evolution and gain-of-function studies using a hypomorphic *GARS1* allele and yeast growth assays. Here, we will present our unpublished data on development of our humanized model and on initial assessments of pathways that can improve *GARS1* function. These studies will aid in defining the molecular mechanism of *GARS1*-related CMT and in identifying the genetic underpinnings responsible for the observed clinical heterogeneity.

Mendelian Phenotypes Posters - Thursday

PB1808*. Defining the nuclear genetic architecture of a maternally-inherited mitochondrial disorder.

Authors:

R. Boggan¹, Y. Ng¹, L. S. Taylor^{1,2}, I. G. Franklin¹, C. L. Alston¹, E. L. Blakely¹, B. Büchner³, E. Bugiardini⁴, K. Colclough⁵, G. S. Gorman¹, C. Feeney¹, M. G. Hanna⁴, A. T. Hattersley⁶, T. Klopstock^{7,8}, C. Kornblum⁹, M. Mancuso¹⁰, K. A. Patel⁶, R. D. S. Pitceathly⁴, C. Pizzamiglio⁴, H. Prokisch^{11,12}, J. Schäfer¹³, A. M. Schaefer¹, M. H. Shepherd¹⁴, A. Thaele¹⁵, D. M. Turnbull¹, C. E. Woodward¹⁶, R. McFarland¹, R. W. Taylor^{1,2}, H. J. Cordell¹⁷, S. J. Pickett¹; ¹Wellcome Ctr. for Mitochondrial Res. and Inst. for Translational and Clinical Res., Newcastle upon Tyne, United Kingdom, ²NHS Highly Specialised Mitochondrial Diagnostic Lab., Newcastle upon Tyne Hosp. NHS Fndn. Trust, Newcastle upon Tyne, United Kingdom, ³Dept. of Neurology Friedrich-Baur-Inst., Univ. Hosp. of the Ludwig-Maximilians-Univ. (LMU Klinikum), Munich, Munich, Germany, ⁴Dept. of Neuromuscular Diseases UCL Queen Square Inst. of Neurology and The Natl. Hosp. for Neurology and Neurosurgery, London, United Kingdom, ⁵Exeter Genomics Lab. Royal Devon and Exeter NHS Fndn. Trust, Exeter, United Kingdom, ⁶Inst. of BioMed. and Clinical Sci. Univ. of Exeter Med. Sch., Exeter, Devon, United Kingdom, ⁷Dept. of Neurology Friedrich-Baur-Inst. Univ. Hosp. of the Ludwig-Maximilians-Univ. (LMU Klinikum), Munich Cluster for Systems Neurology (SyNergy), Munich, Germany, ⁸German Ctr. for Neurodegenerative Diseases (DZNE), Munich, Germany, ⁹Dept. of Neurology Univ. Hosp. Bonn, Bonn, Germany, ¹⁰Neurological Inst. of Pisa, Pisa, Italy, ¹¹Inst. of Human Genetics Sch. of Med. Technische Univ. München, München, Germany, ¹²Inst. of Neurogenomics Helmholtz Zentrum München - German Res. Ctr. for Environmental Hlth., Neuherberg, Germany, ¹³Dept. of Neurology, Univ. sklinikum Carl Gustav Carus, Technische Univ. Dresden, Dresden, Germany, ¹⁴Inst. of BioMed. and Clinical Sci., Univ. of Exeter Med. Sch., Exeter, Devon, United Kingdom, ¹⁵Dept. of Neurology, Martin-Luther-Univ. Halle-Wittenberg, 06120 Halle (Saale), Halle, Germany, ¹⁶Neurogenetics Unit, The Natl. Hosp. for Neurology and Neurosurgery, London, United Kingdom, ¹⁷Population Hlth.Sci. Inst., Newcastle Univ., Newcastle upon Tyne, United Kingdom

Abstract Body:

Mitochondrial function is under bi-genomic control; pathogenic variants in the nuclear and mitochondrial genomes (mtDNA) can result in clinical mitochondrial disease. The most common cause of multi-system adult mitochondrial disease (mtDNA variant m.3243A>G; NC_012920.1) is associated with extensive unexplained clinical heterogeneity. Variant m.3243A>G allele level, age and sex explain only a small proportion of this variability (pseudo-R² range = 0.05-0.26), whereas high to moderate estimates of heritability for some m.3243A>G-related phenotypes provide evidence for the influence of unidentified nuclear factors.

Using Haseman-Elston regression-based genetic linkage analysis in a cohort of 208 individuals from 83 pedigrees, we explored the nuclear genetic architecture of eleven phenotypes related to the m.3243A>G variant. For three phenotypes (migraine, cardiovascular involvement, and gastro-intestinal disturbance), no regions of interest were identified; simulation results suggest that any nuclear genetic contribution is highly polygenic in origin.

For seven phenotypes (cerebellar ataxia, chronic progressive external ophthalmoplegia, diabetes, myopathy, psychiatric disturbance, ptosis, and stroke-like episodes), at least one region of interest (LOD > 1.8) was identified. Seven regions of interest were identified for encephalopathy, including two (LOD > 3.3) on chromosomes 7 and 11, suggesting that a small number of nuclear factors play a key role in the development of this severe neurological phenotype.

This work describes the genetic architecture of the nuclear factors that influence m.3243A>G-related disease, revealing that different phenotypes are influenced by nuclear variation in different ways. Using

ASHG 2022 Annual Meeting Poster Abstracts

the results of this work to inform future studies will enable the further elucidation of the underlying genetic architecture of this complex disease via genome-wide association studies, polygenic scores, and whole-genome sequencing. This work will build a better understanding of disease development and progression, and will have a tangible impact on patients and patient care.

Mendelian Phenotypes Posters - Wednesday

PB1809. Delayed developmental milestones in *MBD5*-associated neurodevelopmental disorder (MAND) associated with 2q23.1 deletions highlight need for early diagnosis

Authors:

L. Zhan¹, S. H. Elsea², S. V. Mullegama¹; ¹Sam Houston State Univ. Coll. of Med., Conroe, TX, ²Baylor Coll. of Med., Houston, TX

Abstract Body:

Introduction: *MBD5*-associated neurodevelopmental disorder (MAND) is a group of conditions characterized by developmental delay, speech impairment, seizures, hypotonia, sleep disturbances, intellectual disability, and abnormal behaviors. MAND can be caused by haploinsufficiency due to single-nucleotide variants, microdeletions, or microduplications in *MBD5*, a dosage-sensitive gene which encodes a protein involved in gene activity regulation. Recent studies have shown that genes involved in speech and motor development are dysregulated in individuals with *MBD5* variants. Thus, this study focuses on characterizing motor and speech phenotypes in children with 2q23.1 deletions inclusive of *MBD5*. **Methods:** A survey was administered to caregivers (n=38) of children under 18 years of age with a heterozygous *MBD5* deletion confirmed through clinical genetic testing. Questions covered demographic information, milestone achievement, therapies received, gross and fine motor movement assessment, speech assessment, and other behaviors. **Results:** The mean age of diagnosis was 3.17±2.88 years. The mean age at the time of survey was 7.03±4.20 years. The majority of the cohort did not meet major developmental milestones for gross motor skills on time, with crawling and standing not achieved by 63.6% (20/32) and walking not achieved by 55.5% (19/35). Further, 84.3% (26/31) missed the speaking milestone. Multiple therapeutic interventions were engaged by all respondents, with 84.6% (32/38) receiving at least three types. These therapies aimed to improve motor control associated with feeding, walking, running, and full limb movements. Despite multiple approaches for therapeutic intervention, a majority of the cohort still reported difficulties with coordination, fine motor movements, and speech development. **Conclusions:** The motor phenotype observed with *MBD5* haploinsufficiency includes gait abnormalities, poor coordination, difficulty with fine motor control, and difficulty swallowing. Speech is markedly impaired, with severely delayed development and inappropriate control of tempo, volume, and pitch in children that are verbal. These missed milestones were apparent by 1 year of age, but most children were still not diagnosed after 3 years of age. These data highlight the need to define the underlying cause of MAND and target critical milestones earlier in the child's life. Earlier genetic evaluation for children who miss key milestones would lead to earlier diagnosis and would offer education and specific interventions to families navigating these complex syndromes, improving outcomes in these populations.

Mendelian Phenotypes Posters - Thursday

PB1810. Deleterious *SNAPC4* Variants are Associated with a Neurodevelopmental Disorder.

Authors:

F. G. Frost¹, M. Morimoto¹, P. Sharma¹, N. Belnap², D. Calame³, Y. Uchiyama^{4,5}, N. Matsumoto⁵, M. Oud⁶, E. A. Ferreira⁷, V. Narayanan², S. Rangasamy², M. Huentelman², K. Ramsey², L. T. Emrick^{3,8}, I. Sato-Shirai^{9,10}, S. Kumada⁹, N. I. Wolf¹¹, P. J. Steinbach¹², Y. Huang¹, Undiagnosed Diseases Network, J. L. Murphy¹, J. R. Lupski^{3,8}, G. Vezina¹³, E. F. Macnamara¹, D. R. Adams^{1,14}, M. T. Acosta¹², C. J. Tiff^{1,14}, W. A. Gahl^{1,1}, M. V. Malicdan^{1,14}; ¹NIH Undiagnosed Diseases Program, Bethesda, MD, ²Translational Genomics Res. Inst., Phoenix, AZ, ³Baylor Coll. of Med., Houston, TX, ⁴Yokohama City Univ. Hosp., Yokohama, Japan, ⁵Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, ⁶Radboud Univ. Med. Ctr., Nijmegen, Netherlands, ⁷Amsterdam Univ. Med. Ctr., Amsterdam, Netherlands, ⁸Texas Children's Hosp., Houston, TX, ⁹Tokyo Metropolitan Neurological Hosp., Tokyo, Japan, ¹⁰Shimada Ryoiku Ctr. Hachioji for Challenged Children, Tokyo, Japan, ¹¹Amsterdam UMC, locatie VUmc, Amsterdam, Netherlands, ¹²NIH, Bethesda, MD, ¹³Children's Natl. Med. Ctr., Washington, DC, ¹⁴NIH Office of the Clinical Director, Bethesda, MD

Abstract Body:

Introduction: Small nuclear RNAs (snRNAs) play integral roles in RNA splicing and are critical for cell function. Transcription of snRNAs is initiated by the small nuclear RNA activating protein complex (SNAPc), which binds to a promoter element upstream of snRNA genes and recruits transcriptional machinery. Impaired transcription of SNAPc subunits has been shown to reduce snRNA transcription; however, the role of SNAPc in mammalian cells remains to be elucidated. Moreover, variants in the genes encoding SNAPc have not been associated with human disease. Here, we present 6 individuals from 5 families with bi-allelic deleterious variants in *SNAPC4*, which encodes a key SNAPc subunit.

Methods: Clinical and biochemical phenotyping was performed on all probands. Exome or genome sequencing of the probands and available family members was performed to identify candidate disease-causing variants. The functional impact of *SNAPC4* variants on RNA splicing was assessed using patient-derived fibroblasts and a *SNAPC4*-deficient HeLa cell line, generated using CRISPR-Cas9 technology. **Results:** Six individuals with bi-allelic *SNAPC4* variants presented with a progressive spasticity disorder, which was characterized by normal development until one year, when developmental delays began to be evident, as well as spasticity. Affected individuals also showed cerebellar or cerebral atrophy on brain MRI, deep tendon reflex abnormalities, dysarthria, and gait alterations. Bi-allelic *SNAPC4* variants segregated with disease, most of which are likely loss-of-function. Premature stop codons were introduced by 5 variants, including 1 nonsense and 4 intronic, which were validated by cDNA sequencing of patient-derived cell lines. All 5 missense variants were predicted to be deleterious by *in silico* tools, with 3 located in the DNA-binding domain. Patient cells showed decreased SNAPC4 and snRNA expression and disruptions in global splicing patterns. *SNAPC4*-deficient HeLa cells showed similar changes, confirming those consequences were specific to *SNAPC4* deficiency. **Conclusion:** Our work supports the role of *SNAPC4* in snRNA transcription and splicing in mammalian cells through functional studies in HeLa and patient-derived cells. The common clinical phenotypes, alteration of *SNAPC4* expression, and global splicing dysregulation in patient cells support bi-allelic variants in *SNAPC4* as a compelling explanation for disease in our cohort. The etiology of neurological disease may be linked to the importance of alternative splicing in the development and function of the nervous system, but future experiments will be required to assess this hypothesis.

Mendelian Phenotypes Posters - Wednesday

PB1811. Deletions of 14q32.2 result in severe neurodevelopmental outcomes and multiple congenital anomalies: three affected males and review of the literature.

Authors:

J. Black¹, M. de Koning², R. Lebel¹, S. Smith¹, M. Byler¹, A. Van Haeringen², C. Ruivenkamp², H. Goel^{3,4}; ¹SUNY Upstate Med. Univ., Syracuse, NY, ²Leiden Univ. Med. Ctr., Leiden, Netherlands, ³HNELHD-Hunter Genetics, Newcastle, Australia, ⁴Univ. of Newcastle, Callaghan, Australia

Abstract Body:

Deletions of 14q32.2 have previously been reported, and most often involve the imprinted region of chromosome 14, giving rise to paternal or maternal UPD(14)-like phenotypes. Few individuals with deletions that spare the imprinted region have been reported. Deletion size and gene involvement have been highly variable.

We report three males with similar size deletions (2.5 to 3.9 Mb), two of which do not involve the imprinted region. These deletions overlapped for 13 different genes, three of which are associated with autosomal dominant conditions: *BCL11B*, *CCNK*, and *YY1*. All three patients present with prenatal and postnatal growth restriction, severe intellectual and developmental disability, feeding problems, cryptorchidism, hypotonia, and similar dysmorphic features (microstomia, micrognathia, wide nasal bridge, small nasal passages, and low set ears). The two with deletions sparing the imprinted region shared additional features including recurrent infections, strabismus, and kidney problems.

Other reported patients with similar 14q32.2 deletions show some clinical overlap with our patients, but none have the same combination or severity of features. Those with smaller deletions involving *BCL11B* and *CCNK* but not *YY1* had severe intellectual and developmental disability but no congenital abnormalities, while others with smaller deletions of only *YY1* had mild to moderate neurodevelopmental problems and minor dysmorphic features. We propose that deletions involving these three genes result in a discrete clinical entity entailing severe neurodevelopmental phenotype, characteristic facial features, and multiple congenital anomalies.

Mendelian Phenotypes Posters - Thursday

PB1812*. Delineation of gain-of-function MYCN-induced novel megalencephaly syndrome and possible out-of-brain complications implicated by a mouse model.

Authors:

Y. Nishio^{1,2,3}, K. Kato^{1,2,3}, H. Futagawa⁴, S. Masuda⁵, F. Tran Mau-Them⁶, S. Otsuji¹, C. Quélin⁷, H. Shawki⁸, H. Oishi⁸, T. Takenouchi⁹, K. Kosaki⁹, Y. Takahashi², S. Saitoh¹; ¹Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, ²Dept. of Pediatrics, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, ³Dept. of Genetics, Res. Inst. of Environmental Med., Nagoya, Japan, ⁴Dept. of Clinical Genetics, Tokyo Metropolitan Children's Med. Ctr., Tokyo, Japan, ⁵Dept. of Hematology and Oncology, Tokyo Metropolitan Children's Med. Ctr., Tokyo, Japan, ⁶Unité Fonctionnelle 6254 d'Innovation en Diagnostique Génomique des Maladies Rares, Pôle de Biologie, CHU Dijon Bourgogne, Dijon, France, ⁷Service de Génétique Clinique, CLAD Ouest, CHU Rennes, Hôpital Sud, Rennes, France, ⁸Dept. of Comparative and Experimental Med., Nagoya City Univ. Graduate Sch. of Med. Sci. and Med. Sch., Nagoya, Japan, ⁹Ctr. for Med. Genetics, Keio Univ. Sch. of Med., Tokyo, Japan

Abstract Body:

Background: *MYCN*, a member of the *MYC* proto-oncogene family, encodes a transcription factor that regulates genes promoting cell growth and proliferation. We previously reported two patients, a 15-year-old boy and an artificially-aborted fetus, with a novel megalencephaly syndrome with ventriculomegaly and polydactyly, who carried a gain-of-function variant in *MYCN* (p.Thr58Met and p.Pro60Leu, respectively). Although the concept of a novel megalencephaly syndrome was established through these patients, additional patients were needed to confirm it and delineate the clinical entity. Furthermore, although a mouse model exhibited a postaxial polydactyly and macrocephaly with increased number of neuronal cells, the comprehensive analysis of other organs was needed to understand possible out-of-brain complications. **Methods:** Exome sequencing was performed to identify a pathogenic variant. We also investigated the multiple organs of knock-in mice (*Mycn*^{WT/T58M}). **Results:** We identified a *de novo* *MYCN* missense variant (p.Thr58Met) in an 8-month-old boy. The patient exhibited megalencephaly, ventriculomegaly, postaxial polydactyly and dysmorphic facial gestalt, that were the very same phenotypes as those of the 1st patient. Furthermore, as with the 1st boy who was diagnosed with neuroblastoma that was successfully treated, the 3rd patient was diagnosed with neuroblastoma, that was already present at birth, when he was 3 months old. Through the analysis of a mouse model other than a macrocephaly and polydactyly, we identified the phenotypes in kidney and female reproductive systems possibly due to the over-proliferation of the tissues that forms these organs. **Conclusions:** These results delineate gain-of-function *MYCN*-induced novel megalencephaly syndrome. The out-of-brain phenotypes of the *Mycn* gain-of-function mouse model suggest the possible complications for clinicians carefully to examine.

Mendelian Phenotypes Posters - Wednesday

PB1813. Developmental Delay at the Single Cell Level in Prader-Willi Syndrome

Authors:

L. Reiter¹, D. Johnson¹, D. Garmire², C. Roco³, A. Victor¹; ¹Univ Tennessee HSC, Memphis, TN, ²Univ. of Michigan, Ann Arbor, MI, ³Parse BioSci.s, Seattle, WA

Abstract Body:

Prader-Willi Syndrome (PWS) is a neurodevelopmental disorder defined by a range of phenotypes including pronounced developmental delay, intellectual disability, sleep disorders and increased autism risk. Motor and language development is impaired in most PWS individuals and about 10-20% of PWS adults experience cycloid psychosis. The heterogeneity of the cognitive phenotypes in PWS warrant further study into the early neural development of PWS. We have established the largest collection of dental pulp stem cell (DPSC) lines from the deciduous teeth of PWS subjects. We regularly differentiate these stem cells into cortical-like neurons. Using neuronal cultures derived directly from DPSC of individuals with PWS and neurotypical controls, we performed single cell sequencing (scRNA-seq) via the SplitSeq pipeline (Parse Biosciences). Enrichment analysis on single neuron sets revealed shared pathways among the differentially expressed transcripts in PWS versus neurotypical controls, including the regulation of neurogenesis and neuronal differentiation. Expression of critical neurodevelopmental genes *NRXN1*, *NRXN2*, and *NAV2* was significantly increased in neurotypical control versus the PWS neurons. Several of these transcripts have now been validated at the protein level in independent control and PWS DPSC-derived neuronal lines. These preliminary results suggest that the PWS neurons display a delayed neurogenesis phenotype, at the single neuron level, that may explain the cognitive developmental delay in PWS and reveal potential therapeutic targets. We are currently increasing both the number of independent cell lines and the number of individual neurons being sequenced in order to further elucidate the mechanisms contributing to developmental delay in PWS.

Mendelian Phenotypes Posters - Thursday

PB1814. Diagnosis, treatment, and follow-up care for patients with Barth syndrome is presented through the results from 1:1 interviews of pediatric cardiologists.

Authors:

D. Curtis¹, M. Windrem², G. Kelly³; ¹Curtis Analytic Partners, Philadelphia, PA, ²Stealth BioTherapeutics, Needham, MA, ³Stealth BioTherapeutics, Needham, MA

Abstract Body:

Background: Barth syndrome (BTHS) is a rare genetic disease characterized by numerous impairments which are associated with substantial morbidity and mortality driven by heart disease. BTHS often manifests at an early age. To improve outcomes, additional information is needed regarding the level of knowledge pediatric cardiologists have identifying and managing patients with BTHS. After conducting a survey of 200 pediatric cardiologists to increase understanding of BTHS patient management, we conducted interviews with a small subset of those physicians who recently managed at least 1 BTHS patient.

Methods: Pediatric cardiologists were invited via email to participate in interviews, followed by a phone screening. Qualitative 1:1 interviews were conducted virtually in October 2021. Interviews were approximately 1-hour in duration and included discussion around BTHS patient engagement, management, and referrals to additional physicians.

Results: Most (n=4/6) respondents were currently involved with 1 or 2 BTHS patients, with the remaining respondents (n=2/6) having notable previous involvement. First patient interaction occurred at ages ranging from prenatal through toddler. Patient involvement typically began in the hospital setting, with the most serious phenotype initially presenting with the following symptoms: neutropenia, low muscle tone, low white blood cell count, skeletal abnormalities, renal dysfunction, arrhythmia, high potassium, failure to thrive, cardiomyopathy, and heart failure. The number of BTHS patients treated remained constant at 1-3 per year. At time of diagnosis, additional medical care team included ECHO cardiologists, neonatologists, and neurologists. Initial diagnostic tests included complete blood work, ECHO, genetic testing, and cardiopipin analysis. Various medications (i.e., ACE inhibitors, digoxin, and beta blockers) and supplements were used for medical management of BTHS. Respondents noted that after leaving the hospital, patient management was variable with most stable BTHS patients returning every 3 months for follow-up. Most respondents (n=5/6) identified the general pediatrician as the overall medical manager of patients with BTHS.

Conclusions: Pediatric cardiologists are aware of, diagnose, and treat patients with the most serious presentation of BTHS (i.e, heart disease/cardiomyopathy) and, along with pediatricians, would benefit from additional training to recognize and diagnose milder phenotypes. To increase understanding of trends in BTHS diagnosis and management, additional interviews will be conducted with a larger population of medical professionals and BTHS caregivers.

Mendelian Phenotypes Posters - Wednesday

PB1815. Diagnostic rate and clinical utility of whole genome sequencing in adults with intellectual disability

Authors:

A. Sabo, S. Pereira, S. Dugan, P. Trotter, M. Gingras, D. Murdock, B. Yuan, D. Muzny, R. Gibbs; Baylor Coll. of Med., Houston, TX

Abstract Body:

Establishing a genetic diagnosis for individuals with intellectual disability (ID) benefits patients and their families as it may inform the prognosis, lead to appropriate therapy, and facilitate access to medical and supportive services. Exome and genome sequencing are routine in a diagnostic setting for pediatric patients, but many adults with ID are lacking a comprehensive genetic evaluation. The families are not readily accessing these tests due to a lack of access to care, insurance coverage, or knowledge of new opportunities.

Our unique recruitment strategy involved partnering with service and education providers as well as three clinical sites. We utilized genome sequencing and analysis, identified SNVs, indels, and CNVs, and performed clinical variant interpretation for each recruited family. Families where genome sequencing analysis identified pathogenic or likely pathogenic variants discussed the results and any potential medical follow-up with a board-certified clinical geneticist. We formally evaluated clinical utility, including measures related to diagnosis, patient clinical management, and familial and psychosocial implications using C-GUIDE questionnaire.

We were able to establish a molecular diagnosis in 45% (13/29) of individuals. Our results include *de novo* variants in *MED13L*, *POGZ*, *BRAF*, *EHMT1*, *SATB2*, and *DNMT3A*, homozygous or compound heterozygous variants in *TUSC3* and *VRK1*, and heterozygous variants with unknown inheritance in *KANSL1*, *SATB1*, *ASH1L*, *CHD4*, and *TBL1XR1*. Clinicians completed a C-GUIDE questionnaire for five individuals. A measure of utility of the genetic testing scores ranged from 16-21; the mean score was 18.6 (SD=2.1; on a scale of -2 - 32, higher scores indicating more utility). When asked whether the genetic testing prompted better care for the patient or their family overall, clinicians reported it prompted better care in three cases and somewhat better care in one case; the clinician was unsure in the remaining case.

Our study suggests a high yield of genome sequencing as a diagnostic tool in adult patients with ID who have not undergone comprehensive sequencing based genetic testing. Currently, medical insurance providers rarely cover the expense of exome or genome sequencing testing for adults due to perceived limited clinical utility. Research studies like ours that include assessment of clinical utility will be critical in the evaluation of the usefulness of genetic testing.

Mendelian Phenotypes Posters - Thursday

PB1816. Diagnostic utility of comprehensive genomic and transcriptomic profiling of anorectal malformations.

Authors:

S. Ramadesikan¹, M. Marhabaie¹, R. Supinger¹, R. Williamson¹, E. Varga¹, P. White^{1,2}, D. C. Koboldt^{1,2}, R. Wood^{3,4}, E. Mardis^{1,2}; ¹The Steve and Cindy Rasmussen Inst. for Genomic Med., Nationwide Children's Hosp., Columbus, OH, ²Dept. of Pediatrics, The Ohio State Univ. Coll. of Med., Columbus, OH, ³Dept. of Pediatric Colorectal & Pelvic Reconstructive Surgery, Nationwide Children's Hosp., Columbus, OH, ⁴Dept. of Surgery, The Ohio State Univ. Coll. of Med., Columbus, OH

Abstract Body:

Anorectal malformations (ARM) constitute a group of congenital defects of the gastrointestinal and urogenital systems. They affect males and females, with an estimated worldwide prevalence of 1 in 3,000 live births. These malformations are clinically heterogeneous and can occur as part of syndromic conditions (rare) or as a non-syndromic entity. Despite the well-recognized heritability of non-syndromic ARM, the etiology in most patients is unknown. As an international leader in pediatric colorectal/pelvic reconstructive surgery, our institution follows around 1,500 ARM patients and adds 150-200 cases per year. We recently established a multidisciplinary, IRB-approved translational research study to interrogate the molecular landscape of ARM through comprehensive molecular profiling of patients, family members, and (when available) surgically resected tissue. To date, we have enrolled and sequenced 16 individuals from four multiplex families representing some rare, and diverse ARM types, as well as 3 surgically resected tissue specimens. In one of the families, we identified a novel paternally inherited heterozygous *CDX2* variant (c.722A>G (p.Glu241Gly)), that was present in all 3 affected siblings. *CDX2* encodes a transcription factor and is considered the master regulator of gastrointestinal development. The variant is absent from population databases and predicted to be damaging by most *in silico* pathogenicity tools. It maps to the homeobox domain of the protein, which is critical for interaction with DNA targets. So far, only 2 other reports implicate variants in *CDX2* with ARM. Remarkably, patients described in these studies exhibit similar clinical phenotypes and genetic alterations in *CDX2*. Our finding provides a potential molecular diagnosis for our family's condition and supports the role of *CDX2* in anorectal anomalies. It also highlights the clinical heterogeneity and variable penetrance of ARM predisposition variants, a well-documented phenomenon. Finally, it underscores the diagnostic utility of genomic profiling of ARM to identify the genetic etiology of these defects.

Mendelian Phenotypes Posters - Wednesday

PB1817. Dilated cardiomyopathy in a 3-month-old female: The answer in genetics.

Authors:

M. Patel, Y. Martinez-Fernandez, J. Bolivar, P. Jayakar; Nicklaus Children's Hosp., Miami, FL

Abstract Body:

Introduction:Dilated cardiomyopathy (DCM) in infancy has a broad differential diagnosis leading to challenges in work-up, treatment, and counseling. Herein described is a case of a 3-month-old female who presented with DCM due to a rare genetic variant.

Case:A 3-month-old female with no significant medical history, presented with decreased PO intake for 48hr, cough, and increased work of breathing for 5 weeks. She has history of a URI one month prior. Tachycardia and a gallop were noted on physical exam. CXR showed cardiomegaly with vascular congestion. She had abnormal troponin and BNP. TEE revealed severe MR, markedly reduced LV systolic function, and severely dilated LA and LV. She was started on a heart failure regimen with clinical improvement. CTA showed a normal origin of coronary arteries. Infectious workup (CMV, Adenovirus, Coxsackie, Enterovirus, Hepatitis, Parvovirus, Toxoplasma) was negative. Cardiomyopathy panel positive for -Variants of Uncertain Significance (VUS), HRAS- c.284A>G (p.Gln95Arg and RYR2- c.7955T>C (p.Leu2652Pro)- both inherited from an asymptomatic mother.

WGS (whole genome sequencing) TRIO- identified 2 VUS- compound heterozygous variants in the autosomal recessive RPL3L gene; maternally inherited missense variant, c.173G>A (p.Arg58Gln) and paternally inherited missense variant, c.922G>A (p.Asp308Asn). The patient continues with depressed LV function, is followed outpatient with cardiology and genetics, and is referred to the heart transplant team.

Discussion:DCM in infancy is often idiopathic. Even as more genes are identified, the yield of genetic testing remains low, especially in nonsyndromic cases. RPL3L encodes for the 60S ribosomal protein L3-like protein, which is highly expressed in striated muscles.

Ganpathy et al in 2020 reported neonatal-onset severe cardiomyopathy, with rapid progression to cardiac decompensation and death unless the patient undergoes heart transplantation. Similar to our case, Ganpathy describes a patient of 2.5 month old with emesis and tachypnea, found to have severely reduced LV function requiring a heart transplant by 5 months of age. A multidisciplinary team approach and early heart transplant referral are key. This case highlights the importance of WES/WGS in the evaluation, treatment, and counseling of infantile DCM.

Conclusion:Most cases of DCM in infancy are idiopathic, limiting disease-specific treatment, followed by inborn errors of metabolism. Early diagnosis is imperative to treatment, such as early referral for heart transplant. We therefore suggest genetic evaluation should be part of the initial evaluation of DCM with non-syndromic DCM in infancy.

Mendelian Phenotypes Posters - Thursday

PB1818. *Drosophila* model of *de novo* *MRTF-B* variants highlights the critical role of actin regulation.

Authors:

J. Andrews¹, J-W. Mok¹, O. Kanca¹, C. Tiffit², E. Macnamara³, B. Russell⁴, S. Yamamoto⁵, H. Bellen¹, S. Nelson⁶, M. Malicdan², M. Wangler⁷; ¹Baylor Coll. of Med., Houston, TX, ²NIH/NHGRI, Bethesda, MD, ³NIH, Bethesda, MD, ⁴UCLA Sch. of Med., San Luis Obispo, CA, ⁵Baylor Coll. of Med., HOUSTON, TX, ⁶UCLA Med Ctr, Los Angeles, CA, ⁷Baylor Coll. Med., Houston, TX

Abstract Body:

Myocardin-related transcription factor B (MRTFB) is a member of a family of genes which serve to potentiate serum response factor (*SRF*)-dependent transcription and is highly conserved in both vertebrate and invertebrate model organisms. MRTF-B is not currently associated with a human disease but has been shown to be highly expressed in all human tissues save the lung. Here we report a pair of probands with a *de novo* variants in the second RPEL domain of *MRTF-B* with intellectual disability, minor dysmorphic features, expressive language delay, impulse control issues, and fine motor delay. We have generated a *Drosophila* model of the variants (p.R104G and p.A91P) found in these probands using the UAS/Gal4 system to drive the expression of human cDNA within the fly. Expression of either of the putatively pathogenic variants within wing tissues via a *Nubbin-Gal4* driver was sufficient to induce significant morphological changes in the fly wing, including truncations of wing veins, expansion of intervein space, loss of crossveins, and blistering. Conversely, expression of a reference human *MRTF-B* cDNA produced only minor changes in the posterior crossvein. Expression of *Drosophila Mrtf* using the *Nubbin-Gal4* driver produced a similar change in crossvein length as was observed with the human reference. To identify if this change in wing morphology was due to a lack of RPEL domain functionality, a *UAS-MRTF-B^{ΔRP}* line was established which lacked the entire 2nd RPEL domain. Expression of this line within the flies wings resulted in damage which was indistinguishable from the changes in morphology caused by the expression of either the p.R104G or p.A91P variants. In *Drosophila*, the *SRF* ortholog, *blistered (bs)*, is known to suppress wing vein formation and promote the development of intervein cells. Exogenous co-expression of *bs* and *Mrtf* has been previously shown to significantly alter wing morphology; therefore we expressed the human reference and p.R104G variant cDNA lines concurrently with a *UAS-bs* line. We found that wing morphology was highly disrupted when *bs* and the reference human cDNA were co-expressed, while the co-expression of p.R104G variant cDNA and *bs* was lethal. As the interaction between *MRTFB* and *SRF* is dependent on actin binding within the RPEL domain, *MRTFB* variants could disturb actin binding. We identified a significant decrease in the ability of the p.R104G variant to bind to actin, suggesting that a lack of regulation may cause its effects on the fly wing. Our results show that the p.R104G *MRTFB* variant disrupts actin binding and underlies a novel disorder.

Mendelian Phenotypes Posters - Wednesday

PB1819. Effect of FOXO3 and Air Pollution on Cognitive Function: A Longitudinal Cohort Study of Older Adults in China

Authors:

J. Ji; Tsinghua Univ., Beijing, China

Abstract Body:

Forkhead Box O 3 (FOXO3) genotype is strongly associated with human longevity and may be protective against neurodegeneration. Air pollution is a risk factor for cognitive decline and dementia. We aimed to study the individual and combined effects of FOXO3 and air pollution on cognitive function in a large prospective cohort with up to 14 years of follow-up. We measured cognitive function and impairment using the Mini-Mental State Examination (MMSE). We used tagging SNPs rs2253310, rs2802292, and rs4946936 to identify the FOXO3 gene, of which roughly half of the population had the longevity-associated polymorphism. We matched annual average fine particulate matter (PM_{2.5}) concentrations within a 1 km² grid. We conducted cross-sectional and longitudinal analyses using multivariable linear and logistic regression models and generalized estimating equations. At baseline, carriers of the longevity-associated homozygous minor alleles of FOXO3 SNPs had a higher MMSE score than the carriers of homozygous major alleles. In the longitudinal follow-up, carriers of FOXO3 homozygous minor alleles had lower odds of cognitive impairment compared with noncarriers. Higher PM_{2.5} was associated with a lower MMSE score and higher odds of cognitive impairment. The positive effects of FOXO3 were the strongest in females, older people, and residents in areas with lower air pollution.

Mendelian Phenotypes Posters - Thursday

PB1820*. Effective Genetic Diagnosis for Unilateral and Asymmetric Pediatric Hearing Loss via Exome Sequencing

Authors:

S. Redfield¹, J. Perry¹, A. Oza¹, S. Rouse¹, C. Stewart¹, H. Khela¹, V. Albano¹, E. Shearer², M. Kenna¹; ¹Boston Children's Hosp., Boston, MA, ²Boston Children's Hosp., Dept. of Otolaryngology, Boston, MA

Abstract Body:

The current standard of care for the evaluation of bilateral symmetric pediatric sensorineural hearing loss (SNHL) includes genetic testing. Current clinical algorithms do not recommend genetic testing as an initial test for the etiological evaluation of asymmetric or unilateral SNHL, given previously reported low diagnostic rates for these populations. However, these studies largely did not focus on a pediatric population, include parental testing, use exome sequencing (ES), or have a large sample size. Our goal was to determine the diagnostic yield of ES for pediatric patient with asymmetric and unilateral SNHL compared to bilateral SNHL in a large cohort of pediatric patients. ES was performed from buccal-derived DNA for pediatric patients with confirmed SNHL without a known genetic or environmental etiology. Biological relatives were also tested (typically trios). ES mapping and variant calling, including copy number variants, was performed with the DRAGEN Bio-IT Platform (Illumina). Primary variant analysis focused on 366 known and candidate hearing loss genes. ES was performed for 218 probands and 333 relative (551 participants, including 130 trios). This cohort was clinically heterogenous in terms of SNHL phenotypes, including probands of varied laterality of SNHL, configuration of audiogram, age of onset of SNHL, and presence of other clinical features. A genetic cause of SNHL was identified for 31.2% of probands (n=68) with causative variants in 37 genes. The overall genetic diagnostic rate was 40.7% for bilateral, 23.1% for asymmetric, and 18.0% for unilateral, with syndromic diagnoses made in 20.8%, 33.3%, and 54.5% of cases in each group, respectively. In several cases of syndromic SNHL diagnosis, a genetic syndrome had not been suspected and no other clinical features were appreciated prior to study enrollment. We identified a genetic cause of SNHL in a significant percentage of pediatric patients with asymmetric and unilateral SNHL. Syndromic SNHL was more common in these cases compared to bilateral SNHL. Increased access to genetic testing for patients with all SNHL phenotypes will facilitate tailored intervention, early referral to appropriate specialists, and improved prognostic and recurrence information for families.

Mendelian Phenotypes Posters - Wednesday

PB1821. Ellis-Van Creveld Syndrome: Clinical and Molecular Analysis of 50 Individuals

Authors:

M. Aubert-Mucca¹, C. HUBER², G. Baujat¹, C. Michot¹, M. Zarhrate³, M. Bras⁴, L. Boutaud⁵, Clinical Contributors, V. Malan⁵, T. Attie-Bitach⁵, V. Cormier-Daire^{1,2}; ¹Reference Ctr. for Constitutional Bone Diseases, Genomic Med. and Rare Disorders Dept., Necker-Enfants Malades Univ. Hosp., AP-HP, Paris, France, ²Paris Cité Univ., INSERM UMR 1163, Imagine Inst., Necker Enfants Malades Univ. Hosp., Paris, France, ³Genomics Core Facility, Imagine Inst. - Structure Fédérative de Recherche Necker, INSERM U1163 and INSERM US24/CNRS UMS3633, Paris Cite Univ., Paris, France, ⁴Bioinformatics Platform, Imagine Inst.-Structure Fédérative de Recherche Necker, Paris Cite Univ., Paris, France, ⁵Genomic Med. and Rare Disorders Dept., Necker-Enfants Malades Univ. Hosp., AP-HP, Paris, France

Abstract Body:

Background:

Ellis-Van Creveld syndrome (EVC) is one of the entities belonging to the skeletal ciliopathies short-rib-polydactyly subgroup. Major signs are ectodermal dysplasia, chondrodysplasia, polydactyly, and congenital cardiopathy, with a high degree of variability in phenotypes ranging from lethal to mild clinical presentations. The *EVC* and *EVC2* genes are the major genes causative of EVC syndrome. However, an increased number of genes involved in the ciliopathy complex have been identified in EVC syndrome, leading to a better understanding of its physiopathology, namely *WDR35*, *GLII*, *DYNC2LII*, *PRKACA*, *PRKACB*, and *SMO*. They all code for proteins located in the primary cilia, playing a key role in signal transduction of the Hedgehog pathways.

Methods: The aim of this study was the analysis of 50 clinically identified EVC cases from 45 families, to further define the phenotype and molecular bases of EVC.

Results: Our detection rate in the cohort of 45 families was of 91.11%, with variants identified in *EVC/EVC2* (77.8%), *DYNC2H1* (6.7%), *DYNC2LII* (2.2%), *SMO* (2.2%), or *PRKACB* (2.2%). No distinctive feature was remarkable of a specific genotype-phenotype correlation. Interestingly, we identified a high proportion of heterozygous deletions in *EVC/EVC2* of variable sizes (26.92%), mostly inherited from the mother, and probably resulting from recombinations involving Alu sequences.

Conclusion: We confirmed that *EVC* and *EVC2* are the major genes involved in the EVC phenotype, and highlighted the high prevalence of previously unreported CNVs.

Mendelian Phenotypes Posters - Thursday

PB1822. Enrichment of *RAI1* genetic aberrations associated with sleep disturbances in Smith-Magenis syndrome and autism spectrum disorder.

Authors:

A. Kaden, J. Brzezynski, C. Johnson, S. Smieszek, C. Polymeropoulos, G. Birznieks, M. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC

Abstract Body:

Autism spectrum disorder (ASD) comprises a complex of neurodevelopmental disorders primarily characterized by deficits in verbal communication, impaired social interactions, and repetitive behaviors. Previous genetic studies have pointed to hundreds of presumptive causative or susceptibility variants in ASD, making it difficult to find common underlying pathogenic mechanisms and suggesting that multiple different genetic etiologies for ASD influence a continuum of traits.

Smith-Magenis syndrome (SMS) is a rare genetic disorder that results from an interstitial deletion of 17p11.2 and, in rare cases, from a retinoic acid induced 1 (*RAI1*) gene variant. The prevalence is estimated to be 1 in 15,000-25,000 people. Haploinsufficiency of *RAI1* is the primary cause of the neurobehavioral and metabolic phenotype in SMS. Individuals with SMS present with a distinct pattern of mild to moderate intellectual disability and, almost uniformly, significant sleep disturbances.

We conducted a large-scale association analysis of The MSSNG Project whole genome sequencing data to elucidate the prevalence of *RAI1* single-nucleotide variants (SNVs) and copy number variations (CNVs) in the ASD population. We accessed the MSSNG database hosting over 6,080 probands and queried both SNVs and CNVs.

We report a single case of a classic deletion (17p11.2 critical region) and an additional three cases of microdeletions. Moreover, we report two frameshift mutations and one splicing variant. Given that the frequency of SMS is 1 in 15,000 people in the general population, we observe a 2.5x enrichment of the major deletion (1 in 6,080 samples) and a > 5x enrichment of the frameshift variants (2 in 6,080 samples). In a set of 6,080 probands, we also observe 54 unique missense variants in 84 individuals within exon 3 of the *RAI1* gene.

Both ASD and SMS patients suffer from sleep disturbances. In this population of individuals with ASD, we report an enrichment of variants known to cause SMS. We estimate the enrichment to be at least 2.5-fold and potentially higher than 10-fold, considering the types of variants observed in the population. The sleep disturbances seen in individuals with SMS may also be the underlying mechanism for sleep disturbances in at least a subset of individuals with ASD, especially in those with consequential variants in the *RAI1* gene. Further studies are needed to help delineate the role of *RAI1* variants in sleep physiology.

Mendelian Phenotypes Posters - Wednesday

PB1823. Enzymatic testing for MPSI in Kuwait: A Pilot Study toward newborn screening.

Authors:

H. Alsharhan¹, A. Alnassar², A. Alyaqoub², M. Almaie², B. Qadoura², U. M. Elkazzaz², D. G. Ramadan², M. Ayed², B. Albash², R. Alsafi², M. Haider³, G. S. Dhaunsi¹, H. Alkandari², N. Makhseed²; ¹Kuwait Univ., Kuwait, Kuwait, ²Kuwait Ministry of Hlth., Kuwait, Kuwait, ³Kuwait Univ., Faculty of Med., Safat, Kuwait

Abstract Body:

Introduction: Mucopolysaccharidosis type I (MPSI) is an autosomal recessive lysosomal storage disorder characterized by deficiency or absence of α -L-iduronidase (IDUA) enzyme due to pathogenic variants in *IDUA* gene. Early treatment with hematopoietic stem cell transplantation and/or enzyme replacement therapy is associated with improved outcomes in this progressive multisystem disease. The diagnosis is usually delayed due to late presentation and nonspecific symptoms resulting in high morbidity and mortality. The world-wide prevalence of MPSI is estimated to be 1:100,000, however, it is unknown in Kuwait. This ongoing pilot study to screen MPSI involves all Kuwaiti neonates born in Farwaniya Hospital over a period of 12-months. This study is aimed to examine the incidence of MPSI in Kuwait for inclusion in our national newborn screening program to enable early detection and adequate treatment. **Methods:** All Kuwaiti neonates born at Farwaniya Hospital, Kuwait from December 2021 to April 2022 have been screened for MPSI. The screening consisted of determining IDUA enzyme activity in dried blood spots (DBS)-derived samples by Tandem Mass Spectrometry. A follow-up genetic analysis of *IDUA* gene is planned to screen the cases with diminished IDUA enzyme activity as second-tier testing. **Results:** A total of 268 Kuwaiti newborns including 139 (52%) males and 129 (48%) females were screened. None of them had deficient or absent IDUA enzyme activity and thus, no genetic testing was conducted. However, we have diagnosed one additional female baby with MPSI, who belongs to Farwaniya Hospital, but the parents chose to deliver in a private hospital. She presented at age three months with recurrent upper airway infections, snoring and extensive Mongolian spots. The molecular study revealed previously reported pathogenic nonsense variant in *IDUA* c.1882C>T; p.(Arg628Ter), associated with severe phenotype. That being included, MPSI is estimated to be about 0.8% among tested females and 0.4% of all screened cases in Kuwait. **Conclusion:** Our study is the first to evaluate the incidence of MPSI in Kuwait. Given the small number of screened babies and the short study duration thus far, it is premature to calculate the incidence of MPSI. As the study continues and more infants are screened, we will be able to estimate the incidence of the disease in our population correctly. Our data support including MPSI in national newborn screening program to allow early initiation of treatment and thus improve the outcome of the disease.

Mendelian Phenotypes Posters - Thursday

PB1824. ERF-Related craniosynostosis: further delineation of the new craniosynostosis syndrome

Authors:

R. Santillán Martínez, M. Abreu-González, L. Hernández-Ancheyta, S. Martínez-Gutiérrez, E. Rojas-Toledo, C. O. Azotla-Vilchis; Laboratorio de Biología Molecular y Secuenciación Masiva. Genos Médica, Centro Especializado en Genética, Mexico City, Mexico

Abstract Body:

ERF-Related craniosynostosis-4 (CRS4) is a multisutural synostosis predominant characterized by pansynostosis or sagittal/bilamboid with the Crouzonoid triad (ocular hypertelorism, exorbitism and malar hypoplasia). It may also present with Chiari Type 1 malformation, speech delay, learning problems, motor control deficiency, hyperactivity and poor concentration.

CRS4 has been recently recognized to be caused by pathogenic variants in *ERF* (MIM: *611888), coding for ETS2 repressor factor. ETS2 is a transcription factor and protooncogene involved in development, apoptosis, and regulation of telomerase. The protein encoded by *ERF* gene binds to the ETS2 promoter and is a strong repressor of ETS2 transcription.

In this study we present a 5-years-old Mexican female patient referred to genetic evaluation for facial dysmorphism. She has mild plagiocephaly, facial asymmetry, mild frontal bossing, hypertelorism, epicanthic folds, a broad nasal bridge, anteverted nares, malar hypoplasia, arched palate, low-set ears and a normal neurodevelopment. No other anomalies were detected. The family history was relevant, because the mother have the same phenotype.

Due to the clinical suspicion of an autosomal dominant entity, complete exome sequencing was performed (Twist Comprehensive Exome Panel). The bioinformatics pipeline analysis of 1,040 genes associated with facial dysmorphism and bone dysplasia identified the NM_006494.4:c.566_567delGT (p.Cys189*) pathogenic variant in *ERF*. This indel is present in ClinVar database (rs1555750816) and is predicted to result in a loss of the last 360 amino acids of the protein.

The overall prevalence of *ERF* mutations in patients with syndromic craniosynostosis is approximately 2%, and 0.7% in clinically no syndromic craniosynostosis [Glass et al 2019]. These frequencies support CRS4, as a new syndrome, is mostly under diagnosed. This case exemplifies that isolated craniosynostosis and mild crouzonoid phenotypes should be sequenced with a panel that includes *ERF*.

Mendelian Phenotypes Posters - Thursday

PB1826. Exome Sequencing in Understanding the Etiologies of Ataxia in Children

Authors:

S. Phadke, A. Shambhavi, H. Sait, S. Srivastava, A. Moirangthem; Sanjay Gandhi Postgraduate Inst. of Med. Sci., Lucknow, India

Abstract Body:

Background/ Objectives: Childhood ataxia can be due to acquired abnormalities or genetic causes. Ataxia can present in its pure form or in association with other neurological features such as spasticity, brainstem signs, or even a more complex phenotype with additional systemic and neurological features. Triplet-primed PCR (TP-PCR) is a method to detect triplet-repeat expansions which is seen in many genetic ataxias. Here, this study aims to provide information about the utility of exome sequencing in identifying single nucleotide variants in cases with ataxia and expand the genotypic spectrum in children presenting with ataxia in Indian population. **Methods:** Children presenting either with ataxia telangiectasia or complex rare ataxias with unknown etiology based on preliminary evaluation were enrolled. Detailed clinical evaluation and pedigree data was collected. Exome sequencing was done followed by genotype-phenotype correlation. Friedreich ataxia and other triplet repeat disorders confirmed by molecular testing were excluded. **Results:** Fifteen cases with childhood-onset ataxia were recruited with average age at clinical evaluation of 11 years. Consanguinity was present in 1 case. In 4 cases, a provisional diagnosis of ataxia telangiectasia was made, while others had no diagnosis based on clinical evaluation. Exome sequencing was performed where causative variant(s) were identified in 12 cases of which 5 cases had biallelic variants (2 homozygous and 3 compound heterozygous variants) in ATM gene. Out of these 8 variants in ATM gene, 3 were novel (c.5957delT, c.3576+2T>G, c.1236G>A). In 2 cases, variants in ERCC6 gene associated with Cockayne syndrome B were identified. In addition, causative variants (3 pathogenic/likely pathogenic and 2 VUS) were identified in 5 more cases. These included 4 genes for autosomal recessive disorders (C19orf12, PLA2G6, MTHFR, SNX14) and one X Linked gene (ATP2B3). SNX14 causes a rare form of spinocerebellar ataxia (SCA 20) was identified in a child with facial dysmorphism, intellectual disability and ataxic gait. The novel splice variant was identified in SNX14 gene which was functionally validated (PMID: 35195341). Out of 11 cases with autosomal recessive disorders, 7 from non-consanguineous families were homozygous. **Conclusion:** In this cohort of 15 childhood-onset ataxia we have identified variants in 12 cases. Six of the 14 variants identified were novel. Homozygosity is commonly seen in many nonconsanguineous families due to inbreeding. Additionally, our study expands the genotypic and phenotypic spectrum of the rare monogenic forms of ataxia. **Grant:** DBT (BT/PR26428/MED/12/783/2017)

Mendelian Phenotypes Posters - Wednesday

PB1827. Exome sequencing leads to the identification of a rare case of an autosomal dominant non-syndromic hearing disorder in a German family

Authors:

R. Birkenhager, S. Hollander, S. Weis, S. Arndt, A. Aschendorff, A. Knopf; Univ Freiburg, Freiburg, Germany

Abstract Body:

Hearing impairment is the most common sensorineural disorder in humans. Approximately 1 - 3 out of thousand newborns suffer from severe hearing loss or deafness at birth or in the first few years of life. Based on the form of the physiological defect, hearing disorders are classified into conductive hearing loss, sensorineural hearing loss, or a combination of both. Hearing disorders can be caused by environmental factors or viral infections, strong sources of noise, ototoxic substances and genetic causes. About ~ 60% of all prelingual hearing disorders are genetic. Inherited hearing disorders are divided into syndromal (SHL) or non-syndromal (NSHL). Almost 70% of cases of inherited hearing disorders are non-syndromic and mainly due to sensorineural causes. About 80% of the cases follow an autosomal recessive (DFNB) and 18% an autosomal dominant (DFNA) inheritance, about 2% are x-chromosomal (DFNX) or mitochondrial (MT) linked. A total of 187 gene locations have been described to date, for which 143 genes have so far been identified, 44 genes at least still unknown. Autosomal non-syndromic hearing loss (ANSHL) is a genetically heterogeneous sensorineural disorder, with prelingual hearing loss and absence of other clinical manifestations. Based on the clinical diagnosis it is not possible to recognize in which genes mutations are present. The aim of this study is to identify the pathogenic gene in a non-consanguineous German family over three generations, with seven affected members. Hearing testing BERA/Electrocochleography and radiological a high-resolution CT scan was made. Mutational analysis of two affected family members was performed using direct sequencing of the coding exon and intron transitions of the genes *GJB2* and *GJB6*, including deletion analysis. For investigation of autosomal nonsyndromic hearing loss genes, whole exome sequencing was performed, with the "INVIEW HUMAN EXOME" platform; array Agilent Genomics SureSelectXT All Exon V5. No mutations could be identified in the DNFB1 gene locus, containing the genes *GJB2* and *GJB6*. A further targeted analysis of other genes was not possible; therefore complete exome sequencing took place. All known genes for hearing impairment were analyzed. A heterozygous pathogenic variation was only detected in the gene *TMCI*, c.1249G> A, Gly417Arg, confirmed by Sanger sequencing. The inheritance pattern of the mutation in the family indicates a dominant non-syndromic hearing impairment. So far, mutations have been described in the *TMCI* (MIM 606706) gene which, on the one hand, leads to recessive (DFNB7/B11; MIM 600974) or dominant (DFNA36; MIM 606705) inherited hearing impairment.

Mendelian Phenotypes Posters - Thursday

PB1828. Exome sequencing uncovers novel variants in two Malian families with Epileptic encephalopathies

Authors:

A. Maiga¹, D. Mohamed Emile¹, S. Bamba², A. Yalcouye³, S. Diarra⁴, L. CISSE⁵, A. Tamega¹, C. Guinto^{1,6}, G. Landoure^{1,6}; ¹Univ. of Sci., Techniques and Technologies of Bamako, Bamako, Mali, ²USTTB, Bamako, Mali, ³Univ. of Sci., Techniques, and Technologies of Bamako, Mali, Bamako, Mali, ⁴NIH, ROCKVILLE, MD, ⁵Regional Hosp. of Segou, Segou, Mali, ⁶Univ. Hosp. "Point G", Bamako, Mali

Abstract Body:

Introduction : The term “Epileptic encephalopathy” refers to a heterogeneous group of brain affections in which the continuous epileptic activity leads to progressive deterioration of cerebral functions. It is mainly characterized by refractory seizures, severe electroencephalographic (EEG) abnormalities and global developmental delay or decline. Most epileptic encephalopathy cases are clinically sporadic, coinciding with frequently identified de novo variants with the application of next-generation sequencing (NGS). Here we report two novel variants in 2 Malian families with epileptic encephalopathies. **Aim:** To explore the genetic defects underlying Epileptic Encephalopathy in two Malian families **Methods:** We recruited probands and their families after formal consent. Detailed history was collected from each family to construct the pedigrees. Affected individuals went through a careful clinical assessment. When possible, EEG was performed. Peripheral blood was collected for DNA extraction. Whole Exome Sequencing (WES) was done, and Sanger Sequencing protocol is underway for segregation studies. **Results:** 16 participants, including seven affected individuals from two families (F1 and F2) with epileptic syndromes, were recruited. Tonic and/or clonic seizures associated with cognitive and developmental impairments were the main clinical features in patients. EEG was not performed in any of them due to limited access. Using WES, we found two pathogenic novel variants (SYNJ1(NM_003895.3): c.1255C>T in F1; SCN7A(NM_002976.4): c.4045C>A in F2). In silico prediction tools were used to confirm the pathogenicity of these variants. Segregation studies are underway through Sanger sequencing. **Conclusion:** Epileptic encephalopathies are one of the most heterogeneous groups of conditions, both clinically and genetically. We report here two de novo genetic alterations consistent with epileptic encephalopathy, expanding the genetic epidemiology of these conditions. This also highlights the important role of NGS in the diagnosis of complex neurological disorders. Further studies are necessary to evaluate the accuracy of NGS in African cohorts, but this opens the way to potential future therapeutics. **Keywords:** Epileptic encephalopathy, Genetics, NGS, Mali, Africa

Mendelian Phenotypes Posters - Thursday

PB1829. Expanding the Phenotypic spectrum of *SMARCA5*-related neurodevelopmental disorder

Authors:

R. Zambrano¹, H. Meddaugh²; ¹LSUHSC, New Orleans, LA, ²Children's Hosp. of New Orleans, New Orleans, LA

Abstract Body:

SMARCA5-related disorder is a newly described syndrome associated with variable neurodevelopmental differences, poor growth, and distinct dysmorphic craniofacial features. At least 10 affected families have been identified in the medical literature to-date. We describe a patient with a novel, de novo, likely pathogenic variant in *SMARCA5* (c.2172+1G>A) identified through clinical whole exome sequencing. The proband displays significant phenotypic overlap with the *SMARCA5* disease spectrum as described, including global developmental delay, autism spectrum disorder, microcephaly, and failure to thrive. However, the patient also displays additional features including central precocious puberty with advanced bone age, and cardiac anomalies, including atrial septal defect and cardiomyopathy, which have not been reported in the *SMARCA5* literature to-date. We propose an expansion of the phenotypic spectrum of *SMARCA5*-related neurodevelopmental disorder.

Mendelian Phenotypes Posters - Wednesday

PB1830. Expanding the phenotypic spectrum of *TAB2* related disorders.

Authors:

P. Gupta, E. Lopez, L. King, L. Fried; St. Joseph's Children's Hosp., Paterson, NJ

Abstract Body:

Introduction: Frontometaphyseal dysplasia (FMD) is a progressive skeletal dysplasia primarily associated with joint contractures, cervical vertebral fusion, and scoliosis. X-linked FMD caused by mutations in *FLNA* are responsible for 50% of all cases of FMD and termed FMD1. Mutations in *MAP3K7* cause autosomal dominant-FMD and termed FMD2. *TAB2* is a novel gene associated with AD-FMD with a small number of individuals exhibiting a wide range of symptoms with close relation to *MAP3K7*-related FMD. A proposed FMD3 for *TAB2*-FMD has been suggested. Additionally, *TAB2* is also associated with AD polyvalvular syndrome, which is characterized by cardiac valve defects, distinct facial features, short stature, and connective tissue abnormalities. Case Presentation: The patient was born to a 39 year old mother G3P2 at 37 weeks gestation, via repeat C-section. Prenatal history was unremarkable. Birth weight was 7 lbs 5oz. Family history was significant for polydactyly in patient's father, paternal grandfather, and paternal great-grandfather. Bilateral post axial polydactyly was noted at birth which were tied by 3 weeks of age. The patient was seen by a neurologist at 2 weeks of age for hypotonia and a head ultrasound noted mild asymmetry of the ventricles. A sacral dimple was noted on physical exam and an ultrasound of the spine ruled out tethered cord. The initial genetics evaluation was done at 7 weeks of age via telehealth -physical exam noted significant head lag, hypertelorism, flat nasal bridge and micro-retrognathia. A SNP microarray analysis, and SMA deletion/duplication analysis were normal. Follow up genetic evaluation at 1 year of age noted prominent forehead with frontal bossing, bitemporal narrowing, hypertelorism, low-set ears, overfolded pinnae, depressed nasal bridge, short nose, long smooth philtrum, thin upper lip, retrognathia, short sternum, protruberant abdomen and generalized joint laxity. Height was less than the 1st percentile. A skeletal dysplasia panel was ordered and revealed a heterozygous pathogenic variant in the *TAB2* gene, resulting in premature translational stop codon associated with polyvalvular syndrome and FMD. Discussion: The diagnosis of our patient adds to the phenotypic spectrum of *TAB2*-related frontometaphyseal dysplasia (FMD3). There is much overlap in facies when comparing to previous individuals diagnosed with FMD3. Our patient is the 4th individual to have features consistent with FMD3, adding on to the phenotypic spectrum in the literature.

Mendelian Phenotypes Posters - Thursday

PB1831. Expert perspectives on sequencing newborns for treatable genetic conditions

Authors:

N. Gold¹, S. Adelson², N. Shah³, S. Williams², S. Bick⁴, J. Gold⁵, A. Strong⁵, R. Ganetzky⁵, A. Roberts⁴, M. Walker¹, A. Holtz⁴, V. Sankaran⁴, O. Delmonte⁶, W. Tan¹, I. Holm⁴, J. Thiagarajah⁴, J. Kamihara⁷, J. comander⁸, E. Place⁹, J. Wiggs⁸, R. Green²; ¹Massachusetts Gen. Hosp., Boston, MA, ²Brigham and Women's Hosp., Boston, MA, ³Dartmouth Children's Hlth.Ctr., Lebanon, NH, ⁴Boston Children's Hosp., Boston, MA, ⁵Children's Hosp. of Philadelphia, Philadelphia, PA, ⁶NIH, District of Columbia, DC, ⁷Dana Farber Cancer Inst., Boston, MA, ⁸Massachusetts Eye and Ear Inst., Boston, MA, ⁹Massachusetts Eye & Ear, Boston, MA

Abstract Body:

Newborn screening (NBS) is a public health program that has decreased morbidity and mortality among infants with rare disorders. The Recommended Universal Screening Panel for NBS state labs includes 35 core and 26 secondary conditions, but targeted treatments and management approaches for hundreds of additional genetic disorders are available. Efforts such as the NIH-funded BabySeq Project explore the implications of sequencing in healthy infants with the aim of designing strategies for implementation. We designed an online survey for rare disease experts (n = 398), investigating which additional treatable genetic conditions they recommend screening in newborns using genomic sequencing. We also assessed attitudes about the inclusion of disorders that are untreatable, lack confirmatory orthogonal tests, are of low penetrance, or are adult-onset. A total of 227 (57.0%) experts, including directors of genetics and genomics programs accredited by the Accreditation Council for Graduate Medical Education (n = 60), clinicians and academicians specializing in the care of rare disease patients (n = 158), and senior scientists within pharmaceutical companies specializing in rare disease therapeutics (n = 9) participated. We analyzed concordance among experts regarding screening specific genes in newborns, as well as assessed responses to a series of descriptive questions. We generated logistic regression models to predict which types of genes were most likely to be recommended for screening by experts, as well as identify the qualities of experts that predicted their likelihood to recommend the inclusion of additional genetic disorders. Experts agreed with $\geq 90\%$ concordance that 11 genes associated with disorders not currently included in universal NBS programs should be evaluated in presymptomatic infants (OTC, G6PC, SLC37A4, CYP11B1, F8, F9, CYP17A1, SLC2A1, ARSB, RB1, DMD). A total of 50 genes had $\geq 80\%$ concordance and 458 genes had $\geq 50\%$ concordance. Overall, 155 of 176 (88.1%) of experts agreed or somewhat agreed that genomic sequencing should be used to expand the number of conditions included in NBS. While several of these genes are associated with clinical domains that are currently included in NBS, there are multiple candidates that represent new frontiers for NBS, such as hereditary cancer predisposition syndromes. This study highlights that experts endorse the expansion of NBS programs to include additional treatable monogenic disorders, including 11 high-priority conditions which could be efficiently screened using genomic sequencing.

Mendelian Phenotypes Posters - Wednesday

PB1832*. Exploration of Ubiquitin-Proteasome System involvement in neurodevelopmental diseases using cellular models.

Authors:

W. Deb^{1,2}, V. Vignard², T. Besnard^{1,2}, B. Cogné^{1,2}, S. Cuinat¹, L. Florenceau², A. Mollé³, J. E. Stanton^{4,5}, S. mercier^{1,2}, M. Nizon^{1,2}, M. Vincent^{1,2}, B. Isidor^{1,2}, E. Krüger⁶, R. Redon², A. M. Grabrucker^{4,5}, J. Poschmann³, F. Laumonnier^{7,8}, F. Ebstein⁶, S. Küry^{1,2}, S. Bézieau^{1,2}; ¹Nantes Université, CHU Nantes, Service de Génétique Médicale, F-44000 Nantes, France, ²Nantes Université, CHU Nantes, CNRS, INSERM, l'institut du thorax, F-44000 Nantes, France, ³Inst. de Recherche en Transplantation - Urologie – Néphrologie, UMR_S 1064, Ctr. de Recherche en Transplantation et Immunologie, F-44000 Nantes, France, ⁴Dept. of Biological Sci., Univ. of Limerick, Limerick, Ireland, ⁵Bernal Inst., Univ. of Limerick, Limerick, Ireland, ⁶Inst. für Medizinische Biochemie und Molekularbiologie, Univ. of Greifswald, Greifswald, Germany, ⁷UMR 1253, iBrain, Université de Tours, Inserm, Tours, France, ⁸Service de Génétique, Ctr. Hosp.ier Régional Univ.ire, Tours, France

Abstract Body:

Context: The Ubiquitin-Proteasome System (UPS) is a major regulator of intracellular protein degradation -and subsequently proteostasis- in eukaryotic cells. A growing number of the 1200 genes constituting the UPS is associated with NeuroDevelopmental Disorders (NDDs), causing 10 to 15% of them. Our team has contributed to the identification of the first genes involved in proteasome-related NDDs: *PSMD12*, *PSMC3*, *BAP1* and *PSMC5*. Our objective is to better understand the pathophysiological mechanisms in play in patients with a UPS-related NDD (UPS-NDD), using cellular models. **Method:** In collaboration with French and international clinical and molecular geneticists, we have gathered blood samples from UPS-NDD patients, from whom we isolated and expanded T-cells. With this cell model, we have been studying the impact of UPS variants on proteostasis and inflammation, through the assessment of proteasomal function and structure, and the evaluation of type 1 interferon response. We have also been exploring the impact of our variants on the expression and regulation of partner proteins, that can be specific for each gene of interest. To further investigate the mechanisms involved, we have produced neuronal models derived from iPSCs, in order to analyze morphological development and, in parallel, gather multi-omics data to be compared to control cell lines. **Results:** To date, we have gathered in a dedicated biobank 170 samples from patients, and their unaffected relatives as controls, with variants in 30 UPS genes, encoding proteasomal subunits (*PSMA3*, *PSMA5*, *PSMB3*, *PSMB5*, *PSMB8*, *PSMB10*, *PSMC1-5*, *PSMD10-14*, *PSMD4*, *PSMD10-14*), ubiquitin-ligases (*CUL2*, *CUL3*, *CUL4B*) as well as deubiquitinases (*OTUD6B*, *USP7*, *USP8*, *USP11*, *USP14*, *BAP1*), therefore encompassing the 3 major actors of the UPS degradation pathway. Common underlying mechanisms in these UPS-NDDs are related to proteasomal enzymatic function, polyubiquitinated proteins aggregation, and abnormal type 1 interferon response. Preliminary results in Chip-Seq highlight dysregulated expression of genes already known for their role in neurodevelopmental processes and intellectual disability. **Discussion:** Current data suggest an overlap in the molecular mechanisms involved in the different UPS-NDDs explored. We wish to expand our study to all UPS-NDDs, thanks our collaborative network, in order to identify molecular keypoints involved in physiopathological processes. Identifying such markers will hopefully allow us to determine which tissues and cell types are the most relevant, to establish reliable genotype-phenotype correlations, and target specific levers for further therapeutic screening.

Mendelian Phenotypes Posters - Thursday

PB1833. Exploring cancer cachexia on the spatial plane

Authors:

Y. Park, L. Song, B. Paulhus, B. Albuquerque, Z. Wu, J. Kim-Muller, B. Zhang, E. Fauman, K. Hales; Pfizer, Cambridge, MA

Abstract Body:

Cachexia is a serious but underrecognized consequence of many chronic diseases characterized by unintentional weight loss, anorexia, muscle/fat wasting, fatigue, muscle weakness and chronic inflammation. The prevalence of cachexia ranges from 20-30% per year in chronic heart failure and chronic kidney disease to 50-80% in advanced cancers. Cachexia is associated with generally poor quality of life and poor prognosis across all affected patients. Growth differentiation factor 15 (GDF15) is a circulating protein implicated in energy homeostasis and body weight regulation. Circulating levels of GDF15 cause anorexia, emesis and weight loss in preclinical species, and correlate with cachexia and reduced survival in patients with cancer. GDF15 inhibition effectively reverses weight loss, muscle and fat loss in mouse models. Interestingly, it remains unclear whether cachexia and GDF15 inhibition have cell-specific effects and whether muscle weakness and muscle atrophy can both be reversed by GDF15 inhibition. We investigated these questions in a mouse tumor model, TOV21G, using a potent and selective monoclonal anti-GDF15 antibody (mAB2) that neutralized circulating GDF15. mAB2 globally reversed cachexia including weight loss, lean and fat mass loss as well as declined muscle function and physical performance in tumor bearing mice. Comprehensive mapping of tissue, cell and disease architecture via spatial transcriptomics (spRNAseq) offers the promise of understanding their modes of dysfunction in the context of intact tissues. We applied Visium H&E protocol to skeletal muscle from non-tumor-bearing mice and tumor-bearing mice with/without mAB2. We also developed and optimized a novel Visium quadplex IF protocol to probe MHC1, MHC2a, MHC2b and WGA. We profiled spRNAseq data deconvolved at single cell resolution using snRNAseq, validated histological fiber typing within each sample and identified rare cell types such as FAPs and tenocytes. Anatomical features such as vasculature were defined by combining communities of cell types. We investigated whether mAB2 has cell-specific effects on atrophy as measurable by fiber size and muscle weight. We identified several novel genes that are significantly associated with TOV21G-induced tumors and subsequently reversed by mAB2 treatment in a fiber- and anatomical-region-specific manner. Together, our findings suggest that spatial data can be used to identify fiber- and anatomical-region-specific effects of cachexia and mAB2 treatment with histological context. Our study demonstrates the great potential for spatial data to be used in therapeutic target validation and exploration of complex diseases.

Mendelian Phenotypes Posters - Wednesday

PB1834. Exploring the distinct phenotypes of nine Romanian patients harboring a homozygous Arg355* variant in *BBS12*

Authors:

I. Focsa^{1,2}, C. Rusu^{3,4}, M. Pânzaru^{3,4}, L. Butnariu^{3,4}, M. Budisteanu^{5,6,7}, L. Bohiltea¹, A. Tutulan-Cunita², D. Stambouli², S. Khan⁸, E. Davis⁸, M. Balgradean^{1,9}; ¹Univ. of Med. Carol Davila, Bucharest, Romania, ²Cytogenomic Med. Lab., Bucharest, Romania, ³Gr T Popa Univ. of Med. and Pharmacy, Iasi, Romania, ⁴Regional Med. Genetics Ctr., “Sf. Maria” Children’s Hosp., Iasi, Romania, ⁵Psychiatry Res. Lab., “Prof. Dr. Alexandru Obregia” Clinical Hosp. of Psychiatry, Bucharest, Romania, ⁶Med. Genetic Lab., “Victor Babeş “Natl. Inst. of Pathology, Bucharest, Romania, ⁷Faculty of Med., “Titu Maiorescu” Univ., Bucharest, Romania, ⁸Stanley Manne Children’s Res. Inst., Ann & Robert H. Lurie Children’s Hosp. of Chicago, Chicago, IL, ⁹Dept. of Pediatrics and Pediatric Nephrology, Emergency Clinical Hosp. for Children “Maria Skłodowska Curie”, Bucharest, Romania

Abstract Body:

Bardet Biedl syndrome (BBS; MIM 209900) is a rare primary ciliopathy that is distinguished by significant genetic and clinical heterogeneity. The core features of the disease are rod-cone dystrophy, postaxial polydactyly, central obesity, urogenital anomalies, learning difficulties and kidney disease. In addition, the impairment of any organ or system may complicate the clinical picture. Twenty-seven genes have been associated with BBS pathogenesis, of which *BBS12* accounts for about 10% of cases of European origin. Here we report the clinical findings of nine patients harboring a homozygous p.Arg355* variant in *BBS12*. Five females and four males from eight unrelated pedigrees were recruited; their ages are between 2 months and 21 years and two individuals are siblings. Ophthalmological assessment confirmed retinal degeneration in five patients while two additional individuals displayed poor vision, however the latter two cases have not been investigated clinically. The other two patients are under the age (2 months and 1 year 7 months, respectively) of development of rod-cone dystrophy symptoms. Polydactyly was observed in all patients, affecting either all four limbs, only the feet, both feet and one hand or left limbs. Obesity was noted in all individuals, with body mass index exceeding the World Health Organization growth standards with 3.2 to 19 deviations. Learning difficulties were present in seven of seven assessed patients ranging from mild to severe. Genitourinary abnormalities were seen in eight patients while different kidney defects, either structural or functional, affect five of eight assessed individuals. Additional clinical findings including neurodevelopmental and behavioral abnormalities, liver involvement, endocrine and metabolic abnormalities, oral-dental anomalies, craniofacial dysmorphic features, or cardiovascular impairment were variably present in our cohort. Detailed clinical assessment of multiple individuals harboring the same primary causal variant confirms the heterogeneous clinical features among individuals and across families shown in previous reports on BBS.

Mendelian Phenotypes Posters - Thursday

PB1835. Family with case of HMG-CoA lyase deficiency with novel pathogenic variant in *HMGCL*

Authors:

T. Froukh; Philadelphia Univ., Amman, Jordan, Jordan

Abstract Body:

3-Hydroxy-3-methylglutaryl coenzyme A lyase deficiency is a rare autosomal recessive disorder characterized by metabolic acidosis without ketonuria, hypoglycemia, and elevated urinary organic acid metabolites including 3-hydroxy-3-methylglutaric, 3-methylglutaric, and 3-hydroxyisovaleric acids. Clinical signs include irritability, lethargy, coma, hepatomegaly and recurrent vomiting. Here, I report a consanguineous family-first cousin parents- with five children, three of them died shortly after death and before diagnoses, one child is healthy and another (born on October 18 2014) experienced similar symptoms to her three dead siblings and she is still alive. The family remained undiagnosed until the exome sequence was read to the index patient in September 2021. The following homozygous variant was identified in the gene *HMGCL* ENST00000374490:c.498-1G>A probably affecting the splicing and thus the final protein structure. The variant was confirmed by sanger sequencing and found to be heterozygous in the parents and in the healthy sibling and therefore the family is diagnosed with HMG-CoA lyase deficiency and is considered the cause of the dead sibling. However, the currently alive affected patient experienced the symptoms of the disease in the past and is currently experiencing mild symptoms. The likely explanation is a random effect of the variant on splicing which might range from very severe, observed in the passed away children, to mild as observed in the still alive patient.

Mendelian Phenotypes Posters - Wednesday

PB1836. First patient reported with a *TPP1* missense mutation pathogenic variant predicted in silico causing *CLN2*

Authors:

L. Moreno Giraldo¹, E. Austin Ward²; ¹Univ. del Valle, Univ. Santiago de Cali, Univ. Libre, Cali, Colombia, ²Hosp. Especialidades Pediátricas (HEPOTH) CSS, Panamá. Hosp. Pacífica Salud, Inst. de Investigaciones Científicas y Servicios de Alta tecnología. Panamá, Panama, Panama

Abstract Body:

Patient 6 years old female, second child born from a nonconsanguineous couple, by cesarean section because of a previous cesarean delivery. Her development was normal until 3 years when she began with tonic-clonic seizures, initially treated with Levetiracetam 4mg/kg. At this age, simple Brain MRI showed a generalized increase in the arachnoid space. An Epilepsy Panel at Invitae™ was performed at age 6. This report included one pathogenic variant and other four variants categorized initially as VUS at that moment. Both the pathogenic and three VUS were in heterozygosis. A variant in *TPP1* gene was reported in homozygosis (c.614T>C (p. Ile205Thr). The bioinformatic analysis of all the variants found, using bioinformatic software, reported a pathogenic clinical significance in the variants reported in the *ARSA*, *COL18A1* and *TPP1* genes; the variant in the *ADAR* gene had benign clinical significance, and the variant in *PCLO* was classified as Unclear Significance; In addition, all variants are currently reported. We present a non-previously reported pathogenic variant of *TPP1* in a patient, although predicted by in silico analysis. *TPP1* is the gene responsible for the enzymatic functioning of tripeptidyl peptidase 1, found in lysosomes. Tripeptidyl peptidase 1 acts as a peptidase, cleaving peptides into groups of three amino acids. This case is especially important for our Latin-American region, which is genetically sui generis and for the patient's country of origin, being the first patient with *CLN2* diagnosed to date.

Mendelian Phenotypes Posters - Thursday

PB1837. First report of severe hemolysis associated with *SCARB1*

Authors:

P. Tanpaiboon^{1,2}, M. Gotesman², E. Panosyan², X. Qing³, J. M. J. Graham², H. J. Lin²; ¹Quest Diagnostics, San Juan Capistrano, CA, ²Dept. of Pediatrics, Harbor UCLA Med. Ctr., Torrance, CA, ³Dept. of Pathology, Harbor UCLA Med. Ctr., Torrance, CA

Abstract Body:

Scavenger receptor class B member 1 is encoded by the *SCARB1* gene on chromosome 12q24.31. The receptor is essential for reverse cholesterol transport by mediating uptake of high-density lipoprotein (HDL) by the liver and free cholesterol efflux from tissues to HDL. Three variants (c.335C>T, c.523A>G, and c.889C>T) in *SCARB1* have been associated with high HDL levels and subclinical atherosclerosis in both homo and heterozygous adults. *SCARB1* variants have not been associated with any other significant phenotypes in humans. We report a neonate with severe hemolysis, elevated cholesterol and HDL levels, and biallelic *SCARB1* variants.

The male infant was born at 36 weeks' gestation. He developed severe jaundice and anemia within the first 24 hours of life. The highest total bilirubin level was 17 mg/dL, and the lowest hemoglobin was 7.4 g/dL. A peripheral blood smear showed numerous acanthocytes. Results of Coombs tests (direct and indirect), G6PD and pyruvate kinase enzyme assays, and hemoglobin electrophoresis were normal. Exome sequencing showed compound heterozygous variants of uncertain significance in *SCARB1*. One variant was a maternally inherited ~7.10 kb deletion encompassing exons 3 to 6 with breakpoints in introns 2 and 6. The second variant was a paternally inherited c.1133C>T change (p.T378M), which has been reported in the compound heterozygous state with another *SCARB1* variant in an individual with hypercholesterolemia. Lipid profiles collected after exome sequencing showed elevated total cholesterol (281 and 241 mg/dL; normal 64-237) and HDL (84 and 68 mg/dL; normal 12-71) levels at 8 and 11 months of age, respectively. The child was doing well at 2.5years, with normal blood counts, he did not receive any treatment after discharge.

The above testing excluded enzymopathies, membranopathies, and hemoglobinopathies as causes of the infant's nonimmune hemolysis and acanthocytosis. Elevated cholesterol and HDL levels were consistent with levels reported for adult patients and support the pathogenicity of these *SCARB1* variants found by exome sequencing.

To our knowledge, this is the first report of hemolytic anemia associated with *SCARB1* variants. Liao et al. (2015) found that *SCARB1* knockout induced hemolysis in hypercholesterolemic mice, hypothesizing that high plasma levels of free cholesterol caused hemolysis by disrupting erythrocyte membranes. Our findings suggest a link between *SCARB1* variants and hemolysis-related membranopathies in humans; functional studies are needed to assess whether such variants have a causative role

Mendelian Phenotypes Posters - Wednesday

PB1838. Fracture prevalence in children diagnosed with Ehlers-Danlos Syndrome and Generalized Joint Hypermobility

Authors:

F. Yeung, J. Smith, **R. Mendoza**, C. O'Connor, A. Howard, J. Sorbara, S. Schwartz; Hosp. for Sick Children, Toronto, ON, Canada

Abstract Body:

Background: Given the substantial consequences for patients and their families, physicians must consider medical conditions that may predispose to injuries seen in cases of suspected child abuse. Recently, hypermobile Ehlers-Danlos Syndrome (hEDS) and hypermobility have been raised as a possible cause for fractures in infants. However, there is no clear evidence to support this hypothesis in the limited pediatrics literature, and no fractures have been documented in infants <1 year of age with EDS. Despite this, the small sample size of this literature continues to be a limitation and hEDS and hypermobility continue to be brought up in the legal setting as a possible cause of unexplained fractures in infants. Thus, we sought to add to the existing literature by assessing fracture prevalence and fracture characteristics in a population of children diagnosed with EDS and Generalized Joint Hypermobility (GJH) by a specialized EDS clinic.

Methods: This retrospective descriptive study identified children aged 0-17 years who were assessed and diagnosed by the EDS clinic with EDS or GJH from April 1st, 2017 - December 15th, 2021. This clinic adhered to the 2017 international EDS diagnostic criteria. Cases were included if they had a history of one or more fractures. Cases were excluded if there were concurrent medical diagnoses and/or use of medications that were known to be associated with bone fragility. Data collection included the subtype of EDS, fracture location, fracture type, age of sustaining each fracture, and injury mechanism. Descriptive statistics were used to analyze the data.

Results: 30 cases of EDS and 60 cases of GJH were identified. Of the EDS patients, 6 were hypermobile type. Four patients were excluded due to underlying medical conditions. Fracture prevalence in the EDS population was 26.7% (8/30). 2 of 6 patients with hypermobile EDS had fractures. Of the fractures with a reported age at time of injury, none occurred at < 2 years of age. In the GJH group, fracture prevalence was 25% (15/60) and no fractures occurred at < 17 months of age. In both groups, all fractures occurred in the limbs, with the exception of 4 fractures which had clear accidental explanations (broken nose after punch, fall on trampoline, fall during cheerleading). There were no rib or skull fractures. Most fractures were the result of an identifiable accidental injury event.

Conclusion: This study supports that children with EDS or GJH do not have an increased prevalence of unexplained fractures in infancy. All fractures occurred in ambulatory age groups. Most fractures were in the limbs and/or were the result of a clearly identified accidental injury event.

Mendelian Phenotypes Posters - Thursday

PB1839. Fragile X Syndrome in a female patient: A case of strong family history and developmental delay.

Authors:

B. Marbaker, R. Lebel, M. Byler, N. Brescia; Upstate Med. Univ., Syracuse, NY

Abstract Body:

Introduction Fragile X Syndrome (FXS) is caused by a triplet repeat expansion in the *Fragile X Mental Retardation 1 (FMR1)* gene on the X chromosome. It manifests with intellectual and developmental disability (especially speech delays and autism spectrum behaviors), macroorchidism in adult males, and a long face with large ears and prognathism. Given its X-linked recessive inheritance pattern, FXS more commonly affects males and is one of the most common inherited causes of intellectual and developmental disability in this population. Females who inherit a full mutation can also exhibit the characteristic developmental phenotype.

Case Description A 15mo female patient with an extensive family history of FXS was referred to genetics after developmental delay and hypotonia were noted. The patient was born at 39 weeks to a 32 year old G4P3>4 mother and a 58 year old father via spontaneous vaginal delivery. Ultrasound and maternal serum screen were unremarkable prior to delivery and there were no pre- or post-natal complications. She exhibited no apparent dysmorphic features and had been growing appropriately but had mild diffuse hypotonia and was delayed in reaching speech and motor milestones. Family history is significant for *FMR1* premutation (70 repeats) in her mother and FXS in two maternal half-brothers and a maternal half-sister. Her mother's paternal half-sisters (monozygous twins) are each premutation carriers, and each has a son affected with FXS. There is no known family history of premature ovarian failure or tremor/ataxia. The patient's Fragile X repeat analysis confirmed a diagnosis of FXS with 24 CGG repeats on one X chromosome and >200 repeats on the other.

Discussion The typically milder phenotype and wide range of severity in female patients with FXS is related to the degree of X inactivation in females who inherit the full mutation. If X inactivation is skewed toward silencing the wild-type *FMR1* allele, less protein product is produced and a more severe phenotype would be expected. Although some female patients affected with FXS display the characteristic facial features, intellectual or developmental delay may be the sole sign of FXS in a female patient. Genetic testing should be considered in any such patient.

Conclusion Fragile X Syndrome is an important diagnostic consideration for any child, male or female, who presents with intellectual or developmental delay. Molecular testing and genetic counseling should be considered for patients and their families in cases of unexplained intellectual or developmental disability, or where a family history of FXS, POI, or FXTAS exists.

Mendelian Phenotypes Posters - Wednesday

PB1840. Frameshift *PPP1R12A* pathogenic variant in a mexican patient with differences of sex development (DSD), middle line defects and hemangiomas.

Authors:

S. Contreras-Capetillo¹, M. Abreu-Gonzalez², Y. Centeno-Navarrete³, S. Ferro-Muñoz⁴; ¹Univ. Autonomous of Yucatan, Merida, Mexico, ²Genos Medica, Ciudad de México, Mexico, ³Hosp. Gen. Dr. Agustín O'Horán, Merida, Mexico, ⁴Hosp. Lic. Ignacio García Téllez, Merida, Mexico

Abstract Body:

Differences of sex development (DSDs) are heterogeneous group of congenital conditions in which chromosomal, gonadal, or anatomical sex do not match. An accurate diagnosis is essential for the care and quality life of the patient, so it must be evaluated by a multidisciplinary team between genetics, endocrinology, urology and surgery. Recently, Hughes et al. reported pathogenic variants in the *PPP1R12A* gene associated with defects in the embryogenesis of the central nervous system and the urogenital system. Here we describe a female assigned patient with hypertelorism, multiple skin hemangiomas and ambiguous genitalia in whom a c.1880delC in *PPP1R12A* was detected. This report main objective is to expand the phenotype of *PPP1R12A* pathogenic variant.

Mendelian Phenotypes Posters - Thursday

PB1841. Frequency of Y chromosome microdeletion and its correlation with spermatogenesis defect in north Indian infertile males.

Authors:

H. Sharma¹, R. Mavaduru², S. Singh², R. Prasad²; ¹AIIMS Bathinda, Bathinda, India, ²PGIMER Chandigarh, Chandigarh, India

Abstract Body:

Deletion in the long arm of Y chromosome is considered as the most common genetic cause of defective spermatogenesis. However, due to variable environmental influence and genetic factors the frequency of the Y chromosome microdeletion is heterogeneous in different population. Further, there is paucity of data pertaining to significant correlation between phenotypic and genotypic profile among Indian fertile males with Y-chromosome microdeletion. Hence, the study was aimed to ascertain the frequency of Y chromosome microdeletion in 292 idiopathic cases of male infertility from northern region of India and an attempt was also made to define the subgroup of patients which are at higher risk of harbouring Y chromosome microdeletion based on their phenotype. Frequency of Y chromosome microdeletion in north Indian infertile males was found to be about 8.5%, with azoospermia factor (AZFc) region as the most susceptible region for microdeletion. Comparatively microdeletion is more common in patients with nonobstructive azoospermia than oligozoospermia (9.2% versus 7.1%). Statistical analysis also revealed that patients with hormonal FSH level between 20 and 40 mIU/mL have more chances of harbouring microdeletion. Hence, we conclude, incidence of Y chromosome microdeletion are more common in northern region of India with AZFc region as most susceptible for deletion. Therefore, all infertile couples specifically those with higher FSH level should be advised for Y chromosome microdeletion screening before undergoing any ART procedure.

Mendelian Phenotypes Posters - Wednesday

PB1842. Functional analysis of *EZH1-A678G* in *Drosophila melanogaster*

Authors:

S. Jangam^{1,2}, L. C. Briere³, K. Jay^{1,2,4}, J. Andrews^{1,2}, M. Walker⁵, F. High⁶, L. Rodan⁷, Undiagnosed Disease Network, H. Bellen^{1,2}, S. Yamamoto^{8,2}, D. Sweetser^{6,3}, M. Wangler^{1,2,4}; ¹Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, ²Jan and Dan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, ³Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁴Genetics and Genomics Program, Baylor Coll. of Med., Houston, TX, ⁵Dept. of Neurology, Div. of Neurogenetics, Child Neurology, Massachusetts Gen. Hosp., Boston, MA, ⁶Div. of Med. Genetics & Metabolism, Massachusetts Gen. Hosp., Boston, MA, ⁷Dept. of Neurology, Boston Children's Hosp., Boston, MA, ⁸Dept. of Molecular and Human Genetics, Baylor Coll. of Med., HOUSTON, TX

Abstract Body:

Genetic mutations in DNA histone modifiers have been shown to disrupt key signaling pathways leading to developmental disorders. *EZH1* (*Enhancer of Zeste 1, Polycomb Repressive Complex 2 Subunit*), a component of Polycomb Repressive Complex-2 (PRC2), promotes maintenance of embryonic stem cell pluripotency by modifying histones at H3 Lysine27 (H3K27). The methyltransferase activity of *EZH1* containing PRC2 complex suppresses downstream target genes by H3K27-trimethylation (H3K27me3). *EZH1* has not yet been associated with any disorders. Here we report a novel *de novo* variant in *EZH1*, *p.Ala678Gly*, resulting in a novel neurodevelopmental phenotype. The patient, evaluated at 2.5 years of age, has a severe global developmental delay, mixed tone, proximal muscle weakness, and intermittent exotropia. While there is no history of seizures, EEG was indicative of mild diffuse cerebral dysfunction. The patient also has mild short stature with preserved head circumference, small hands and feet, mild dysmorphisms, and hypopigmentation. This variant, *EZH1-A678G*, is in a well-conserved motif, the SET domain, which is required for methyltransferase activity. Human *EZH1* is homologous to the *Drosophila Enhancer of zeste (E(z))* gene, and the residue is conserved throughout the species including flies. To study its functional aspect *in vivo*, we generated transgenic flies expressing wildtype (*E(z)WT*) and the variant (*E(z)-A691G*) under the constitutively active tubulin promoter. The resulting variant flies showed a significant bang-sensitive phenotype similar to seizures in humans. Knowing the methyltransferase activity of *E(z)*, we quantified H3K27me3 in those flies and observed hyper H3K27me3 in the variant when compared to wildtype or control. This indicates a gain-of-function mechanism of the variant. In the presence of one *E(z)* null allele (sensitized background), the wildtype transgene resulted in a reduction in H3K27me3, but the variant did not show a significant change. Complete loss of *E(z)* in the fly is lethal, *E(z)WT* rescued the lethality completely, while *E(z)-A691G* rescued it partially. These data suggest that this variant is dominant-negative gain-of-function in nature. Interestingly, while this variant has never been observed in *EZH1*, the human paralogue, *EZH2*, has the identical amino acid variant *EZH2-A677G*, which also results in hyper H3K27me3, supporting our *Drosophila* studies. In conclusion, the *de novo* incidence and evolutionary constraint of this variant, along with the *in vivo* functional studies strongly suggest a novel syndrome for *EZH1* related disorders.

Mendelian Phenotypes Posters - Thursday

PB1843. Functional characterization of recurrent truncating variant in *UBAPI* associated with hereditary spastic paraplegia.

Authors:

S. Gu¹, N. Ho¹, I-J. Lin¹, K-K. Lee¹, L. Meng²; ¹The Chinese Univ. of Hong Kong, Hong Kong, Hong Kong, ²Baylor Coll. Med., Houston, TX

Abstract Body:

Hereditary spastic paraplegias (HSPs) are a group of genetic neurodegenerative diseases resulting from the axonal degeneration of the corticospinal upper motor neurons. Dismayingly however, the molecular diagnosis of more than half of HSP patients and the pathogenicity mechanisms of most HSP genes remain unknown. Recently, our group and others independently identified a novel gene, *UBAPI* (ubiquitin associated protein 1), as causative for autosomal-dominant HSP via clinical exome sequencing analyses. All affected individuals showed highly comparable clinical presentations of non-syndromic HSP, with almost identical disease onset age of around eight years old. To understand the mechanism underlying *UBAPI*-associated disease, we conducted cellular and model organism studies to authentically reproduce the pathogenicity observed in patients. Specifically, applying CRISPR/Cas9 genome editing, we established human embryonic stem cells (hESCs) and a mouse model harboring a patient-specific frameshift variant that was observed in more than half of *UBAPI*-HSP patients. *UBAPI* encodes a highly conserved and ubiquitously expressed protein. It is a component of the endosome-specific ESCRT-I complex that functions to sort ubiquitinated cargo and to maintain endosomal ubiquitin homeostasis. This two base-pair deletion led to stable expression of a truncated protein, which retained the ESCRT-I binding domain yet lost the ubiquitin interaction domains. We therefore proposed that the truncated protein exert a dominant-negative effect on its normal function, as it could still be recruited by the ESCRT-I complex yet lost the ability to interact with ubiquitinated proteins. Consistent with the proposed mechanism, we observed vesicle accumulation and enriched ubiquitinated proteins in mutant cells comparing to wildtype cells. Additionally, mutant hESCs-derived upper motor neuron organoids demonstrated reduced neurite outgrowth and axon length, resembling the hallmark phenotype in HSP patients. Further, mutant mice exhibited hindlimb motor defects starting from 6 months old, which were illustrated by beam walking and hindlimb clasping tests. Currently, besides identifying and quantifying the accumulated ubiquitinated proteins in mutant cells through mass spectrometry, we are also investigating the distinct pathogenicity of heterogeneity in mouse brains upon *Ubap1* mutation via single-nucleic transcriptome analysis. Our studies will elucidate not only the pathogenicity of *UBAPI*, but also its role in maintaining endosomal dynamics, providing an authentic preclinical model system for future therapeutic studies of HSP.

Mendelian Phenotypes Posters - Wednesday

PB1844. Gaps in the phenotype descriptions of ultra-rare genetic conditions: Systematic review and recommendations

Authors:

A. Almail¹, A. Jamjoom², A. Pan³, N. Jewitt⁴, C. Diskin⁴, D. Baribeau⁵, G. Costain²; ¹Univ. of Toronto, Temerty Faculty of Med., Toronto, ON, Canada, ²The Hosp. for Sick Children, Toronto, ON, Canada, ³Univ. of Toronto, Dept. of Molecular Genetics, Toronto, ON, Canada, ⁴The Hosp. for Sick Children, Dept. of Pediatrics, Toronto, ON, Canada, ⁵The Hosp. for Sick Children, Dept. of Psychiatry, Toronto, ON, Canada

Abstract Body:

Background: Genome-wide sequencing and genetic matchmaker services are propelling a new era of genotype-first ascertainment of novel ultra-rare diseases. The degree to which reported phenotype data address informational priorities for clinicians and families is unclear. We hypothesize that current phenotype descriptions limitedly address questions of adaptive functioning, feeding and growth, medication use, and proxies for quality of life.

Methods: We systematically reviewed the literature to identify reports of novel genetic conditions published from 2017-2021 where ascertainment was genotype-first. Reports were assessed by two independent raters regarding the adequacy of phenotype data provided in these domains: (I) Development, cognition, adaptive functioning, behavior, and mental health; (II) Feeding, growth, and nutrition; (III) Medication use and treatment history; and (IV) Pain, sleep, and quality of life.

Results: In total, 200 of 3243 screened publications met inclusion criteria. Preliminary analysis of the reported phenotype data revealed superficial descriptions and numerous gaps in the four phenotype domains in the majority of publications. Common issues included lack of detail regarding the severity of developmental delays / intellectual and developmental disabilities, use of general descriptions like “feeding difficulties” and “behavior problems”, and little to no information about past treatment trials (e.g., anti-epileptic and psychotropic medications).

Conclusion: Phenotype information relevant to clinical management, genetic counseling, and the stated priorities of patients and families is lacking in many descriptions of new ultra-rare genetic diseases. We propose a checklist to guide phenotype data collection and reporting.

Mendelian Phenotypes Posters - Thursday

PB1845. *GATA3*-related pedigrees showing characteristic audiologic profile and inheritance pattern.

Authors:

B. Kim¹, J. Han^{1,2}, Y-M. Kim¹, B. Choi²; ¹Chungnam Natl. Univ. Coll. of Med./Chungnam Natl. Univ. Sejong Hosp., Sejong, Korea, Republic of, ²Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of

Abstract Body:

GATA binding protein 3 (*GATA3*) is a protein encoded by *GATA3* gene, which is associated with autosomal dominantly-inherited Hypoparathyroidism, Deafness, and Renal Dysplasia (HDR) syndrome (MIM no. 146255). The phenotypes of HDR syndrome are known to vary and sometimes symptoms related to hypoparathyroidism and renal disease are subclinical, which makes it challenging to diagnose the syndrome in the patient predominantly presenting with hearing loss. Here, we report three sporadic cases who first presented with asymmetric, low-frequency hearing loss, and eventually are diagnosed as having *GATA3* variants-related HDR syndrome by molecular genetic diagnosis and luciferase assay.

A 14-year-old girl and a 21-year-old man presented with asymmetric moderate degree low-frequency hearing loss. Screening panel of deafness genes followed by exome sequencing revealed two potential candidate variants of *GATA3* (c.404del:p.Pro135Argfs*60 and c.1015T>C:p.Cys339Arg (novel variant) respectively) in each family. Luciferase assay using Wildtype and two *GATA3* variants co-transfected with *KLRC1* promoter into HEK293T cell line, showed reduced luciferase activity in these two variants than wildtype, further confirming the pathogenic potential of the two variants. Another 1-year-old boy was referred for hearing evaluation and asymmetric moderate hearing loss was found, although frequency-specific hearing threshold was not adequately obtained because of young age. Exome sequencing identified c.1099C>T:p.Arg367* of *GATA3* as a causative variant. Segregation study confirmed de novo inheritance pattern in three pedigrees and subsequent laboratory and radiologic evaluation revealed varying degree phenotypes of HDR syndrome in all three patients, which are mostly subclinical, but should be adequately managed.

Taken together, we report three *GATA3* pedigrees, initially presenting with hearing loss, finally diagnosed of HDR syndrome. Luciferase assay assisted proving the pathogenicity of each variant and de novo inheritance pattern and asymmetric, low-frequency audiogram pattern observed in these pedigrees could be a notable clue for further pursuit of *GATA3*-related HDR syndrome. This genetic diagnosis-driven revelation of varying degree symptoms of HDR syndrome should also be noted from clinical perspectives.

Mendelian Phenotypes Posters - Wednesday

PB1846. Genealogy as a predictor of disease progression in patients with myotonic dystrophy type 1: A demonstration of the power of intersectorial research.

Authors:

J. Bouchard¹, C. Moreau¹, L. Gagnon¹, A. Girard², H. Vezina¹, A. Barry¹, C. Gagnon³, É. Duchesne¹, S. Girard⁴, ¹Université of Québec at Chicoutimi, Chicoutimi, QC, Canada, ²Univ. of Quebec in Chicoutimi, Chicoutimi, QC, Canada, ³Univ. of sherbrooke, Chicoutimi, QC, Canada, ⁴Université du Québec à Chicoutimi, Chicoutimi, QC, Canada

Abstract Body:

Background/Objectives: Myotonic dystrophy type 1 (DM1) is an autosomal dominant disease and is the most common form of muscular dystrophy in adults. DM1 is caused by an abnormal repetition of a CTG nucleotide triplet located on the DMPK gene. DM1 is classically categorized into five phenotypes based on the age of onset and the number of CTG repeats. The presence and severity of signs and symptoms, as well as their progression, vary greatly not only between but also within the phenotypes. However, there is a lack of knowledge in identifying the predictors of multisystemic impairments associated with DM1. The objective of this study was to verify if genealogical kinship-based clustering could be used to better understand the severity and progression of physical impairments in DM1. Our research team proposes an innovative method combining a large genealogical dataset with an extensive DM1 patients' clinical collection. **Methods:** The cohort is composed of 200 participants from the Saguenay-Lac-St-Jean region in Quebec who participated in a longitudinal study over 20 years. Our team has collected several clinical measures related to the progression of the disease including grip strength, forced vital capacity, walking speed, balance and upper- and lower-limb maximal muscle strength measurement. We have also conducted a deep characterization of CTG repeat length for each patient including the presence of interruptions. Additionally, we took advantage of the BALSAC population database (based on linked civil and church records) to reconstruct the genealogies of the participants. Genealogical kinship-based clustering was performed at different generations in order to identify individuals and families with high levels of kinship. **Results:** As previously reported, we observed that the CTG repeat number did not correlate strongly with severity. Correlations with muscle impairment measurements were tell us more. Using our deep genealogical records, we were able to identify several clusters of families with higher kinship values. We currently investigate if certain family clusters present different profiles of disease severity, CTG repeats and muscle impairments. Additionally, the genealogical clusters at different generations will be compared with the severity of the disease to verify if they could become a marker for earlier therapies in order to improve the prognosis in DM1. **Conclusion:** We report here one of the first studies that aims to bridge genealogical structures with clinical measures in DM1. This research will increase our understanding of a disease caused by a nucleotide repeat in a population with a strong founder effect.

Mendelian Phenotypes Posters - Wednesday

PB1848. Genetic analysis reveals that GNE Myopathy remains an underdiagnosed neuromuscular disorder

Authors:

F. Rossignol, P. Leoyklang, M. Malicdan, C. Ciccone, Q. Yuan, J. Jang, A. Bowling, N. Carrillo, W. Gahl, M. Huizing; NHGRI, NIH, Bethesda, MD

Abstract Body:

GNE encodes the rate-limiting, bifunctional enzyme of sialic acid biosynthesis, UDP-*N*-acetylglucosamine 2-epimerase/*N*-acetylmannosamine kinase (*GNE*). Bi-allelic variants, mostly missense, underlie recessively inherited *GNE* myopathy, an adult-onset progressive myopathy due to decreased sialylation of muscle glycoconjugates. The recent expansion of molecular testing and inclusion of *GNE* in neuromuscular disease panels greatly increased diagnosis and identification of new *GNE* gene variants. Since the clinical presentation at onset of disease is non-specific and the majority of *GNE* variants are missense, the interpretation of pathogenicity of each variant for a correct diagnosis remains a challenge. To support our *GNE* myopathy natural history study and clinical trial with the sialic acid precursor *N*-acetylmannosamine (ManNAc) (ClinicalTrials.gov: NCT01417533, NCT04231266), we assembled a comprehensive list of 285 *GNE* gene variants associated with *GNE* myopathy reported in the literature to date, including pathogenicity classification according to American College of Medical Genetics (ACMG). We also listed the bi-allelic variants of all reported genotyped *GNE* myopathy patients (1221 cases worldwide). Reported *GNE* variants not associated with *GNE* myopathy, including those causing the dominantly inherited disorder sialuria or those associated with thrombocytopenia, are also discussed, as they may obscure the *GNE* myopathy diagnosis.

Based on allele frequencies in gnomAD, the world-wide prevalence of *GNE* myopathy was estimated to be at least 2.5/1,000,000. This unrecognized high prevalence (~ 20,000 patients worldwide; ~850 patients in USA) confirms suspicions that many patients escape diagnosis. Indeed, our ongoing natural history study revealed significant diagnostic delay (~9 years) after initial symptoms in most patients, due to the rare nature of the disease and the lack of conclusive, inexpensive diagnostic tests. Genetic testing for pathogenic, bi-allelic *GNE* variants ultimately confirms the diagnosis.

GNE myopathy should be considered in any young adult with distal, lower extremity muscle weakness. Delayed diagnosis causes emotional hardship for the patient, postpones proper disease management, and influences eligibility to enroll in clinical trials. Awareness among physicians including neurologists and geneticists for *GNE* myopathy is essential for the identification of new patients, preferably in early stages of disease. This will support understanding early pathomechanisms of the disease and success of ongoing treatment trials.

Mendelian Phenotypes Posters - Thursday

PB1849. Genetic and clinical landscape of childhood cerebellar hypoplasia and atrophy.

Authors:

M. Sakamoto¹, N. Tsuchida^{1,2}, Y. Uchiyama^{1,2}, E. Koshimizu¹, A. Fujita¹, K. Hamanaka¹, K. Misawa¹, S. Miyatake^{1,3}, T. Mizuguchi¹, N. Miyake^{4,1}, N. Matsumoto¹; ¹Dept. of Human Genetics, Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, ²Dept. of Rare Disease Genomics, Yokohama City Univ. Hosp., Yokohama, Japan, ³Clinical Genetics Dept., Yokohama City Univ. Hosp., Yokohama, Japan, ⁴Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan

Abstract Body:

[Purpose] Cerebellar hypoplasia and atrophy (CBHA) in children is an extremely heterogeneous group of disorders, though few comprehensive genetic studies have been reported. [Methods] Using exome sequencing, patients with CBHA in 176 families were genetically examined and patients who have disease-causing variants were clinically evaluated. [Results] In 95 of the 176 families (54%), disease-causing variants were identified. After excluding six families unsuitable for further analysis, 47 patients from 41 families were categorized as having specific diseases associated with CBHA, while the remaining 51 patients from 48 families had true CBHA with and without supratentorial lesions (CBHA+S [N = 15] and CBHA-S [N = 36], respectively). In the 48 families, 26 aberrant genes were identified, of which 20 (76.9%) caused disease in a single family each, with long tail distribution. *CACNA1A*, *ITPR1*, and *KIF1A* were the most prevalent genes (8+8+6=22, 45.8%). Of the 26 aberrant genes, 14 were functionally annotated to cerebellar atrophy. CBHA-S was more clinically milder than CBHA+S. Notably, *ARG1* and *FOLR1* variants were identified in two families, for whom medical treatments were available. [Conclusion] This cohort revealed a wide genetic and clinical diversity of CBHA by exome sequencing, which highlights the importance of comprehensive genetic analyses. Furthermore, molecular-based treatment was possible for two families. Acknowledgments: We would like to thank all the following doctors: Sasaki M, Ishiyama A, Komaki H, Saito T, Takeshita E, Shimizu-Motohashi Y, Haginoya K, Kobayashi T, Goto T, Tsuyusaki Y, Iai M, Kurosawa K, Osaka H, Tohyama J, Kobayashi Y, Okamoto N, Suzuki Y, Kumada S, Inoue K, Mashimo H, Arisaka A, Kuki I, Saijo H, Yokochi K, Kato M, Inaba Y, Gomi Y, Saito S, Shirai K, Morimoto M, Izumi Y, Watanabe Y, Nagamitsu S, Sakai Y, Fukumura S, Muramatsu K, Ogata T, Yamada K, Ishigaki K, Hirasawa K, Shimoda K, Akasaka M, Kohashi K, Sakakibara T, Ikuno M, Sugino N, Yonekawa T, Gürsoy S, Cinleti T, Kim CA, Teik KW, Yan CM, Haniffa M.

Mendelian Phenotypes Posters - Wednesday

PB1850. Genetic and phenotypic heterogeneity of childhood-onset hearing loss and implications for the success of cochlear implants.

Authors:

R. Carlson¹, T. Walsh¹, J. Mandell¹, A. Abu Rayyan¹, D. Horn², H. Ou², K. Sie², L. Mancl¹, J. Rubinstein¹, M-C. King¹; ¹Univ. of Washington, Seattle, WA, ²Seattle Children's Hosp., Seattle, WA

Abstract Body:

In the US, most childhood-onset hearing loss is genetic and highly heterogeneous with more than 120 causal genes and thousands of different causal alleles known. Cochlear implants are the management of choice for many children with hearing loss, and their success may be associated with the cause of the hearing loss. In a cohort of 449 children from 406 families treated for hearing loss at Seattle Children's Hospital or the University of Washington between 2019 and 2021, we evaluated the genetic causes of hearing loss and how these causes might contribute to cochlear implant success. Genomic DNA was evaluated by targeted sequencing of 191 genes and by structural variant analysis. Longitudinal audiologic testing was used to evaluate severity and progression of hearing loss, and longitudinal speech perception testing was used to evaluate success of cochlear implants. A genetic cause of hearing loss was found for 53% (213/407) of families, with causal mutations in 49 different genes. Age-specific audioprofiles over time revealed gene-dependent and allele-dependent variation in severity, affected frequencies, and progression with age. Hearing loss was significantly progressive for children with hearing loss due to mutations in *MYO6*, *OTOA*, *SLC26A4*, *TMPPRS3*, or severe loss-of-function mutations in *GJB2*. Cochlear implant success was high overall, with 89% of implanted patients scoring at least 60% on adult-level speech perception tests. Even so, level of success varied significantly by genotype (ANCOVA $P < 0.0001$). Cochlear implants benefitted children with hearing loss regardless of cause, but level of speech perception after implant varied by causal gene. Therefore, management of childhood-onset hearing loss can benefit from cochlear implantation undertaken in the context of genetic diagnosis. Supported by research grants from the Bloedel Foundation, the Ben B. Cheney Foundation, and by NIH/NIDCD 5R01DC011835.

Mendelian Phenotypes Posters - Thursday

PB1851. Genetic insights from consanguineous cardiomyopathy families.

Authors:

Y. Jamshidi¹, O. Boleti², P. Najarzadeh³, F. Norouzi⁴, S. Minaee⁵, K. Salih⁶, M. Taherpour⁵, H. Birjandi⁷, B. Alizadeh⁷, A. Salih⁶, M. Bijari⁸, R. Maroofian⁹, E. Ghayoor^{10,1}, J. Kaski¹¹, F. Vakilian⁸; ¹St George's Univ. of London, London, United Kingdom, ²Great Ormond Street Hosp. for Children, Ctr. for Inherited Cardiovascular Diseases, London, United Kingdom, ³Next Generation Genetic Polyclinic, Mashhad, Iran, Islamic Republic of, ⁴Dept. of Cardiology, Faculty of Med., Mashhad Univ. of Med. Sci., Mashhad, Iran, Islamic Republic of, ⁵Dept. of Cardiovascular Diseases, Razavi Hosp., Mashhad, Iran, Islamic Republic of, ⁶Dept. of Pediatrics, Coll. of Med., Sulaimani Med. Univ., Sulaimania, Iraq, ⁷Dept. of Pediatric Diseases, Faculty of Med., Mashhad Univ. of Med. Sci., Mashhad, Iran, Islamic Republic of, ⁸Faculty of Med., Mashhad Univ. of Med. Sci., Mashhad, Iran, Islamic Republic of, ⁹Dept. of Neuromuscular Diseases UCL Queen Square Inst. of Neurology, Univ. Coll. London, London, United Kingdom, ¹⁰Dept. of Med. Genetics, Next Generation Genetic Polyclinic, Mashhad, Iran, Islamic Republic of, ¹¹Ctr. for Inherited Cardiovascular Diseases, Great Ormond Street Hosp. and UCL Inst. of Cardiovascular Sci., Zayed Ctr. for Rare Disease Res., London, United Kingdom

Abstract Body:

Background: Inherited cardiomyopathies are a prevalent cause of heart failure and sudden cardiac death. Both hypertrophic (HCM) and dilated cardiomyopathy (DCM) are genetically heterogeneous and typically present with an autosomal dominant mode of transmission. However, in highly consanguineous populations, such as in the Middle East, there is a burden of autosomal recessive disease variants.

Methods: The study included 9 un-related probands from consanguineous Middle Eastern families presenting with HCM/DCM. Phenotypic assessment and diagnoses were based on cardiological investigations and clinical data, including echocardiography, resting and ambulatory 12-lead electrocardiogram, cardiac catheterization and cardiac MRI. Whole exome sequencing was carried out in the proband followed by bioinformatic and co-segregation analysis to predict the potential pathogenicity of candidate causal variants.

Results: We identified homozygous missense variants in *LDB3*, *TNNI3K*, *DOLK*, *DSP*, *DSC2*, *RBCK1* and *MYH6* linked with a dilated phenotype, in *NRAP* linked with a hypertrophic phenotype, and in *KLH24* linked with a mixed phenotype of dilated/hypertrophic and non-compaction features. Co-segregation analysis by Sanger sequencing in available family members confirmed autosomal recessive inheritance associated with disease presenting either in early childhood or early adulthood.

Conclusions: Our findings add to the mutational spectrum of recessive cardiomyopathies, supporting inclusion of *KLH24* and *NRAP* as disease-causing genes for cardiomyopathy. We also provide evidence for novel (recessive) modes of inheritance of well-established genes such as *TNNI3K* and *LDB3*, expanding our knowledge of the clinical heterogeneity of cardiomyopathies. A greater understanding of the genetic causes of cardiomyopathy, particularly in underrepresented populations such as the Middle East, has major implications for diagnosis and screening.

Mendelian Phenotypes Posters - Wednesday

PB1852. Genetic Mapping Identifies A Founder Mutation of *GHRHR* Gene in Pakistani Family with Isolated Growth Hormone Deficiency Type 4

Authors:

M. Khan¹, S. Ahmad², M. Ali², S. Abbas², I. Ahmed¹, S. Nawaz¹, M. Ziab¹, K. Fakhro¹, A. Akil¹; ¹Sidra Med., Doha, Qatar, ²Gomal Univ., D.I.Khan, Pakistan

Abstract Body:

Background Isolated growth hormone deficiency (IGHD) is characterized by growth and developmental failure in children due to disruption of the growth hormone (GH) or growth hormone releasing hormone (GHRHR). Patients with IGHD type IV have short stature, reduced serum growth hormone (GH) levels, and delayed bone age. **Objectives** Genetic mapping of familial IGHD cases to identify the mutation and explore its functional impact through *in silico* approaches. **Methods** In the present study, we ascertained a consanguineous IGHD family comprising of four patients. Clinical and radiological studies of the affected individuals were performed for deep phenotyping of the disease. Whole exome sequencing (WES) and Sanger sequencing was carried out to identify the disease-causing mutation. *In silico* studies involved protein structural modeling, docking and simulation analysis using different computational tools. **Ethical Statement** The study is approved by the ethical review board of Gomal University, D.I.Khan, and patients were enrolled signing the informed written consent. **Result** We investigated the phenotypic spectrum, hormonal profile, and the molecular basis of IGHD4 (MIM# 618157). All affected individuals presented short stature, without any gross skeletal anomalies. Biochemical findings demonstrate a significant reduction in the level of serum growth hormone. Genetic mapping revealed a nonsense mutation [NM_000823:c.G214T:p.(Glu72*)] in the 3rd exon of *GHRHR* gene (MIM#139191). The identified variant segregated in all patients, but it was absent in the normal individuals. The substituted amber codon (UAG) presumably truncates the protein by deleting the C-terminus GPCR domain. The protein modeling studies have shown a drastic effect of this mutation on GHRHR structure and its interaction with growth hormone releasing hormone and thus confirms its involvement in disease etiology. The identified protein truncation mutation (Glu72*) has previously been mapped in three South Asian families, which suggest the founder effect of this mutation in IGHD. **Conclusion:** Our study reports a founder mutation of *GHRHR* gene in an extended consanguineous family, and confirms the body of evidence that p.Glu72* disrupt the receptor protein, affect the hormonal signaling and causes IGHD type IV. Moreover, the deep phenotype analysis, and hormonal profiling will help in establishing the genotype-phenotype correlation.

Mendelian Phenotypes Posters - Thursday

PB1853. Genetic modifiers in Niemann-Pick type C1 disease.

Authors:

D. Sitarska^{1,2}, K. Bodzon^{1,3}, L. Sivitskaya², A. Lugowska¹; ¹Dept. of Genetics, Inst. of Psychiatry and Neurology, Warsaw, Poland, ²Genomed S.A., Warsaw, Poland, ³Dept. of Biology, Univ. of Warsaw, Warsaw, Poland

Abstract Body:

Niemann-Pick disease type C1 (NPC1) is a genetically determined neurodegenerative metabolic disease characterized by the accumulation of non-esterified cholesterol and glycosphingolipids in the lysosomes. It is an autosomal recessive disease that results from pathologic mutations in the *NPC1* gene. NPC1 is exceptionally heterogeneous in age of onset and symptoms, which may vary from perinatal period to adulthood. Similarly, patients' lifespan can range from a few days to even tens of years. Most patients develop hepatosplenomegaly. Among the neurological symptoms the most characteristic are: cerebellar ataxia, dysarthria, dysphagia and developmental delay or progressive dementia. The majority of patients show a characteristic symptom - vertical supranuclear gaze palsy (VSGP). Cataplexy, epileptic seizures and dystonia are also common and psychiatric symptoms, present in patients with late onset of the disease include e.g. psychosis, paranoid delusions or schizophrenia. The severity of the NPC1 disease depends primarily on the type of genetic variants in the *NPC1* gene. However, some patients with the same pathologic mutations display different phenotypes, even within the same family. We hypothesised, that there are some other variants (genetic modifiers) that may suppress or enhance the severity of the disease, resulting in the variability of phenotypic outcomes in NPC1 patients. We performed NGS analyzes in 20 patients with NPC1 of various types. DNA samples were grouped according to the similarity of the genotype. The most frequent variant in our cohort was c.3019C>G p.Pro1007Ala, including 5 unrelated homozygous patients. The age of onset of the first symptoms in these patients ranges from 3 months of age (seizures) to 25 years (VSGP). In order to elucidate the potential influence of molecular background onto the clinical outcome of NPC1 disease, the NGS data were studied and the identified possible modifier genes variants will be presented. This work was supported by National Science Centre (<https://www.ncn.gov.pl>) grant no. 2019/35/N/NZ2/03102.

Mendelian Phenotypes Posters - Wednesday

PB1854*. Genetic modifiers of nutritional status in CF are both unique to CF and shared with the general population

Authors:

H. Ling¹, M. A. Aksit¹, E. W. Pugh¹, R. G. Pace², F. Onchiri³, K. Raraigh¹, F. Wright⁴, A. Faino³, M. Bamshad⁵, R. Gibson³, M. Knowles², Y-H. Zhou⁴, G. Cutting¹, S. Blackman¹; ¹Johns Hopkins Univ., Baltimore, MD, ²Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ³Seattle Children's Hosp., Seattle, WA, ⁴North Carolina State Univ., Raleigh, NC, ⁵Univ. of Washington, Seattle, WA

Abstract Body:

Background: Individuals with cystic fibrosis (CF) often struggle to maintain good nutritional status, and this contributes to lung disease and mortality. Paradoxically, some with CF have obesity. After accounting for CFTR genotype severity, variation in nutritional status (e.g. body mass index, BMI) is largely determined by genes other than *CFTR* (genetic modifiers). We report an initial genome-wide association (GWAS) study for BMI on >4000 individuals with CF in the CF Genome Project (CFGP). **Methods:** Longitudinal height, weight, demographics, and covariates were ascertained from medical records and the CFF Patient Registry 2017. The trait used in the analysis (AvgBMIz) was derived by calculating the average of per-quarter BMI z-scores. GWAS used linear mixed modeling with fixed effect of study site, first 4 principal components, birth cohort, and pancreatic sufficiency (PS) status. Relatedness was modeled as a random effect. **Results:** In an analysis of 4101 CF individuals with severe CFTR genotypes, genome-wide significant association was observed for variants in the *FTO* gene (rs28567725, $P=3.36e-8$; MAF = 0.41) and *ADAMTS5* (rs162500, $P=4.92e-10$; MAF = 0.006, $\beta=-0.7$). The variants in *ADAMTS5* fall on a haplotype that is more common in TOPMed African (MAF = 18.3%) than European (MAF = 0.6%) samples. Among 46 unrelated carriers of a rs162500 minor allele, 28 (61%) were non-EUR vs 4% in the entire CFGP samples. To assess overlap with BMI modifiers in the general population, a polygenic risk score was calculated using 941 variants identified from 700,000 non-CF Europeans (Yengo 2018); this PRS was strongly associated with AvgBMIz ($P=4.9e-30$; $r^2=0.03$). **Conclusions:** Some genetic modifiers of BMI in the CF population are shared with the general population, and others appear unique to CF. *ADAMTS5* variants are not known to be associated with BMI in the general population, but functional studies suggest a role in liver integrity in a mouse model of high-fat diet-induced obesity. Understanding the genetic contributions to BMI in CF can identify mechanisms contributing to undernutrition or obesity in people with CF, who have high metabolic needs and consume high-calorie, high-fat diets. *Supported by CFF CUTTINI8XX1, BAMSHA18XX0, and KNOWLE18XX0.*

Mendelian Phenotypes Posters - Wednesday

PB1855. Genetic underpinnings of moderate to severe hearing loss in singleton individuals born to consanguineous couples in Pakistan

Authors:

H. Khan¹, F. Muzaffar¹, G. H. Seo², S. Naz¹; ¹Sch. of Biological Sci., Univ. of the Punjab, Lahore, Pakistan, ²3Billion inc., Seoul, Korea, Republic of

Abstract Body:

Hearing loss is a common sensory defect which affects over 5% of the world's population. Genetic studies comprising hearing loss in individuals with no previous history of deafness are relatively uncommon in Pakistan. We recruited 22 singletons born to consanguineous couples with no family history of deafness. Audiometry was completed in ambient noise conditions. Exome sequencing was performed on extracted DNA. Most of the participants had non-progressive moderate to severe degree of sensorineural hearing loss. The exome data were filtered and variants with frequency of less than 0.01 in public databases including gnomAD were assessed using automatic variant prioritization system by 3billion and Franklin software. Pathogenicity predictions by multiple software and conservation scores were considered. Analyses revealed both known and novel variants in the samples of the participants. Pathogenic variants in *SLC26A4* were found in nine individuals out of which a single missense variant c.1337A>G:p.(Gln446Arg) was observed in more than half of the participants. Different pathogenic variants of *CDH23* and *OTOF* were identified in two affected individuals each. Homozygous missense, nonsense, and frameshift variants in *CLND14*, *LHFPL5*, *RDX* and *USH1G* were found in samples of four affected individuals while compound heterozygous missense variants in *MYO15A* were identified as the cause of the hearing loss for one participant. Interestingly, a missense variant in *HGF* was identified which was predicted to be pathogenic and could perhaps be the cause of hearing loss of the participant. Previously, only non-coding variants of *HGF* have been implicated in deafness. Exome sequencing failed to pinpoint the cause of hearing loss for three individuals. Our study strengthens the involvement of previously described variants of genes implicated in hearing loss as well as describes novel pathogenic variants in deafness genes. It further suggests that either some deafness genes remain to be identified or the variants were not detected due to limitations of exome sequencing. Whole genome sequencing could be employed in future to detect pathogenic variants in undiagnosed samples.

Mendelian Phenotypes Posters - Thursday

PB1856. Genetics of intellectual disability in pakistani consanguineous families.

Authors:

M. Rasheed¹, V. Khan¹, R. Harripaul², M. Malik¹, M. Zahid¹, Z. Ullah¹, J. Vincent³, S. Leal⁴, M. Ansar⁵; ¹Quaid-i-Azam Univ., Islamabad, Pakistan, ²CAMH, Toronto, ON, Canada, ³Ctr Add/Mental Hlth, Clarke Div, Toronto, ON, Canada, ⁴Columbia Univ., New York, NY, ⁵Quaid-i-Azam Univ, Islamabad, Pakistan

Abstract Body:

Intellectual disability (ID) is a neurodevelopmental disorder which is characterized by less than average intelligence and compromised adaptive behavior. ID affects 1-3% population worldwide and is much prevalent in certain populations. ID is genetically heterogeneous with over 1500 genes currently known and presence of different co-morbidities further adds to the genetic complexity. The present study focused on ten consanguineous families (Family A to J) from Pakistan with autosomal recessive and X linked ID inheritance with an aim to identify underlying genetic causes of the disease. The clinical investigation of the families revealed presence of additional clinical features in some families such as seizures (Family A, B, G), hypotonia (Family A), facial dysmorphism (Family A, C, G, H, J), camptodactyly (Family B), autistic features (Family F), behavioral abnormality (Family A, E, F, H, J), speech impairment (Family A, B, C, D, F, G, H, I), lethargy (Family F, G), microcephaly (Family I), ataxia (Family A, I) and ambulation delay (Family J). The genetic analysis was performed by using genome wide SNP microarray and exome sequencing which helped in the identification of homozygous missense variants in three novel genes, *ZBTB11* (p.H880Q), *ELFN1* (p.P50H) and *CCS* (p.117A), in family A, B and G(b), respectively. *ZBTB11* acts as a transcription factor which is localized in nucleolus and the variant, p.H880Q, identified in family A results in its mislocalization resulting in loss of function. *ELFN1* functions as trans-regulator of *mGluR7* in excitatory post-synaptic sites and *Elfn1* knockout mice exhibits seizures along with hyperactivity and these clinical features were also observed in our patients of family B. In family C and F, novel non-sense homozygous variants were identified in previously known ID genes i.e. *METTL4* (p.R84*) and *C5orf42* (p.Q2871*). In family E and I, a 9 bp and a 4 bp homozygous deletions were identified in *ARX* (p.R483fs*46) and *BCKDK* (p.D216Lfs*65), respectively. In family D, a novel homozygous ~0.2 Mb deletion was identified which resulted in the deletion of exon 6 of *ILIRAPL1*. In family G(a), H and J, recurrent mutations were identified in *GNE* (p.Y156H), *RSRC1* (c.532-1G>A) and *LRP2* (p.D3779N). The *GNE* missense variant, p.Y156H, is previously reported to cause hereditary inclusion body myopathy however, the clinical profile in family G(a) was ID with seizures. These gene mutations are believed to cause ID by disrupting cellular processes such as neuronal functioning (*ELFN1*, *ARX*, *ILIRAPL1*, *C5orf42*, *LRP2*), protein synthesis (*ZBTB11*, *RSRC1*, *METTL4*) and metabolism (*BCKDK*, *GNE*).

Mendelian Phenotypes Posters - Wednesday

PB1857. Genome sequencing identifies coding and non-coding variants in hereditary deafness missed by exome sequencing.

Authors:

M. Ramzan¹, D. Duman², L. C. P. Hendricks¹, M. F. Zafeer¹, G. Bademci¹, M. Tekin¹; ¹Dr. John T. Macdonald Fdn. Dept. of Human Genetics and John P. Hussman Inst. for Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, ²Univ. of Ankara, Ankara, Turkey

Abstract Body:

The recent implementation of advanced sequencing technologies has significantly contributed to the molecular study of hearing loss. It accelerated the identification of causative variants and allowed the discovery of novel candidate genes involved in the auditory process. In this study, probands of 162 families from different ethnicities, referred with presumably non-syndromic autosomal recessive hearing loss were subjected to Exome Sequencing (ES) after excluding common *GJB2* variants. A total of 69 variants in 26 genes were detected and confirmed with Sanger sequencing which are considered to explain their phenotype. This allowed us to solve only 37% of the cases. Thirty multiplex families which remained unsolved after ES, underwent Genome Sequencing (GS). We solved 21 (70%) of these families after data analysis. In total, 27 variants (14 novel) were identified with GS, of which four are intronic and predicted to be splice disrupting, 13 are missense, nine are leading to frameshifts or stop codons and one was a large deletion in *GJB6*. Six coding variants were missed by ES because of poor coverage and the rest were missed due to insufficient evidence for pathogenicity at the time of ES. Our results suggest that GS can yield improved detection of deafness variants compared to ES.

Mendelian Phenotypes Posters - Thursday

PB1858. Genomic and transcriptomic characterization of a novel *RYR1* variant in the French-Canadian population

Authors:

M. Labrecque^{1,2}, J. Hauteclouque^{1,3}, A. Rihoux^{1,3}, J-D. Brisson^{4,5}, C. Morin⁴, J. Mathieu^{4,5}, A. J. Parker^{1,3}, C-T. E. Nguyen^{3,6}, M. Tétreault^{1,3}; ¹CRCHUM, Montréal, QC, Canada, ²Dept. of Bioinformatics, Université de Montréal, Montréal, QC, Canada, ³Dept. of NeuroSci.s, Université de Montréal, Montréal, QC, Canada, ⁴Neuromuscular Diseases Clinic, Integrated Univ. Hlth.and Social Services Ctr., Saguenay-Lac-Saint-Jean, QC, Canada, ⁵Faculty of Med. and Hlth.Sci., Université de Sherbrooke, Sherbrooke, QC, Canada, ⁶Sainte-Justine Hosp. Ctr., Montréal, QC, Canada

Abstract Body:

Neuromuscular diseases are a group of diseases that affect the muscles, and the main symptom is muscle weakness. Myopathies are a type of neuromuscular diseases that affects the skeletal muscles and lead to a wide range of symptoms. *RYR1* related myopathies (RYR1-RM) are amongst the most common subtypes and is associated with phenotypes ranging from malignant hyperthermia to severe myopathies. The *RYR1* gene which codes for the ryanodine receptor 1 protein plays an important role in contraction of muscles. An uncharacterized mutation, p.E176K, has been identified in 28 individuals from 6 families in Quebec that show an inter- and intra-familial phenotypic variability. Our hypothesis is that the mutation is pathogenic and that an expression profile could explain the heterogeneity observed. To test the pathogenicity of p.E176K, a *C. elegans* model was made using CRISPR-cas9 genome editing. The p.E182K change in *C. elegans* UNC-68 is equivalent to p.E176K in humans. We then evaluated lifespan, reaction to aldicarb and motility in the modified worms. In a pilot project for the study of expression and pathways for RYR1-RM in humans, an RNA-seq pipeline was used on 4 patients with a myopathic phenotype and 2 controls. The RNA was extracted from muscle biopsies and sequenced. The differentially expressed genes (DEGs) were obtained with NOISeq and a gene enrichment exploratory analysis was made with GSEA to find differentially expressed pathways. In the *C. elegans* model, the lifespan of the mutant worms was significantly shorter than wildtype. In the presence of aldicarb, the muscle is not able to relax which eventually lead to paralysis. The paralysis took significantly longer in mutant worms compared to wildtype. For the motility, the mutant worm compared to the wildtype had a significant reduction of thrashing frequency and swim speed. In the exploratory analysis of DEGs in humans we found 36 upregulated genes and 168 downregulated genes, including *RYR1* and *ACHE* (Acetylcholinesterase). Afterwards, we found 32 upregulated pathways, including myogenesis in which *RYR1* and *ACHE* are implicated, and 18 that were downregulated. In the *C. elegans* model, all tests points to the p.E182K mutation being pathogenic which means the homologous p.E176K human mutation is also likely pathogenic. Interestingly, the delayed paralysis in mutant worms when presented with aldicarb shows a defect in synaptic transmission and possibly a lack of acetylcholinesterase that normally allows the muscle to relax. In humans, we also found that *RYR1* and *ACHE* were less expressed in patients, which corroborates the findings in the *C. elegans* model and offer an insight into a possible treatment for RYR1-RM.

Mendelian Phenotypes Posters - Wednesday

PB1859. Genomic ascertainment and reverse phenotyping diabetogenic variants in a phenotypically unselected cohort.

Authors:

C. Wilczewski¹, R. Kumar¹, R. Muniyappa², K. I. Rother³, D. Ng¹, K. Lewis⁴, P. A. Chan¹, H. Shiferaw¹, R. Semple⁵, C. Turner¹, A. Katz¹, L. Biesecker¹; ¹Natl. Human Genome Res. Inst., Bethesda, MD, ²Natl. Inst. of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, ³Eunice Kennedy Shriver Natl. Inst. of Child Hlth. and Human Dev., Bethesda, MD, ⁴NIH, Bethesda, MD, ⁵Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract Body:

Diabetes mellitus is defined simply by hyperglycemia, but is etiologically highly heterogeneous. In some cases, it is caused by monogenic variants in genes involved in insulin release, insulin receptor signaling, or lipodystrophy. As research into the phenotypic consequences of these variants has primarily relied on evaluating clinically affected individuals, our understanding of the phenotypic consequences of these variants in the broader population may be hampered by phenotypic selection bias. We sought to understand the effect of these variants in a phenotypically unselected population by genomically ascertaining individuals heterozygous for predicted deleterious variants in genes associated with metabolic disease, then recalled these individuals to perform reverse phenotyping for diabetes and lipodystrophy. This included collecting medical and family history, physical examination, clinical urine and blood-based laboratory assays, glucose tolerance tests, body fat quantification, and liver MRI. Out of 1514 exome sequencing datasets collected through the ClinSeq study, we identified 19 individuals with variants of interest (6 missense and 13 predicted loss of function) in 14 genes. Hemoglobin A1C levels from electronic health records indicated diabetes in one individual (7.6%) and prediabetes in 12 individuals (5.7-6.4%). Out of eight participants to complete reverse phenotyping, three had been previously diagnosed with type 2 diabetes and one had been diagnosed as prediabetic. Clinical examinations showed two genomically ascertained participants had monogenic diabetes, changing one diagnosis and making one new diagnosis. One participant had diabetes consistent with a previous diagnosis of familial partial lipodystrophy type 3. Five participants had evidence of monogenic impact on metabolism without meeting a diagnosis of diabetes. We highlight one individual harboring an *HNF1A* missense variant (NM_000545.8:c.788G>A, p.Arg263His) previously diagnosed with type 2 diabetes at 31 years old; his diagnosis was subsequently updated to Maturity Onset Diabetes of the Young (MODY). Cascade testing led to a diagnosis of MODY in his daughter. This work demonstrates how genomic ascertainment and reverse phenotyping can interrogate the clinical yield of variants associated with diabetes and lipodystrophy in a phenotypically unselected population, yielding more accurate estimates of variant pathogenicity in the broader population. Ultimately, this informs how large-scale genomic screening can enable precision medicine that impacts treatment and prevention of one of the most common chronic health conditions in the United States.

Mendelian Phenotypes Posters - Thursday

PB1860. Genomic region identified on chromosome 22 contributing to lymphedema in Phelan-McDermid syndrome

Authors:

S. Sarasua¹, M. Smith¹, C. Rogers², K. Phelan³, L. Boccuto¹; ¹Clemson Univ., Clemson, SC, ²Greenwood Gen Ctr, Greenville, SC, ³Florida Cancer Specialists, Fort Myers, FL

Abstract Body:

Lymphedema can be a painful and disabling co-morbid condition with many genetic syndromes including Phelan-McDermid syndrome (PMS). The cause of lymphedema is often unknown, it can be difficult to treat, and it tends to receive less research attention than other acute conditions. PMS is a rare neurodevelopmental disorder caused by deletions of varying size within chromosome band 22q13 or pathogenic variants of the *SHANK3* gene. The severity and type of clinical phenotypes in PMS are also highly variable. Parents and caregivers of individuals with PMS expressed concern that lymphedema was a problem for many people but had received little research in the PMS research community. Therefore, we sought to assess the prevalence and epidemiology of lymphedema in PMS and investigate genotype-phenotype correlations. To do this we assessed reports from the Phelan-McDermid Syndrome International Registry, now known as the PMS DataHub, on more than 400 individuals with information on this condition. Participants in the registry sample tended to be young with a median age of 7 years (range 1-49). We found a formal diagnosis of lymphedema was present in approximately 5% of people with PMS, with the highest prevalence in those over 18 years of age. Frequently co-occurring with lymphedema were cellulitis, pitting, and swelling in extremities. Lymphedema was more common in those with a deletion on 22q13 compared to a *SHANK3* pathogenic variant. Further, individuals with lymphedema also tended to have larger genomic deletions than those without lymphedema suggesting the contribution of additional gene(s). Therefore, an association analysis was conducted using Plink software which identified a specific genomic region at 22q13.31 with a major candidate gene for lymphedema. The results of this investigation can be used to help patients and families, as well as their treating clinicians, to understand lymphedema as a co-morbid condition and provide direction for targeted therapy research. Understanding the function of the candidate gene will also be useful in better understanding the genetic underpinnings of lymphedema in other genetic conditions. This research highlights the utility of assessing genotype-phenotype correlations in deletion syndromes, the use of large patient registries, and the benefits of patient-oriented approaches to research projects.

Mendelian Phenotypes Posters - Wednesday

PB1861. GGC repeat expansion within *NOTCH2NLC* causes behavioral deficits and neurodegeneration in a mouse model of Neuronal Intranuclear Inclusion.

Authors:

Y. Kang¹, Q. Liu², K. Zhang², Y. Li¹, P. Deng², Y. Li¹, Y. Tian², Q. Sun², Y. Tang², K. Xu², Y. Zhou², J-L. Wang², J. Guo², J-D. Li², K. Xia², Q. Meng³, E. Allen¹, Z. Wen¹, Z. Li⁴, H. Jiang², L. Shen², R. Duan², B. Yao¹, B. Tang², Y. Pan², P. Jin¹; ¹Emory Univ., Atlanta, GA, ²Central South Univ., Changsha, China, ³The First Affiliated Hosp. of Univ. of South China, Hengyang, China, ⁴The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract Body:

GGC repeat expansion within *NOTCH2NLC* gene has been identified as the genetic cause of neuronal intranuclear inclusion disease (NIID). To understand the molecular pathogenesis of NIID, here we establish both a transgenic mouse model and a human neural progenitor cell (hNPC) model. Expression of the *NOTCH2NLC* gene with expanded GGC repeats produces widespread intranuclear and perinuclear polyglycine (polyG), polyalanine (polyA) and polyarginine (polyR) inclusions, leading to severe neurodegeneration, motor dysfunction and cognitive deficits, which faithfully mimics the clinical manifestations and pathological features associated with NIID. Furthermore, conserved alternative splicing events was identified between the NIID mouse and hNPC models, among which are the enrichment of the binding motif of hnRNPM, an RNA-binding protein known as alternative splicing regulator. Expanded *NOTCH2NLC*-polyG and -polyA could interact with and sequester hnRNPM while overexpression of hnRNPM could ameliorate the cellular toxicity. These results together suggest that dysfunction of hnRNPM-mediated splicing could play an important role in the molecular pathogenesis of NIID.

Mendelian Phenotypes Posters - Thursday

PB1862. GWAS of cleft palate trios reveals novel associations and subtype-specific effects.

Authors:

K. Robinson¹, T. Mosley¹, T. H. Beaty², A. Butali³, C. Buxo⁴, G. Shaw⁵, J. Hecht⁶, L. Moreno⁷, J. Murray⁸, H. Brand⁹, S. Weinberg¹⁰, M. Marazita¹¹, E. Leslie¹; ¹Emory Univ., Atlanta, GA, ²Johns Hopkins Univ., Baltimore, MD, ³Univ. of Iowa, Coralville, IA, ⁴Univ. of Puerto Rico, San Juan, PR, ⁵Stanford Univ., Stanford, CA, ⁶Univ Texas McGovern Med Sch, Houston, TX, ⁷Univ Iowa, Iowa City, IA, ⁸U of Iowa, Iowa City, IA, ⁹MGH, Wilmington, MA, ¹⁰Univ of Pittsburgh, Pittsburgh, PA, ¹¹Univ Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA

Abstract Body:

Orofacial clefts (OFCs) are the most common craniofacial birth defects, occurring in approximately 1 in 1000 live births. OFCs are divided into cleft lip with or without cleft palate (CL/P) and cleft palate only (CPO), with CPO accounting for approximately 30% of all OFCs. For several reasons, CPO is considered etiologically distinct from CL/P, one of which includes differences in genetic risk factors. Compared to the dozens of risk loci associated with CL/P, only a few have been identified for CPO via common variant analyses. Historically, genome-wide studies have evaluated CPO as a whole, rather than by heterogeneous subtypes which include a spectrum of clefts affecting the hard and/or soft palate. However, differences in timing of midline fusion and variation in transcriptomic profiles between the anterior hard palate and the posterior soft palate suggest there may be distinct genetic drivers of each CPO subtype. To investigate this further, we sequenced 365 trios with CPO (118 unclassified, 69 hard and soft, 136 soft only, and 42 hard only) followed by transmission disequilibrium tests for genome-wide association in three groups: any cleft palate (ACP, n=365), any cleft of the hard palate (CHP, n=111), and any cleft of the soft palate (CSP, n=205). No SNP achieved genome-wide significance ($p < 5 \times 10^{-8}$), though a total of 39 loci reached suggestive significance ($p < 1 \times 10^{-5}$) in at least one analysis. Interestingly, only three of these loci overlapped (all shared between ACP and CSP), supporting the hypothesis of distinct risk factors for hard versus soft palatal CPO. The most significant locus was identified in the ACP group, and was closest to the gene *HOMER1* ($p = 4.20 \times 10^{-7}$, OR 3.29, 95% CI 2.02-5.36). Although not previously associated with CPO, Homer1 in mice is expressed in developing embryonic craniofacial tissue, especially within the palatal rugae. The top locus for CHP was on 9q33.3 ($p = 4.46 \times 10^{-7}$, OR 15.50, 95% CI 3.71-64.76), a region that when deleted results in a syndrome that often includes craniofacial dysmorphism and/or OFCs. Lastly, the top locus of interest for CSP was closest to *FAM120A* ($p = 2.38 \times 10^{-6}$, OR 5.57, 95% CI 2.49-12.46), a scaffold protein involved in many roles, including cell adhesion and apoptosis. Cumulatively, we have identified several novel risk loci yielding suggestive evidence of association for CPO in this study, and provided the first evidence for subtype-specific genetic variants. To further support these findings, future evaluations will include a replication cohort and functional testing to better characterize the role of these novel associations in CPO.

Mendelian Phenotypes Posters - Wednesday

PB1863*. GWAS of Down Syndrome Associated Atrioventricular Septal Defect identifies three novel loci.

Authors:

E. Feldman¹, D. J. Cutler¹, T. C. Rosser¹, S. B. Wechsler¹, L. Sanclemente², A. Rachubinski³, K. R. Rabin², M. Wagner⁴, B. Gelb⁵, J. M. Espinosa³, P. J. Lupo², S. L. Sherman¹, E. J. Leslie¹; ¹Emory Univ., Atlanta, GA, ²Baylor Coll. of Med., Houston, TX, ³Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ⁴Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, ⁵Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Congenital heart defects (CHDs) are the most common structural birth defect, affecting 8-10 of every 1000 live births and accounting for 25% of infant mortality. They represent a group of structural anomalies which range in severity some of which are incompatible with life. Atrioventricular septal defect (AVSD) is a form of CHD occurring in 1 in 5 infants with Down syndrome (DS) compared to 1 in 10,000 infants with euploid chromosome constitution. While increased dosage of chromosome 21 genes clearly contributes to this risk for AVSD, trisomy 21 is not sufficient to cause CHD as about 50% of infants with DS have structurally normal hearts. Thus, additional modifying genetic and environmental factors may exist. In order to characterize the genetic architecture of AVSD, we sequenced genomes of a multiethnic set of 1000 infants including 428 cases with DS and AVSD (AVSD+DS) and 572 controls with DS and a structurally normal heart (NH+DS) as part of the Gabriella Miller Kids First Pediatric Research Program and the INCLUDE (INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndrome) Project. We performed a genome-wide association study (GWAS) for common variants (MAF > 0.05) using logistic regression and adjusting for principal components of ancestry. Although no SNP achieved genome-wide significance ($p < 5 \times 10^{-8}$), three loci achieved suggestive significance ($p < 2 \times 10^{-6}$). The most significant locus was on 12q21.2 ($p = 2.22 \times 10^{-7}$, OR 0.55), near *NAV3*. Although not previously associated with CHD in humans, *nav3*^{-/-} mutant zebrafish were recently reported to have severe cardiac malformations. The other two lead SNPs at 13q21.33 ($p = 1.15 \times 10^{-6}$, OR 1.71) and 19q13.3 ($p = 3.41 \times 10^{-7}$, OR 1.98) are located in the vicinity of multiple genes expressed in the heart or associated with heart defects in human and/or animal models. Taken together, we have identified novel loci that have suggestive evidence for association with AVSD in DS. Additional analyses are underway to replicate these findings and to identify rare variants associated with AVSD and CHD in DS. The results from this study will generate new hypotheses about biological pathways associated with abnormal heart development due to trisomy 21.

Mendelian Phenotypes Posters - Thursday

PB1864*. Haploinsufficiency and loss-of-function of LEF-1 cause a novel Mendelian disorder by dysregulating WNT signaling

Authors:

M. Asif^{1,2,3}, S. Alawbathani^{1,2}, W. Dufour^{4,5}, A-S. Jourdain^{4,6}, G. Baujat⁷, C. Becker¹, B. Budde¹, L. Gallacher^{8,9}, T. Georgomanolis¹, J. Ghomid^{4,5}, W. Höhne¹, S. Lyonnet⁷, I. A. Ba-Saddik¹⁰, S. M. Hanu^{4,5}, S. Motameny¹, A. A. Noegel^{2,3}, L. Pais¹¹, C. Vanlerberghe^{4,5}, P. Wagle¹², S. M. White^{8,9}, M. Willems¹³, P. Nürnberg^{1,3}, F. Escande^{4,5}, F. Petit^{4,5}, M. S. Hussain^{1,2,3}; ¹Cologne Ctr. for Genomics, Cologne, Germany, ²Ctr. for Biochemistry, Med. Faculty, Univ. of Cologne, Cologne, Germany, ³Ctr. for Molecular Med. Cologne (CMMC), Univ. of Cologne, Faculty of Med. and Univ. Hosp. Cologne, Cologne, Germany, ⁴Univ. Lille, EA7364 RADEME, F-59000, Lille, France, ⁵CHU Lille, Clinique de génétique Guy Fontaine, F-59000, Lille, France, ⁶CHU Lille, Inst. de Biochimie et Biologie Moléculaire, F-59000, Lille, France, ⁷Hôpital Necker Enfants Malades, Service de génétique, Paris, France, ⁸Victorian Clinical Genetics Services, Murdoch Children's Res. Inst., Melbourne, Australia, ⁹Dept. of Paediatrics, Univ. of Melbourne, Melbourne, Australia, ¹⁰Dept. of Pediatrics, Faculty of Med. and Hlth.Sci., Univ. of Aden, Aden, Yemen, ¹¹Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, Cambridge, MA, ¹²Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), Univ. of Cologne, Cologne, Germany, ¹³Service de génétique, Hôpital Arnaud de Villeneuve, MONTPELLIER, France

Abstract Body:

Background: *LEF1* encodes a transcription factor acting downstream of the WNT- β -catenin signaling pathway. It was recently suspected as a candidate for ectodermal dysplasia in 2 individuals carrying 4q35 microdeletions. We recruited a highly diverse cohort of 12 patients from 5 unrelated families segregating de novo, monoallelic and biallelic variants of LEF1 inherited in both autosomal recessive and dominant fashion. All the patients manifested a novel Mendelian disorder, ectodermal dysplasia associated with limb malformations.

Methods: High-throughput sequencing was employed to delineate the genetic underpinnings of the disease. Cellular consequences were characterized by immunofluorescence, immunoblotting, pulldown assays, and/or RNA sequencing.

Results: Monoallelic variants in LEF1 were detected in 11 affected individuals from 4 unrelated families, and a biallelic variant was detected in an affected individual from a consanguineous family. The phenotypic spectrum includes various limb malformations, such as radial ray defects, polydactyly or split hand/foot, and ectodermal dysplasia. Our functional data indicate that 2 molecular mechanisms are at play: haploinsufficiency or loss of DNA binding are responsible for a mild to moderate phenotype, whereas loss of β -catenin binding caused by biallelic variants is associated with a severe phenotype.

Transcriptomic studies reveal an alteration of WNT signaling. Notably, WNT target genes, HOX genes and several long non-coding RNAs were differentially expressed in two of the patient derived fibroblasts.

Conclusion: Our findings establish mono- and biallelic variants in LEF1 as a cause for a novel syndrome comprising limb malformations and ectodermal dysplasia.

Mendelian Phenotypes Posters - Wednesday

PB1865. Haplotype in interleukin 6 is associated with a reduction of blood transfusion events in sickle cell anemia patients.

Authors:

C. Fong¹, E. Portilla², L. Cifuentes³; ¹Univ. Cooperativa de Colombia, Santa Marta, Colombia, ²Univ. de Nariño, Pasto, Colombia, ³Univ. Cooperativa de Colombia, Pasto, Colombia

Abstract Body:

Sickle cell disease (SCD) is the most commonest monogenic disease in the world, it is caused by the presence of hemoglobin S. SCD causes altered red-blood cells and hemolysis which drive to an inflammatory state that induces several symptoms. Interleukin 6 (IL6) is a central cytokine in the inflammatory response and polymorphisms in this gene are associated with susceptibility to inflammatory and autoimmune diseases. This study aimed to identify polymorphic sites in the IL6 gene in the Colombian population and to verify their influence on symptoms or blood transfusion events in SCD patients. We sequenced the promoter and coding region of IL6 in a sample of the Colombian population to identify genetic variants. Then, we collected clinical information from 112 patients with SCD. We identified three polymorphic SNPs (rs1800796, rs2069832, and rs2069849) which were genotyped in SCD patients, with these SNPs haplotypes were built. To evaluate both SNPs and haplotypes association with clinical data a logistic regression was performed. The symptoms more frequent in this population were pain crisis, osteomyelitis, and dactylitis. 77% of patients suffered a blood transfusion. Neither polymorphism was associated with symptoms or blood transfusion. Eight haplotypes were observed, the haplotype CGT showed a protective effect (OR= 0.4, p-value= 0.001) on blood transfusion. Our results show an association between IL6 and the events of blood transfusion, probably caused for a reduction of IL6 expression or a reduced amount of functional protein. The haplotype in this population identified could be used as a predictor of SCD severity.

Mendelian Phenotypes Posters - Thursday

PB1866. Hemizygous CNV in chromosome region Xq23 is associated with a complex psychiatric and bone mineral density disorder.

Authors:

J. Kapalanga¹, T. M. Parikh²; ¹Western Univ./GBHS, Owen Sound, ON, Canada, ²McMaster Univ., Hamilton, ON, Canada

Abstract Body:

The purpose of this report is to highlight that male inheritance of an X chromosomal copy number variant (CNV) from a clinically normal mother should not always be taken as evidence supporting the interpretation that it is benign. The CNV may have no deleterious impact in the heterozygous mother and a pathogenic effect in the nullizygous male offspring. We report on a hemizygous loss of 0.334 Mb copy number variant (CNV) in chromosome region Xq23, in a male proband with a complex behavioural and emotional profile, and a skeletal deformity. A maternal uncle was also found to have a constellation of similar clinical features. Parental studies by region-specific microarray analysis revealed that this CNV was inherited from a clinically normal mother. The deletion involves the entire length of *LUZP4* (leucine zipper protein 4), as well as the loss of the 5'UTR region of *PLS3* (plastin 3, OMIM 300131). The deletions within the 5'UTR of the *PLS3* gene and of the entire *LUZP4* gene, likely impacted gene expression and protein function of these genes in the proband and his uncle. The *LUZP4* hemizygous variants have been reported in two males with autism spectrum disorder (PMID: 23352160). Further *PLS3* is a an OMIM Morbid gene associated with X-linked dominant susceptibility to low mineral density skeletal disorders (BMND 18, OMIM 300910, OMIM 300910). The patient and his uncle have a constellation of social deficits, behavioral, and emotional features observed in patients with autism spectrum disorder. Both also have skeletal abnormalities attributable to low bone mineral density. This suggests that the CNV is pathogenic in males.

Mendelian Phenotypes Posters - Wednesday

PB1867*. Here, There and Everywhere: Characterizing Variability in Approaches to Ascertaining Individuals with Rare Mendelian Disorders from EHRs

Authors:

O. Veatch¹, J. Mathew², S. Rockowitz³, D. Baldrige⁴, A. Wetzel⁵, M. Niarchou⁶, M. Clarke⁷, S. Berger⁸, K. German⁹, S. Shankar¹⁰, P. Shankar¹⁰, J. S. Cohen¹¹, A. Keith⁴, I. Oh⁴, R. Raizada⁷, S. Soby⁷, M. Kaplarevic⁸, D. Doherty⁹, A. Gropman⁸, C. Smith-Hicks¹¹, J. L. Neul⁶, V. Lanzotti⁴, B. Darbro⁵, Q. Chang², M. Sahin³, M. Chopra³, on behalf of the NICHD IDDRC GREATestEHR Working Group; ¹Univ. of Kansas Med. Ctr., Kansas City, KS, ²Univ. of Wisconsin-Madison, Madison, WI, ³Boston Children's Hosp., Brookline, MA, ⁴Washington Univ. in St. Louis, St. Louis, MO, ⁵Univ. of Iowa, Iowa City, IA, ⁶Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁷Albert Einstein Coll. of Med., Bronx, NY, ⁸Children's Natl. Med. Ctr., Washington, DC, ⁹Univ. of Washington, Seattle, WA, ¹⁰MIND Inst. UC Davis, Sacramento, CA, ¹¹Kennedy Krieger Inst., Baltimore, MD

Abstract Body:

A critical step in translational medicine for rare Mendelian disorders is the rapid identification of research-eligible individuals. Liberating genetic testing results from electronic health records (EHRs) offers the opportunity to find potential subjects and assess rich phenotypic and genetic evidence. However, issues related to institutional differences in EHR configurations, lack of diagnostic codes for these disorders, and storage of genetic test results as unstructured text, PDFs or scanned images complicate developing reliable and efficient search strategies. This, in turn, impedes detection of individuals with rare monogenic disorders for clinical research. To identify methods likely to enhance the use of these data for research and clinical purposes, NICHD-supported Intellectual and Developmental Disabilities Research Centers (IDDRCs) formed the Genetic Reports - Enhancing Access to Test results in the Electronic Health Record ("GREATest EHR") Working Group. To move towards more efficient extraction of test results, eleven sites systematically compared the variability in methods and yield in identifying and extracting these data from EHRs. The four broad approaches used were: 1) manual searches, reviews, and extractions, 2) Natural Language Processing (NLP), 3) direct collaborations with genetic testing labs, and 4) incorporating EHR genomics modules being developed to aid with structured data capture. Of the eleven IDDRCs, eight perform searches for any text matching gene names of interest followed by manual review and extraction of data from each record with a gene name match. One site uses NLP trained to accurately recognize phenotype terms reflecting IDs, and manually reviews records within the subset to confirm genetic testing results and manually extract data. Four sites use a custom database developed by an external or in-house genetic testing lab. Four sites are in the process of implementing a genomics module developed by EHR vendors to streamline collection of structured genetic testing data. Strengths and weaknesses were noted for each approach. Notably, all approaches were time and labor intensive, requiring at least some level of manual review. Ongoing work includes evaluating the yield for each approach and developing more automated methods (e.g., optical character recognition, context specific search strategies) to optimize legacy data extraction across sites. The ultimate goals are to develop best practice guidelines to identify subjects eligible for translational research and to support clinical decisions about interpretation, reanalysis, new testing, and management.

Mendelian Phenotypes Posters - Thursday

PB1868. Heterogeneity of Inherited Cone Dysfunction Disorders with Normal Fundus Appearances

Authors:

K. Joo, H. Kim, W. Se Joon, K. Park; Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of

Abstract Body:

Purpose: To investigate the clinical and genetic characteristics of inherited cone dysfunction disorders with normal fundus appearances. **Methods:** Twenty-four patients with inherited cone dysfunction disorders with normal fundus appearances were evaluated. Exome sequencing targeting 245 candidate genes of inherited retinal diseases and direct Sanger sequencing were performed. Examinations including best-corrected visual acuities (BCVA), retinal morphologic abnormalities analyzed by spectral-domain optical coherence tomography (SD-OCT), longitudinal reflectivity ellipsoid zone (EZ) intensity ratio and electroretinogram (ERG) were performed. **Results:** The median age at the initial examination (n=24) and the final visit (n=17) were 11 years with a range of 4 to 59 years, and 21 years with a range of 8 to 63 years. Fifteen disease-causing disease variants were identified in fifteen genes, including *RGS9BP*, *RBP3*, *CNGA3*, *GNAT1*, *GNAT2*, *GUCY2D*, *CACNA1F*, *RAB28*, *RPGR1P1*, *PRPH2*, *CDHRI*, *RS1*, *PROM1*, *CNGB3*, and *PDE6B*. There were significant negative correlations between the BCVA and external limiting membrane-retinal pigment epithelium thickness (ERT), EZ intensity ratio at the initial examination ($P = 0.002$, $P < 0.001$), final visit ($P = 0.004$, $P = 0.002$), and the interval changes ($P = 0.006$, $P = 0.003$), respectively. ROC curve values for disease progression parameters of ERT differences were AUC 0.909, $P = 0.007$, and EZ intensity ratio differences were AUC 0.939, $P = 0.004$. **Conclusions:** Inherited cone dysfunction disorders with normal fundus appearances show clinically and genetically heterogeneous features. Retinal morphologic and reflectivity changes are identified in SD-OCT images, correlated to the visual function, and considered as predictive parameters for disease progression.

Mendelian Phenotypes Posters - Wednesday

PB1869. High Molecular Diagnostic Yields and New Phenotypic Expansions Involving Anorectal Malformations

Authors:

R. Belanger Deloge; Baylor Coll. of Med., Houston, TX

Abstract Body:

Evidence suggests that genetic factors contribute to the development of anorectal malformations (ARMs). However, the etiology of the majority of ARMs cases remains unclear. Exome sequencing (ES) may be underutilized in the diagnostic workup of ARMs due to uncertainty regarding its diagnostic yield. In a clinical database of ~17,000 individuals referred for ES, we identified 130 individuals with non-isolated ARMs (ARMs+). A definitive or probable diagnosis was made in 45 of these individuals for a diagnostic yield of 34.6% (45/130). The diagnostic yield of individuals who met criteria for VACTERL association was lower than those who did not (26.8% vs 44.1%; $p=0.0437$), suggesting that non-genetic factors may play an important role in this subset of ARMs+ cases. Within this cohort, we identified two individuals who carried de novo pathogenic frameshift variants in *ADNP*, two individuals who were homozygous for pathogenic variants in *BBS1*, and single individuals with carried pathogenic or likely pathogenic variants in *CREBBP*, *EP300*, *FANCC*, *KDM6A*, *SETD2*, and *SMARCA4*. The association of these genes with ARMs was supported by previously published cases, and their similarity to known ARMs genes as demonstrated using a machine learning algorithm. These data suggest that ES should be considered for all individuals with ARMs+ in whom a molecular diagnosis has not been made, and that ARMs represent a low penetrance phenotype associated with Helsmoortel-van der Aa syndrome, Bardet-Biedl syndrome 1, Rubinstein-Taybi syndromes 1 and 2, Fanconi Anemia group C, Kabuki syndrome 2, *SETD2*-related disorders, and Coffin-Siris syndrome 4.

Mendelian Phenotypes Posters - Thursday

PB1870. High prevalence of frontotemporal dementia in females of five Hispanic families with R159H VCP multisystem proteinopathy.

Authors:

A. Shmara¹, I. Gibbs¹, R. Mahoney¹, K. Hurth², V. Goodwill³, A. Cuber⁴, R. Im⁴, D. Pizzo³, J. Brown⁵, C. Laukaitis⁶, S. Mahajan⁷, V. Kimonis⁸; ¹UCI, Irvine, CA, ²USC, Los Angeles, CA, ³UCSD, San Diego, CA, ⁴Western Univ. of HS, Pomona, CA, ⁵Cure VCP Disease, Americus, GA, ⁶Carle Illinois Coll. of Med., Urbana, IL, ⁷Cedars Sinai Med. Ctr., Los Angeles, CA, ⁸Univ CA Irvine, Orange, CA

Abstract Body:

ABSTRACT Background and Objective Missense variants of the valosin-containing protein (*VCP*) gene cause a progressive, autosomal dominant disease termed VCP multisystem proteinopathy (MSP1). The disease is a constellation of clinical features including inclusion body myopathy (IBM), Paget's disease of bone (PDB), frontotemporal dementia (FTD) (IBMPFD) and amyotrophic lateral sclerosis (ALS), typically reported at a frequency of 90%, 42%, 30% and 9% respectively. The Hispanic population is currently underrepresented in previous reports of VCP myopathy. We expand our genotype-phenotype studies in five Hispanic families with the c.476G>A, p.R159H *VCP* variant. Methods We report detailed clinical findings of 11 patients in five Hispanic families with the c.476G>A, p.R159H *VCP* variant. Additionally, we report frequencies of the main manifestations in 39 affected members of the extended family members. We also compared our findings with an existing larger cohort of patients with VCP MSP1. Results FTD was the most prevalent feature reported, particularly frequent in females. PDB was only seen in one patient in contrast to the earlier reported cohorts. The overall frequency of the different manifestations: myopathy, PDB, FTD and ALS in these five families was 39%, 3%, 72% and 8% respectively. The atypical phenotype and later onset of manifestations in these families resulted in a noticeable delay in the diagnosis of VCP disease. Discussion Studying each *VCP* variant in the context of ethnic backgrounds is pivotal in increasing awareness of the variability of VCP-related diseases across different ethnicities, enabling early diagnosis, and understanding the mechanism for these genotype-phenotype variations.

Mendelian Phenotypes Posters - Wednesday

PB1871*. Highly recurrent histone H4 mutations cause a neurodevelopmental disorder

Authors:

L. Bicknell¹, F. Tessadori², K. Duran³, K. Knapp⁴, Histone H4 Clinical Consortium, P. Mace¹, G. van Haaften⁵; ¹Univ. of Otago, Dunedin, Otago, New Zealand, ²Hubrecht Inst.-KNAW, Utrecht, Netherlands, ³Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, ⁴Univ. of Otago, Dunedin, New Zealand, ⁵Univ. Med Ctr. Utrecht, Utrecht, Utrecht, Netherlands

Abstract Body:

Chromatin is essentially an array of nucleosomes, each of which consists of the DNA double-stranded fiber wrapped around a histone octamer. This organization supports cellular processes such as DNA replication, DNA transcription and genome integrity in all eukaryotes. The histone octamer is composed of two each of four different histones, including histone H4. Human histone H4 is encoded by fourteen canonical H4 genes, all of which are independently regulated and differ at the nucleotide level. These 14 genes encode an invariant protein, which is one of the most conserved proteins amongst eukaryotes. Here we present a cohort of 29 subjects with de novo missense variants in six H4 genes (*H4C3*, *H4C4*, *H4C5*, *H4C6*, *H4C9* and *H4C11*) identified by whole exome sequencing and genetic matchmaking. All individuals present with neurodevelopmental features of intellectual disability and motor and/or gross developmental delay, while non-neurological features such as facial dysmorphism, hearing loss and skeletal anomalies are more variable. Ten amino acids are affected, six of which recurrently mutated, and are all located within the H4 core or C-terminal tail. These variants cluster to specific regions of the core H4 globular domain, where protein-protein interactions occur with either other histone subunits or histone chaperones. Intriguingly, no variants were identified in the modification-heavy N-terminal tail of H4. Functional consequences of the identified variants were evaluated in zebrafish embryos, which displayed abnormal general development, defective head organs and reduced body axis length, providing compelling evidence for the causality of the reported disorder(s). While multiple developmental syndromes have been linked to chromatin-associated factors, missense-bearing histone variants (e.g. H3 oncohistones) have only more recently emerged as a major cause of pathogenicity. Our findings establish a broader involvement of H4 variants in developmental syndromes.

Mendelian Phenotypes Posters - Thursday

PB1872. Homozygous *EMCI* variant in four families from the same Kuwaiti tribe with cerebellar atrophy, visual impairment, and psychomotor retardation.

Authors:

M. Alenzi¹, R. Alqusaimi², E. El-anany³, A. Alholle⁴, S. Omar⁴, A. Elmonairy⁴, A. Alahmad⁴, R. Alsafi⁴, H. Alsharhan⁵, L. Bastaki⁴, B. Albash⁴, D. Marafie^{4,5,3}; ¹Farwaniya Hosp., Subah-Alnaser, Kuwait, ²Kuwait Univ., Health Science Center, Kuwait, ³Section of Child Neurology, Dept. of Pediatrics, Adan Hosp., Ministry of Hlth., Hadiya 52700, Kuwait, ⁴Kuwait Med. Genetics Ctr., Ministry of Hlth., Sulaibikhat 80901, Kuwait, ⁵Dept. of Pediatrics, Faculty of Med., Kuwait Univ., P.O. Box 24923, 13110 Safat, Kuwait

Abstract Body:

IntroductionThe subunit 1 of the endoplasmic reticulum (ER)-membrane protein complex is a conserved multi-subunit transmembrane complex. This protein enables energy-independent insertion of newly synthesized membrane proteins into ER membranes, mediating protein folding, phospholipid transfer from ER to mitochondria, and elimination of misfolded proteins. Both monoallelic *de novo* and biallelic *EMCI* variants have been identified to cause cerebellar atrophy, visual impairment, and psychomotor retardation (CAVIPMR) [OMIM # 61687]. Only five families with biallelic *EMCI* variants and CAVIPMR have been reported. Here, we describe four Kuwaiti families from the same tribe, with a homozygous pathogenic missense *EMCI* variant [c.245C>T:p.(Thr82Met)] and CAVIPMR.

MethodsClinical exome sequencing was performed in probands from three families while targeted molecular testing for *EMCI* [c.245C>T:p.(Thr82Met)] variant was performed in the fourth family based on strong clinical suspicion and tribal origin. Sanger sequencing confirmed the variant and segregation with disease in all families.

ResultsSeven individuals from four families affected by homozygous pathogenic *EMCI* variant [c.245C>T:p.(Thr82Met)] were identified. The variant was previously reported pathogenic in a Turkish family with CAVIPMR, and was absent from Kuwait Medical Genetic Centre database thus unlikely representing a population founder allele. The average age at symptom onset was 14.7 weeks, with all families reporting visual abnormalities, hypotonia, and/or global developmental delay (GDD) as the presenting features. Shared clinical features include GDD (7/7), cortical visual impairment (6/6), truncal hypotonia (6/6), failure to thrive (6/6) and hyperreflexia (5/5). Other common features include microcephaly (6/7; 86%), chorea (5/7; 71%), peripheral hypertonia (3/4; 75%), dysmorphism (3/5; 60%) and epilepsy (4/7; 57%). Brain imaging showed cerebellar atrophy in 2/5 (40%) and cerebral atrophy in 1/3 (33%) individuals.

ConclusionThe presence of exact biallelic homozygous *EMCI* variant in four families originating from the same tribe in Kuwait suggests a tribal founder allele. The clinical features herein are consistent with the phenotypic spectrum of *EMCI*-associated CAVIPMR in previous reports. The presence of chorea, first noted here, further expands the phenotypic spectrum. Our findings inform the clinical practice in Kuwait through recommended targeted *EMCI* variant [c.245C>T:p.(Thr82Met)] testing for infants from affected tribe presenting with visual impairment, GDD and hypotonia, and augment global efforts to understand rare Mendelian disorders.

Mendelian Phenotypes Posters - Wednesday

PB1873*. How many diseases can one gene cause: Why mechanism matters for gene curation, variant classification, and clinical management.

Authors:

K. Radtke, M. Rossi, W. Alcaraz, B. Wayburn, D. Thrush, H. Keller, M. Towne, J. Huang; Ambry Genetics, Aliso Viejo, CA

Abstract Body:

It is often thought that pathogenic variants in a gene cause a singular disorder; however, this “one-gene-one-disease” position is not universally true. One gene may cause multiple genetic disorders distinguished by differences in mode of inheritance (MoI), mechanism of disease (MoD), and/or genotype/phenotype correlations. Often, this means a pathogenic variant for one disorder does not cause the other associated disorder(s). An example, loss-of-function (LoF) *KLHL7* variants are associated with autosomal recessive neurodevelopmental disorder (NDD), and dominant negative variants are associated with autosomal dominant retinitis pigmentosa. This creates complexities in variant interpretation, clinical reporting, public variant databases submission, and patient counseling. Curating a combination of gene, MoI, MoD, and clinical presentation is critical for accurate variant classification. While this dilemma is known, the scope has not been thoroughly assessed. We reviewed a large set of NDD genes and all associated disorders. Our analysis did not include clinically distinct disorders with no genotype/phenotype correlation, differences in MoI, or MoD. Genes with more than one genetic disorder were grouped into 4 main categories: - monogenic dosage effect: distinct clinical presentations due to dosage effects (ex. *ATM*) - monogenic multimodal: similar clinical presentation with differences in MoI and MoD (ex. *GRINI*) - allelic: distinct clinical presentations and unique MoD (ex. *SMARCA4*) - unknown: unique clinical presentations but insufficient data about MoD We identified 901 distinct disorders associated with 800 NDD genes (1-4 disorders/gene). 11% (86/800) of genes are associated with >1 genetic disorder, totaling 187 distinct disorders. 2% (14/800) of genes are associated with >2 genetic disorders. These 187 disorders were categorized as: monogenic dosage effect 15%; monogenic multimodal 16%; allelic 50%; unknown 19%. Each of these categories require precise curation and have unique implications on variant interpretation, clinical reports, and patient counseling. Of the 86 genes associated with >1 disorder, we analyzed each disorder’s MoI and MoD. 31% (58/187) have different MoI and 42% (78/187) have different MoD. Further, 28% (53/187) differed in both MoI and MoD. Out of the 187 disorders, 41% (78/187) have a LOF mechanism, and approximately 8% (15/187) have a gain of function mechanism; the remaining 51% of disorders have other or unknown mechanisms. Functional confirmation of non-LoF MoD is highly nuanced but will be critical for accurately defining genetic disorders.

Mendelian Phenotypes Posters - Thursday

PB1874*. Human brains with Tay Sachs disease exhibit altered transcriptomes during fetal development

Authors:

S. Han¹, A. Hirt¹, E-R. Nicoli¹, C. Toro¹, M. Kono², R. L. Proia², C. J. Tiffit¹; ¹Natl. Human Genome Res. Inst., Bethesda, MD, ²Natl. Inst. of Diabetes and Digestive and Kidney Diseases, Bethesda, MD

Abstract Body:

Developing therapy for monogenic disorders requires deep understanding of disease pathogenesis with the ideal goal of compensating for the dysfunctional gene in every relevant cell type and at every time point that the gene is normally expressed. Understanding the temporal expression pattern, and correlating a therapeutic “window of opportunity”, is critical to prevent irreversible consequences. Tay Sachs disease (TSD), a lysosomal storage disorder (LSD) resulting from a deficiency of hexosaminidase A producing storage of GM2 ganglioside, is an appealing candidate for gene transfer therapy. Infants with TSD appear healthy at birth and typically manifest symptoms in the first weeks or months of life, providing a potential therapeutic window before symptom onset. However, accumulation of GM2 ganglioside begins during fetal development, raising the possibility of perturbed development despite the healthy presentation at birth. We confirmed GM2 accumulation in tissues from two 17-week human fetal brains with *HEXA* genotypes consistent with juvenile onset TSD and evaluated the transcriptomes of several brain regions by RNAseq. We found that differentially expressed gene sets were enriched for genes perturbed in neurodegenerative disorders including Alzheimer’s, Parkinson’s, Huntington’s, diseases as well as ALS. Moreover, transcriptomes from all regions of TSD brains appear delayed developmentally as they more closely resemble less differentiated control forebrains rather than the more differentiated regions from which they were sampled. Interestingly, we did not observe upregulation of immune genes, suggesting that the neuroinflammation typical of TSD and many neurodegenerative diseases has not yet begun. And, we did not observe upregulation of lysosomal genes previously noted in postnatal tissues in several LSDs. These data suggest that TSD disrupts normal brain development at least by mid-gestation, even before some of the pathological signs of disease have begun. These findings have obvious implications for gene or cell therapies and their ability to interrupt or correct a disease process that is already in place. The tissues used in this study were from the milder juvenile form of TSD and it is possible that pathology in the more severe infantile form will have even earlier onset and progress more rapidly. Therefore, when designing therapies, it will be important to select appropriate endpoints and, more importantly, set realistic outcome expectations for families even if treatment commences in the first few weeks of life.

Mendelian Phenotypes Posters - Wednesday

PB1875. Human mutations in *Slitrk3* implicated in GABAergic synapse development in mice

Authors:

S. Efthymiou¹, W. Han², M. Ilyas³, J. Li², Y. Yu⁴, M. Scala⁵, N. T. Malintan¹, R. Maroofian⁶, C. Rocca¹, V. Salpietro¹, S. Lakhani⁷, E. J. Mallack⁸, T. Palculict⁹, H. Li¹⁰, G. Zhang¹¹, F. Zafar¹², N. Rana¹², N. Takashima¹³, H. Matsunaga¹⁴, P. Striano⁵, M. F. Lythgoe⁴, J. Aruga¹⁵, W. Lu², H. Houlden¹; ¹UCL Inst. of Neurology, London, United Kingdom, ²Natl. Inst. of Neurological Disorders and Stroke, Bethesda, MD, ³IIUI, Islamabad, Pakistan, ⁴Ctr. for Advanced BioMed. Imaging, London, United Kingdom, ⁵Univ. of Genoa, Genoa, Italy, ⁶Inst. of Neurology UCL, London, United Kingdom, ⁷413 E 69th Street, New York, NY, ⁸Feil Family Brain and Mind Res. Inst., New York, NY, ⁹GeneDX, Gaithersburg, MD, ¹⁰Emory Univ Sch. of Med., Atlanta, GA, ¹¹Emory Univ. Sch. of Med., Atlanta, GA, ¹²Multan Childrens Hosp., Multan, Pakistan, ¹³RIKEN Brain Sci. Inst. (BSI), Wako-shi, Saitama 351-0198, Japan, ¹⁴Nagasaki Univ. Inst. of BioMed. Sci., Nagasaki 852-8523, Japan, ¹⁵Nagasaki Univ. Inst. of BioMed. Sci., Nagasaki, Japan

Abstract Body:

We report on biallelic homozygous and monoallelic *de-novo* variants in *Slitrk3* (*ST3*) in 3 unrelated families presenting with epileptic encephalopathy associated with a broad neurological involvement characterized by microcephaly, intellectual disability, seizures, and global developmental delay. *ST3* encodes for a transmembrane protein that is involved in controlling neurite outgrowth and inhibitory synapse development and that has an important role in brain function and neurological diseases. Using primary cultures of hippocampal neurons carrying patients' *ST3* variants and in combination with electrophysiology, we demonstrate that recessive variants are loss-of-function alleles. By analyzing the development and phenotype of *ST3* KO (*ST3*^{-/-}) mice, we bring additional evidence of enhanced susceptibility to pentylenetetrazole-induced seizure with the appearance of spontaneous epileptiform EEG, as well as developmental deficits such as higher motor activities and reduced parvalbumin interneurons. Taken together, our results exhibit impaired development of peripheral and central nervous system and support a conserved role of this transmembrane protein in neurological function.

Mendelian Phenotypes Posters - Thursday

PB1876. Hypophosphatemia gene panel sponsored program: a comparison of findings from 2019 through 2022 describing *PHEX* variant landscape and molecular diagnosis of XLH.

Authors:

N. Miller¹, E. Rush^{2,3,4}, K. Dahir⁵, O. K. Japalaghi¹, P. Ramesan¹, J. Yong⁶, H. McLaughlin⁷; ¹Ultragenyx Pharmaceutical Inc, Brisbane, CA, ²Univ. of Missouri – Kansas City Sch. of Med., Kansas City, MO, ³Children's Mercy, Kansas City, MO, ⁴Dept. of Internal Med., Univ. of Kansas Med. Ctr., Kansas City, KS, ⁵Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁶Invitae Corp., San Francisco, CA, ⁷Invitae, South Lyon, MI

Abstract Body:

X-linked hypophosphatemia (XLH), a dominant disorder caused by disease-associated variants in the *PHEX* gene, affects males and females of all ages. Patients with XLH have musculoskeletal phenotypes along with below-normal serum phosphate and elevated serum FGF23. **Methods:** Individuals aged ≥ 6 months with a clinical XLH diagnosis or suspicion of genetic hypophosphatemia, as evidenced by 2 or more clinical signs/symptoms or family history of XLH, are eligible for the no-charge genetic testing program. A next generation sequencing panel with copy number variant (CNV) detection initially included 13 genes: *ALPL*, *CLCN5*, *CYP2R1*, *CYP27B1*, *DMP1*, *ENPP1*, *FAH*, *FAM20C*, *FGF23*, *FGFR1*, *PHEX*, *SLC34A3* and *VDR*. *CTNS*, *GNAS*, *ORCL*, and *SLC34A1* were added in SEP 2020. Variants were interpreted according to the joint consensus from ACMG/AMP using Invitae's interpretation algorithm, Sherlock. Here we compare results from 27 FEB - 2 OCT 2019 (first 31 weeks) with results through 31 MAR 2022 (3 years, 4 weeks). **Results:** 317 unrelated probands were tested as of 2 OCT 2019. 244 probands (77%) had a *PHEX* variant: 216 (88%) pathogenic or likely pathogenic (P/LP) and 28 (11%) variants of uncertain significance (VUS). 122 unique *PHEX* variants (P/LP) were detected: 39 SNV (32%), 29 splicing/intronic (24%), 23 CNV (19%), 23 small deletions (19%), 6 small duplications (5%), and 1 each Alu insertion and complex rearrangement (<1% each). Additional data, including family member testing, clinical information, or family history, resulted in reclassification of 6 VUS to P/LP, impacting 21 individuals. Of the 73 (23%) cases where no *PHEX* P/LP/VUS was found, 7 (10%) had molecular diagnoses in other genes: *ALPL*, *ENPP1*, *FGF23*. As of 30 MAR 2022 the number tested increased to 2188 unrelated probands. 1101 probands (50%) had a *PHEX* variant (1034 [94%] P/LP and 67 [6%] VUS). 406 unique *PHEX* variants (P/LP) were detected with a similar distribution as in 2019. Additional data resulted in 32 VUS being reclassified to P/LP, impacting 107 individuals. Of 1087 (48%) individuals where no *PHEX* P/LP or VUS was found, 200 (18%) had molecular diagnoses in other genes: *ALPL* (166), *CLCN5* (1), *CTNS* (2), *CYP27B1* (9), *CYP2R1* (1), *DMP1* (2), *ENPP1* (1), *FGF23* (10), *GNAS* (1), *ORCL* (2), *SLC34A3* (4), and *CTNS/ALPL* (1). **Conclusions:** We report a 3.3-fold increase in unique *PHEX* P/LP variants since 2019. Variant subtype distribution remains similar. *PHEX* VUS decreased from ~11% in 2019 to ~5 % over the same period in 2021. Cascade testing and VUS resolution aids molecular diagnosis of XLH.

Mendelian Phenotypes Posters - Wednesday

PB1877. Identification and Functional Evaluation of Autosomal Recessive Non-Syndromic Hearing Impairment Genes in Rwanda.

Authors:

E. Uwibambe¹, L. Mutesa², A. Wonkam³; ¹Univ. of Rwanda, Kigali city, Rwanda, ²Univ. of Rwanda, Kigali, Kigali City, Rwanda, ³Johns Hopkins Univ. Sch. of Med., Univ Cape Town, Cape Town, South Africa

Abstract Body:

The incidence of hereditary hearing impairment (HI) is higher in developing countries compared to developed countries and more than 120 independent genes have been identified as responsible for almost 50% of profound HI. Nonsyndromic HI is the most common form accounting for 70% of cases of which 80% are autosomal recessive. Reported mutations in GJB2, GJB6, and GJA genes are the most common cause of HI but studies among Cameroonian and South African participants did not identify a significant association, hence the need for further genetic exploration of other responsible genes in the African population. In Rwanda, more than 50% of HI among children has been attributed to hereditary causes but no genetic evidence has been established. This study aims at identifying genes responsible for autosomal recessive non-syndromic hearing impairment (ARNSHI) among the Rwandan population. We will recruit 25 familial cases, 100 isolated cases, and 100 control individuals without HI or a family history of HI. One proband per familial affected family will be recruited together with one unaffected sibling, both parents, and at least one affected relative for blood sampling. Standard audiometry including air and bone conduction will be performed, and clinical history will be assessed to exclude exposure to ototoxic drugs or infections including prenatal exposure. For each family, we will exome-sequence samples of the affected members and use Sanger sequencing to follow up variants segregating in their parents and non-affected sibling, and the controls. At the end of the study, we expect to establish a dataset of genes that cause ARNSHI in Rwanda with possible novel ARNSHI variants that have not yet been found in the African population. This will help to advance the science of HI and establish appropriate medical care including early genetic screening, possible genetic therapies, proper genetic counseling necessary for affected individuals and families as well as plan for prevention and control measures against genetic HI.

Mendelian Phenotypes Posters - Thursday

PB1878. Identification of a novel compound heterozygous mutation in *RyR1* gene in an Indian family affected with congenital myopathy

Authors:

L. Kirola¹, D. Joshi², T. Murayama³, T. Sakurai³, A. Mukherjee¹, M. Mutsuddi¹; ¹Dept. of Molecular and Human Genetics, Inst. of Sci., Banaras Hindu Univ., Varanasi, India, ²Dept. of Neurology Inst. of Med. Sci. Banaras Hindu Univ., Varanasi, India, ³Dept. of Cellular and Molecular Pharmacology, Juntendo Univ. Graduate Sch. of Med., Tokyo, Japan

Abstract Body:

In the ryanodine receptor family (RyR), three genes (*RyR1*, *RyR2*, and *RyR3*) are involved in Ca²⁺ homeostasis, storage, and regulation. The RyR isoforms also play a vital role in extracellular and intracellular calcium signaling in many cell types, including muscles, neurons, and epithelial cells. Mutations in *RyR1* cause a wide range of clinical phenotypes, including several congenital myopathies (CM), central core disease (CCD), and hyperthermia susceptibility. *RyR1*-related CCDs usually show clinical heterogeneity and an early-onset of disease pathogenesis. Here, we present a family that includes unaffected parents and three siblings who are affected with CCD since childhood. The clinical features include lower proximal muscle weakness, difficulties in standing up and climbing, skeletal malformations, and hypotonia since birth. Muscle examinations (e.g., NCV, EMG, and muscle MRI) showed weak muscle intensity and activity. Whole-exome sequencing was performed in all the affected siblings using NovaSeq2000. Bioinformatic analysis and filtering of multiple variants revealed a novel variant in *RyR1*. This compound heterozygous variant (c.A5096G;p.D1699G+c.C5097A:p.D1699E; 13423_13424del:p.K4475Efs*106) has not been reported in public databases and *in silico* analysis predicted the variant to be damaging. Furthermore, this novel variant has segregated within the family and *in silico* protein analysis showed putative changes in the protein activity between the wildtype versus mutant *RyR1*. The functional analysis of the causal *RyR1* variants is being further investigated by our group. This study will help in the prenatal diagnosis of CM in the affected family and will also provide a platform for future therapeutics in *RyR1*-related diseases. Our study will also provide a genotype-phenotype correlation.

Mendelian Phenotypes Posters - Wednesday

PB1879. Identification of a novel pathogenic variant in *HIST1H1E* encoding the H1 histone linker in a patient with a complex phenotype including progeroid feature, intellectual disability and lipodystrophy. Contribution of additional variants to the complexi.

Authors:

S. Mellone¹, D. Vurchio², S. Ronzani¹, T. Daffara³, M. Romanisio³, G. Ceccarini⁴, C. Pelosini⁴, F. Santini⁴, F. Prodam³, M. Giordano⁵; ¹Lab. of Genetics, “Maggiore della Carità” Hosp., Novara, Novara, Italy, ²Lab. of Genetics, “Maggiore della Carità” Hosp. and Dept. of Hlth.Sci., Università del Piemonte Orientale, Novara, Novara, Italy, ³Endocrinology Unit and Dept. of Hlth.Sci., Università del Piemonte Orientale, Novara, Novara, Italy, ⁴Obesity and Lipodystrophy Ctr., Endocrinology Unit, Univ. Hosp. of Pisa, Novara, Italy, ⁵Lab. of Genetics, “Maggiore della Carità” Hosp., Novara and Dept. of Hlth.Sci., Università del Piemonte Orientale, Novara, Novara, Italy

Abstract Body:

Here we report a case of an adult patient with type 2 diabetes, hypercholesterolemia that at physical examination presented: progeroid features, prominent forehead, hypertelorism, broad nasal bridge, baldness, hypotricosis, hearing loss and lost teeth at ten years of age. He showed developmental delay and ID, mild atrial septal defect, mild osteoporosis and a monolateral low testis volume, with subclinical hypergonadotroph hypogonadism. The BMI was normal and subcutaneous fat was underrepresented in leg and arms, but increased in the abdomen. Since the abnormal distribution of fat and metabolic alterations at a young age, we suspected a progeroid lipodystrophy syndrome. Leptin (1.9 mcg/L) was quite undetectable and the genes known to cause classical lipodystrophy syndromes were wild type as well as CGH array. A whole genome sequencing (WGS) of the proband and his parents was thus performed revealing a de novo pathogenic frameshift variant in *HIST1H1E*, p.Ser150fs,c.447dupC, encoding the H1 histone linker that acts as a structural component of the chromatin controlling the packaging of DNA, gene expression and DNA replication/repair. Mutations in this gene perturb multiple cellular process resulting in cellular senescence and alteration of genome methylation pattern and are characteristic of Rahman Syndrome, a recently recognized developmental disorder characterized by ID, dysmorphism and an aging appearance. Variable somatic overgrowth may manifest in early infancy but is not observed in adults that often display decreasing height percentile over time. The WGS analysis also revealed inherited pathogenic variants in the *LDLR* gene explaining hypercholesterolemia, in *OTOGL* responsible for neurosensorial hearing loss and in *FBNI* possibly correlated to the lipodystrophy phenotype. This case represents a complex phenotype that remained without a clinical diagnosis for a long time and that reached a definitive diagnosis on the basis of the results obtained through WGS with at least four genes explaining most of his phenotypic characteristics.

Mendelian Phenotypes Posters - Thursday

PB1880*. Identification of a Novel Therapeutic Target Underlying Atypical Manifestation of Gaucher Disease

Authors:

H. Do¹, S. Kim¹, I. Choi², H. Bae³, G-H. Kim⁴, K. Kyunggon⁵, B. Lee⁶; ¹Asan Inst. for Life Sci., Seoul, Korea, Republic of, ²Asan Med. center, Seoul, Korea, Republic of, ³Med. Genetics Ctr., Asan Med. Ctr. Children's Hosp., Seoul, Korea, Republic of, ⁴Asan Med Ctr, Seoul, Seoul, Korea, Republic of, ⁵Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, ⁶Asan Med Ctr, Seoul, Korea, Republic of

Abstract Body:

Gaucher disease (GD) is a lysosomal storage disorder caused by glucocerebrosidase 1 (GCase) deficiency. This GCase deficiency leads to accumulating glucosylceramide, glucosylsphingosine (Lyso-Gb1), a deacylated form glucosylceramide in the lysosome of macrophages. Therapeutic unmet needs such as lymphadenopathy and neurological manifestations have been observed despite enzyme replacement therapy (ERT), and the role of Lyso-Gb1 in these manifestations has been poorly understood. In the current study, the patients' lymph nodes, plasma, and fibroblasts were examined by multi-dimensional in vitro and in vivo studies. In a type 3 GD patient with severe mesenteric lymphadenopathy despite ERT, atypical Gaucher-like cells were found with intense Lyso-Gb1 accumulation as well as multinucleation, surrounded by fibrous band-like structures. Detailed immunohistochemistry and tissue proteomic analysis revealed altered complement and subsequent aberrant macrophage M1-M2 polarization underlying this aberrant phenotype. In addition, autophagosome formation was increased in these atypical cells, but the autophagy flux was impaired, resulting in low autolysosome formation and a subsequent severe inflammatory response. Importantly, we observed differential expression of proteins related to endothelial mesenchymal transition, in particular, an altered TGF- β signaling pathway. In line with these observations, TGF- β was increased in the plasma and fibroblasts of patients with type 3 GD, which was reduced by high-dose ambroxol treatment. In conclusion, our study demonstrated that complex molecular pathways underlie the unusual progression of GD, leading to ERT unresponsiveness, and a multi-functional therapeutic approach with combination therapy is required to relieve this devastating process.

Mendelian Phenotypes Posters - Wednesday

PB1881. Identification of a pathogenic gene mutation in Med12

Authors:

G. Serrano Rodriguez¹, A. Lebron², G. Arroyo Figueroa², N. Arciniegas³, S. Carlo⁴, A. Cornier⁵; ¹Ponce Hlth.Sci. Univ., Ponce, PR, ²Univ. of Puerto Rico, San Juan, PR, ³Mayaguez Med. Ctr., Cabo Rojo, PR, ⁴Ponce Hlth.Sci. Univ., Cabo Rojo, PR, ⁵San Jorge Children s Hosp./ Ponce Hlth.Sci. Univ., San Juan, PR

Abstract Body:

This case report presents a 12 year old boy with learning disabilities such as speech delay, fine, gross motor problems, autism, hypotonia, low ear set, and keratoconus. The patient was referred for genetic evaluation that resulted in genetic variants on PRMD5 and a novel pathogenic variant on Med12 p.(Asn1591Ser). The patient mother and sister were also evaluated for the MED12 variation. Med 12 is a component of the multiprotein complex known as Mediator, which is involved in the transcription regulation of several cellular processes. Mutations in different regions of Med 12 are related to x-linked intellectual disability syndromes and other abnormalities cataloged as Med-12 related disorders. We proposed that the variable expressivity observed on Med 12-related disorders is a consequence of allelic heterogeneity supporting a continuum of symptoms instead of different syndromes. New variants on Med12 are continuously broadening its genome architecture and phenotype expression, allowing for better genetic counseling and management of these patients.

Mendelian Phenotypes Posters - Thursday

PB1882. Identification of *INHBA*-related neurodevelopmental disorder

Authors:

T. Pierson^{1,2,3}, K. L. Park⁴, H. May⁵, M. Ross⁶, A. Iglesias⁷, F. Faletra⁸, L. Musante⁹, K. Grand¹⁰, J. A. Golden¹¹, J. M. Graham¹⁰; ¹Dept. of Pediatrics, Cedars-Sinai Med. Ctr., Los Angeles, CA, ²Dept. of Neurology, Cedars-Sinai Med. Ctr., Los Angeles, CA, ³Board of Governors Regenerative Med. Inst., Cedars-Sinai Med. Ctr., Los Angeles, CA, ⁴Univ. of Colorado Sch. of Med., Aurora, CO, ⁵Columbia Univ. Irving Med. Ctr., New York, NY, ⁶Columbia Univ. Med. Ctr., New York, NY, ⁷Columbia Univ. Med. Ctr., HASTINGS ON HUDSON, NY, ⁸IRCCS Burlo Garofolo, Trieste, Italy, ⁹Inst. for Maternal and Child Hlth.- IRCCS “Burlo Garofolo”, Trieste, Italy, ¹⁰Med. Genetics, Dept. of Pediatrics, Cedars-Sinai Med. Ctr., Los Angeles, CA, ¹¹Dept. of Pathology, Vice Dean Res. and Graduate Ed., Burns and Allen Res. Inst., Los Angeles, CA

Abstract Body:

Activins and Inhibins are composed of homo- and hetero-dimers of beta or alpha inhibin subunits. Beta-A and beta-B subunits, which are encoded by the *INHBA* and *INHBB* genes, can homodimerize to form the activins (Activin AA, AB, or BB); while the alpha subunit, encoded by *INHA*, heterodimerizes with either beta subunit to form inhibins (inhibin A or B). These compounds are part of the TGF-beta superfamily and play critical roles in development. In general, Activin homodimers are antagonized by Inhibin heterodimers. *INHBA* is located at 7p14.1; it has 3 exons and initiation of the beta-A coding sequence begins in exon 2. Mice deficient in the beta-A subunit (*Inhba*-null) have abnormal tooth development as Activin A is thought to regulate keratocyte development. Activin A may be involved in myofibroblast transdifferentiation as well. *Inhba*-null mice develop to term, but die within 24 hours after birth with cleft palate and absent whiskers and lower incisors, suggesting beta-A is involved in craniofacial development as a result of deficiency of either Activin A, AB, or Inhibin A. Ablation of *Inhba* expression in fetal mouse Leydig cells resulted in testis cord dysgenesis due to decreased Sertoli cell proliferation. We have identified 4 unrelated patients with de novo heterozygous variants in *INHBA* via exome sequencing (three missense variants involving residues within the mature domain and prodomain and a nonsense mutation involving the signal sequence). Subjects possessing the three missense variants had similar phenotypes associated with mild/moderate intellectual disability (ID) and primary generalized epilepsy. The subject possessing the nonsense variant phenotype had mild ID without epilepsy and a normal EEG. Dysmorphic facial features included hypertelorism and prominent ear lobes in two of the missense cases. The single male had delayed puberty, and 66% of females had endocrine abnormalities, although stature was normal. Tooth development was not examined in detail. Functional studies of these variants are planned in the future.

Mendelian Phenotypes Posters - Wednesday

PB1883. Identification of molecular causes of recessively inherited ataxic and neuropathic disorders in consanguineous Pakistani families.

Authors:

F. Aslam^{1,2}, M. Wajid³, E. Wohler⁴, N. Sobreira⁴, G. H. Seo⁵, W. Ji², S. Lakhani², S. Naz¹; ¹Univ. of the Punjab, Lahore, Pakistan, ²Yale Univ., New Haven, CT, ³Univ. of Okara, Okara, Pakistan, ⁴John Hopkins Univ., Baltimore, MD, ⁵3billion inc, Seoul, Korea, Republic of

Abstract Body:

Autosomal recessive ataxia is characterized primarily by gait and balance problems. Neuropathy is a condition in which nerve problems are manifested as distal muscle weakness, hand and foot deformities, sensory loss and dysphonia. We recruited seven consanguineous families having patients with symptoms of neuropathy or ataxia for the identification of possible genetic defects underlying their disease conditions. Exome sequencing was completed for multiple individuals from each family. Web-based server wANNOVAR, online Franklin tool and 3billion automated prioritization system were used for data annotation, prioritizing homozygous, exonic and splicing variants with frequency less than 0.01 in all public databases including gnomAD. Conservation of affected amino acids was checked by multiple protein alignments using Clustal Omega. Candidate variants were checked for segregation with the disease phenotype by Sanger sequencing of DNA from all participants. In members of five out of seven families, we identified variants, classified as pathogenic by multiple online prediction tools. These included homozygous nonsense, frameshift and missense variants in four genes: two families with different variants in *GDAP1*, and one family each with variants in *AFG3L2*, *MFSD8* and *SETX*. Variants of *GDAP1* were associated with neuropathic disorder while those of *AFG3L2* and *MFSD8* led to ataxic disorder without neuropathy. *SETX* variant was linked to ataxia with associated neuropathic syndrome. Pathogenic variants in *MFSD8* have been previously associated with childhood onset ceroid neuronal lipofuscinosis that cause vision problems and early death. Patients evaluated in this study reported no vision problems and were alive in their late fifties. No genetic cause was identified in the patients from two families, each having symptoms of ataxia, dysarthria, strabismus and spasticity. However, in one of these two families, we identified a frameshift variant in *ALS2* in one patient, but the same variant was not identified in his affected first cousins. In summary, exome sequencing identified the molecular causes of recessive ataxia or neuropathy in most of the recruited families and has broadened the clinical spectrum of *MFSD8* and mutational spectrum of others.

Mendelian Phenotypes Posters - Thursday

PB1884. Identifying modifier genes in a PIGA-CDG pedigree with reduced penetrance

Authors:

H. Thorpe, B. Pedersen, J. Bonkowsky, A. Quinlan, C. Chow; Univ. of Utah, Salt Lake City, UT

Abstract Body:

Phosphatidylinositol glycan class A (PIGA) encodes the catalytic component of the enzyme in the first step of GPI-anchor biosynthesis. GPI-anchors act as membrane anchors for over 150 proteins involved in signal transduction, immune response, and cellular communication among other functions. Loss of function mutations in *PIGA* lead to PIGA deficiency (PIGA-CDG), an ultra-rare disease, typically presenting with seizures, hypotonia, and neurodevelopmental delays. PIGA deficiency is an X-linked recessive congenital disorder of glycosylation (CDG). We identified two brothers (proband) with PIGA-CDG, presenting with mild developmental delay, epilepsy, and autism. Both probands carry the novel, rare *PIGA*^{S132C} variant, a predicted damaging variant not found in the gnomAD database. Confirming this diagnosis, both probands also show a 50% decrease in GPI-anchor protein on the cell surface. Strikingly, the maternal grandfather and a great uncle both also carry *PIGA*^{S132C}, but neither presents with symptoms associated with PIGA-CDG. We hypothesized that there might be a modifier segregating in the family that contributes to this reduced penetrance. Using whole genome sequencing and pedigree analysis, we identified all the possible susceptibility variants found in the probands and not in carriers and all the possible protective variants found in the carriers and not in the probands. This list of potential candidates included a single, damaging rare variant in two other genes also involved in GPI-anchor biosynthesis, *PIGS* and *DPM1*. To functionally test our predicted modifiers, we are using a *Drosophila* eye-based model of PIGA-CDG. We created double knockdowns of *PIGA* and the candidate modifiers and compared the eye sizes of the double knockdowns to the PIGA eye model and single knockdowns of the candidate modifiers. Functional and genetic analyses in this *Drosophila* model indicates that susceptibility and protective modifier genes are plausible human modifiers and could explain the incomplete penetrance in this family. The identification and study of rare disease modifier genes in human pedigrees may lead to pathways and targets that may be developed into therapies.

Mendelian Phenotypes Posters - Wednesday

PB1885. Identifying phenotypic expansions for congenital diaphragmatic hernia plus (CDH+) using DECIPHER data.

Authors:

A. Hardcastle¹, W. Chung², C. Shaw³, A. M. Berry³, I. Campbell⁴, X. Zhao³, P. Liu³, A. Gerard⁵, J. Rosenfeld³, S. Sisoudiya³, A. Hernandez-Garcia³, S. Loddo⁶, S. Tommaso⁷, A. Novelli⁷, M. Dentici⁷, R. Capolino⁷, M. Digilio⁷, L. Graziani⁷, C. Rustad⁸, K. Neas⁹, G. Ferrero¹⁰, A. Brusco¹⁰, E. Gregorio¹¹, D. Wellesley¹², C. Beneteau¹³, M. Joubert¹³, K. Van Den Bogaert¹⁴, A. Boogaerts¹⁴, D. McMullan¹⁵, D. Scott³; ¹Brigham Young Univ., Provo, UT, ²Columbia Univ., New York, NY, ³Baylor Coll. of Med., Houston, TX, ⁴Children's Hosp. of Philadelphia, Philadelphia, PA, ⁵Texas Children's Hosp., Houston, TX, ⁶CSS-Mendel Inst, Rome, Italy, ⁷IRCCS, Rome, Italy, ⁸Oslo Univ. Hosp., Hosle, Norway, ⁹Genetic Hlth.Service NZ, Wellington, New Zealand, ¹⁰Univ. of Torino, Torino, Italy, ¹¹Città della Salute e della Scienza Univ. Hosp., Torino, Italy, ¹²Princess Anne Hosp., Southampton, United Kingdom, ¹³Nantes Université, Nantes, France, ¹⁴Univ. Hosp. Leuven–KU Leuven, Leuven, Belgium, ¹⁵Birmingham Women S NHS foundation Trust, Birmingham, IN, United Kingdom

Abstract Body:

Congenital diaphragmatic hernia (CDH) can occur in isolation or in conjunction with other birth defects (CDH+). A molecular etiology can only be identified in a subset of CDH cases. This is due, in part, to an incomplete understanding of the genes that contribute to diaphragm development. Here, we used clinical and molecular data from 36 individuals with CDH+ who are catalogued in the DECIPHER database to identify genes that may play a role in diaphragm development and to discover new phenotypic expansions. Among this group, we identified individuals who carried putatively deleterious sequence or copy number variants affecting *CREBBP*, *SMARCA4*, *UBA2*, and *USP9X*. The role of these genes in diaphragm development was supported by their expression in the developing mouse diaphragm, their similarity to known CDH genes using data from a previously published and validated machine learning algorithm, and/or the presence of CDH in other individuals with their associated genetic disorders. Our results demonstrate how data from DECIPHER, and other public databases, can be used to identify new phenotypic expansions and suggest that *CREBBP*, *SMARCA4*, *UBA2*, and *USP9X* play a role in diaphragm development.

Mendelian Phenotypes Posters - Thursday

PB1886. *In Silico* genomic interrogation reveals 16 autosomal dominant neurodevelopmental candidate genes at 1p13.3

Authors:

A. Ben Mahmoud¹, K. Jun², V. Gupta³, P. Shastri⁴, J. Rosenfeld⁵, A. de la Fuente¹, Y. Park¹, K. Shin¹, C. Kim⁶, A. da Cruz⁷, I. Pinto⁸, L. Minasi⁸, A. da Cruz⁸, L. Faivre⁹, P. Callier¹⁰, C. Racine¹¹, L. Layman¹², I-K. Kong¹³, W-Y. Kim¹⁴, H-G. Kim³; ¹Qatar BioMed. Res. Inst., Doha, Qatar, Qatar, ²Dept. of Lab. Med., Inje Univ. Haeundae Paik Hosp., Busan, Korea, Korea, Republic of, ³Qatar BioMed. Res. Inst., Doha, Qatar, ⁴Dept. of Internal Med., Univ. of Toledo, Toledo, OH 43614, USA, Toledo, USA, OH, ⁵Baylor Coll. of Med., Houston, TX, ⁶Unidade de Genética do Inst. da Criança - Hosp. das Clínicas HCFMUSP, Faculdade de Med., São Paulo, Brazil, Brazil, ⁷Human Cytogenetics and Molecular Genetics Lab., Goiás, Brazil, Brazil, ⁸Sch. of Med. and Life Sci., Pontifical Catholic Univ., Goiás, Brazil, Brazil, ⁹CHU Dijon, Dijon, France, ¹⁰CHU Le Bocage, Dijon, France, ¹¹CHU de Dijon, DIJON, France, ¹²Med Coll GA at Augusta Univ., Augusta, GA, ¹³Dept. of Animal Sci., Div. of Applied Life Sci., Gyeongsang Natl. Univ., Jinju, Korea, Korea, Republic of, ¹⁴Dept. of Biological Sci., Kent State Univ., Kent, USA, OH

Abstract Body:

Genome-wide chromosomal microarray is extensively used to detect copy number variations (CNVs), which can explain undiagnosed microdeletion and microduplication syndromes. These small unbalanced chromosomal structural rearrangements ranging from 1 kb to 10 Mb comprise up to 15% of human mutations leading to monogenic or contiguous genomic disorders. Albeit rare, CNVs at 1p13.3 cause a variety of neurodevelopmental disorders (NDDs) including development delay (DD), intellectual disability (ID), autism, epilepsy, and craniofacial anomalies (CFA). Most of the 1p13.3 CNV cases reported in the pre-microarray era encompassed a large number of genes and lacked the demarcating genomic coordinates, hampering the discovery of positional candidate genes within the boundaries. Here, we present four subjects with 1p13.3 microdeletions displaying DD, ID, autism, epilepsy, and CFA. *In-silico* comparative genomic mapping with previous reported three CNV subjects with 22 unreported DECIPHER CNV cases has resulted in the identification of four different sub-genomic loci harboring five positional candidate genes for DD, ID and CFA at 1p13.3. Most of these genes have pathogenic variants reported and their interacting genes are involved in NDDs. RT-qPCR in various human tissues revealed a high expression pattern in the brain and fetal brain, confirming their functional roles in NDDs. Interrogation of variant databases and interacting protein partners led to the identification of another set of 11 potential candidate genes, which might have been disrupted by the position effect of these CNVs at 1p13.3. Our studies define 1p13.3 as a genomic region harboring 16 NDD candidate genes and underscore the critical roles of small CNVs in *in silico* comparative genomic mapping for disease gene discovery. Our candidate genes will help accelerate the isolation of pathogenic heterozygous variants from the whole exome/genome sequencing (WES/WGS) database.

Mendelian Phenotypes Posters - Wednesday

PB1887. Incontinentia pigmenti female with the *IKBKG/NEMO*del4-10 deletion: A mosaic form.

Authors:

M. Ursini¹, E. Spinosa¹, M. Salvia¹, C. Casale¹, A. Pescatore¹, A. Torella^{2,3}, G. Piluso², V. Nigro^{2,3}, V. Piccolo⁴, A. Diociaiuti⁵, M. El Hachem⁵, F. Fusco¹; ¹IGB ABT CNR, Naples, Italy, ²Dept. of Precision Med., Univ. of Campania "Luigi Vanvitelli", Naples, Italy, ³Telethon Inst. of Genetics and Med. (TIGEM), Naples, Italy, ⁴Dermatology Unit, Univ. of Campania Luigi Vanvitelli, Naples, Italy, ⁵Dermatology Unit, Bambino Gesù. Children's Hosp.—IRCCS, Rome, Italy, Rome, Italy

Abstract Body:

Incontinentia pigmenti (IP; OMIM#308300) is an X-linked dominant disease, generally lethal in male, caused by mutations in *IKBKG/NEMO* gene, essential for NF-κB activation. The IP phenotype is characterized by typical skin lesions and by neuroectodermal defects that contribute to a wide variability of disease severity.

Here we report the case of IP female with neurological and ocular impairment, carrying the recurrent deletion in *IKBKG/NEMO* gene as a somatic mutation. Moreover, the patient showed also a de novo pathogenic variant c.1708_1709del, p.Ser570fs*27 in *MED13L* gene.

The IP locus analysis revealed the presence of de novo *NEMO*del4-10 deletion and excluded the presence of the risk alleles for IP (*MER67*dup, *NEMO*del). Somatic mosaicism was supported by quantitative analysis of the ratio of allele mutated versus wild-type allele in genomic DNA from blood. Consistent with somatic mosaicism, the sample of patient had lower ratios (30 %) of mutant versus wild-type allele compared to the fully heterozygote IP female control.

It has been previously reported that postzygotic/somatic mosaicism for the *IKBKG/NEMO* LoF mutation causes IP in male.

We here report that low level mosaicism for a post-zygotic/somatic *IKBKG/NEMO* mutation causes IP also in female. The somatic *IKBKG/NEMO* mutation in general is not detectable in male peripheral blood while is detectable in female most likely due to X-inactivation. The contribution of constitutive *MED13L* mutation in the female patient (NM_015335.5: c.1708_1709del) will be discussed.

Mendelian Phenotypes Posters - Thursday

PB1888. Integrating Genomic and Phenotypic Analyses of Autonomic Nervous System Dysfunction in a Rare Neurological Disease Cohort

Authors:

E. Rivera-Munoz¹, A. Jolly², H. Du¹, J. R. Lupski¹, J. E. Posey¹, Z. H. Coban-Akdemir³; ¹Baylor Coll. of Med., Houston, TX, ²1 Baylor Plaza, Houston, TX, ³Baylor Coll. of Med.; Human Genetics Ctr., Houston, TX

Abstract Body:

Autonomic nervous system dysfunction or dysautonomia represents a broad group of disorders that includes many diverse phenotypes affecting single - or multiple- organ systems within an affected individual. Many individuals with dysautonomia have neurocardiac involvement characterized by postural orthostatic tachycardia syndrome (POTS), orthostatic hypotension, and vasovagal syncope. Nevertheless, additional features have been described including abnormal temperature regulation, chronic fatigue, gastrointestinal issues, and dysregulation of other involuntary processes regulated by the parasympathetic and sympathetic nervous systems. Autonomic perturbations have multiple known etiologies ranging from aberrant development, adult stem cell maintenance, defects in ion channels, metabolism of neurotransmitters, inflammation, and neurodegeneration. Despite these mechanistic insights, many families with dysautonomia remain without a molecular diagnosis and limited efficacious treatments. Through the Baylor Hopkins Center for Mendelian Genomics (BHCMG) and Baylor College of Medicine Genomic Research to Elucidate the Genetics of Rare (BCM-GREGoR) databases, we have identified 38 unrelated probands and families with exome sequencing (ES) data across a range of clinical phenotypes that are indicative of autonomic nervous system dysfunction. This cohort was reported to have individuals with different organ systems involved including: 60.5% (23/38) cardiovascular, 57.9% (22/38) neurological, 39.5% (15/38) musculoskeletal, 36.8% (14/38) pulmonary, and 26.3% (10/38) gastrointestinal abnormalities. Using a quantitative phenotypic similarity score to compare patients' phenotypes, we have identified subtle phenotypic differences among our cohort, that may potentially inform distinct subgroups of dysautonomia in the context of rare disease. By leveraging existing genomic knowledge sources, we have identified putative candidate, rare variants (MAF \leq 0.1%) predicted to be loss of function in genes with relevant phenotypes in model organisms and the literature (*GXYLT1*, *ATP7B*, *CENPJ*, and *ARG1*). These molecular findings in rare disease may further and more broadly inform autonomic nervous system dysfunction and its clinical presentations beyond neurocardiogenic manifestations. We propose that parallel integration of gene- and phenotype-first analyses offers a productive method for investigation of human disease traits displaying substantial genotypic and phenotypic heterogeneity.

Mendelian Phenotypes Posters - Wednesday

PB1889. Intraflagellar transport protein RABL5/IFT22 is required for normal visual function and retinal photoreceptor survival

Authors:

Y. Yang, Z. Xianjun; UESTC, Chengdu, China

Abstract Body:

Purpose

The outer segment (OS) and inner segment (IS) of retinal photoreceptors are connected by a highly modified connecting cilium, which serves as a conduit to mediate the transport of photopigment molecules to the OS. RABL5/IFT22, a small GTPase-like component of the intraflagellar transport (IFT) complex B, which has been implicated in cilium construction and serves important functions in ciliary cargo selection/transport by recruits the BBSome to basal body. Considering the vital roles of ciliary transport in photoreceptors, we investigated the effect of genetic deletion of *Ift22* on the visual function and pathological changes in mouse.

Methods

We first constructed *Ift22* knockout (KO) mouse model. Electroretinograms (ERGs) were recorded to evaluate the visual function. H&E staining was carried out to verify the loss of rod cells in outer nuclear layer (ONL). To specific investigate the roles of in rod and cone photoreceptors, the conditional *Ift22*-knockout mouse models using the Rho-Cre and HRGP-Cre lines were also generated. Immunofluorescent labeling of several antibodies was performed to reveal detailed pathological alterations caused by *Ift22* deficiency. Transmission electron microscopy (TEM) was also employed to explore the ultrastructural pathological changes. Finally, label-free quantitative proteomics were used to investigate the underlying functional mechanisms behind the retinal pathologies.

Results

IFT22 was abundantly expressed in the retinal photoreceptor layer. ERG recordings manifested a reduction in the amplitudes of a- and b-waves in KO mice. Accordingly, H&E staining revealed a gradual thinning of ONL and shortening of OS in KO mice. Specific depletion of *Ift22* in rod cells causes extensive rod degeneration accompanied with ectopic accumulation of multiple OS proteins in the IS. Likewise, *Ift22* deficiency in cone cells led to the mislocalization of cone opsin and progressive death of cone cells. Mechanistically, label-free quantitative proteomics illustrated that IFT22 depletion resulted in reduced content of multiple phototransduction- and cilium-associated proteins. Western blot further verified that the levels of phototransduction signaling proteins CNGA1, GNAT1 and BBSome components BBS2, BBS4, BBS5 were reduced in *Ift22*-deficient retinas, although both immunostaining and TEM revealed normal cilium organization.

Conclusions

Our data demonstrate that IFT22 plays an important role in maintaining ciliary transport in retinal photoreceptors, and is essential for photoreceptor function and survival.

Mendelian Phenotypes Posters - Thursday

PB1890. Investigating SLC6A1 targeted therapeutics in *Drosophila*.

Authors:

K. Jay¹, K. Pham¹, J. Andrews¹, S. Jangam¹, N. Gogate¹, R. German¹, F. Alavi Naini¹, J. Constantino², H. Dierick¹, H. Bellen¹, M. Wangler¹; ¹Baylor Coll. of Med., Houston, TX, ²Washington Univ., St. Louis, MO

Abstract Body:

Our understanding of the clinical phenotype associated with *SLC6A1* (*Solute Carrier 6, Member 1*) related disorders is rapidly increasing. Patients experience genetic pleiotropy and variable expressivity of symptoms including developmental delay, epilepsy, intellectual disability, motor dysfunction, and autism spectrum disorder. *SLC6A1* encodes the gamma aminobutyric acid (GABA) transporter type 1 (GAT1) protein and contributes to maintaining homeostasis of excitatory and inhibitory neurotransmitters. GAT1 is responsible for the reuptake of the inhibitory neurotransmitter GABA at the synapse which is essential to protect from seizure activity. The most prevalent phenotype of *SLC6A1* patients is pediatric onset absence seizures that progress with age to myoclonic atonic seizures. Seizures are managed with general anti-epileptics, but many patients are not responsive to standard therapeutics. Furthermore, GAT1 is a potential therapeutic target. For this study, we recruited 11 patients through the Brain Gene Registry with *SLC6A1* related disorders. We completed a genotype-phenotype correlative study and identified patients who were responsive and non-responsive to current therapeutics. The average number of therapeutics attempted was 4.6 per patient and several patients experienced exacerbated symptoms or adverse side effects before identifying the most effective treatment. Stronger alleles that impede protein function should result in increased GABA in the presynaptic cleft and increased downstream GABA receptor signaling. I hypothesize increasing activity of the functional GAT1 protein will improve synaptic homeostasis and suppress seizure activity in *SLC6A1* patients. We have characterized loss-of-function phenotypes associated with loss of the *Drosophila* ortholog *Gat*. We found that *Gat* mutant flies are bang sensitive analogous to seizure activity in humans, exhibit motor dysfunction, and have a reduced lifespan. We are currently developing patient specific custom constructs to insert into the *Drosophila* genome to express our variants of interest. Our short-term goal is to evaluate the ability of the variants to rescue loss-of-function phenotypes. Our long-term goal is to identify novel *SLC6A1* modulators and to apply these custom models to predict patient response to therapeutics. Identifying therapeutics that will be most effective in suppressing seizure phenotypes by genotype will eliminate several rounds of trial and error for the patient and result in improved outcome for those living with *SLC6A1* related disorders.

Mendelian Phenotypes Posters - Wednesday

PB1891. Investigating the role of seryl-tRNA synthetase in mitochondrial biology and human recessive disease

Authors:

C. Del Greco, J. Kitzman, A. Antonellis; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Aminoacyl-tRNA synthetases (ARSs) are essential, ubiquitously expressed enzymes that charge tRNA molecules with amino acids in the cytoplasm and mitochondria. All 37 human ARSs are nuclear-encoded and 17 charge tRNA exclusively in the mitochondria to allow for the translation of the thirteen mitochondrial-encoded proteins, which are involved in oxidative phosphorylation. All 17 mitochondrial ARSs have been implicated in human recessive diseases with a broad range of clinical phenotypes that often affect tissues with high energy demands. Mitochondrial seryl-tRNA synthetase (*SARS2*) has been implicated in disease phenotypes including progressive spastic paresis and HUPRA (hyperuricemia, pulmonary hypertension, renal failure in infancy, and alkalosis) syndrome; however, there is no explanation for this clinical variability. Additionally, the number of disease-causing variants is small ($n=6$) and the spectrum of potentially pathogenic *SARS2* variants is incomplete. To address these issues, we will perform a massively parallel assessment of all possible loss-of-function variants in the *SARS2* open reading frame. We generated a Hap1 cell line that contains: (1) a randomly integrated, doxycycline-inducible wild-type copy of *SARS2*; and (2) a 2kb deletion at *SARS2* that ablates gene function. These cells will be transduced with lentiviral cDNA libraries containing all possible *SARS2* variants. Since *SARS2* is an essential enzyme, we will sequence the transduced cDNA from cells in the presence of doxycycline (during expression of the inducible wild-type copy of *SARS2*) and in the absence of doxycycline (in which the variant *SARS2* from the cDNA is the only copy of *SARS2* being expressed). This will allow us to catalog loss-of-function variants by identifying variants that are reduced in the cell population. This comprehensive variant-to-function map will expand the spectrum of potentially pathogenic *SARS2* variants to aid in future diagnosis, and will identify amino acid residues that are important for *SARS2* function.

Mendelian Phenotypes Posters - Thursday

PB1892*. Investigating tissue specific defects in mitochondrial bioenergetics and quality control: new implications for cellular pathogenesis and therapeutic targeting in Barth syndrome

Authors:

O. Sniezek¹, N. Senoo², G. Butschek³, A. Anzmann¹, S. Claypool², H. Vernon¹; ¹Johns Hopkins Sch. of Med., Dept. of Genetic Med., Baltimore, MD, ²Johns Hopkins Sch. of Med., Dept. of Physiology, Baltimore, MD, ³Johns Hopkins Sch. of Publ. Hlth., Dept. of Molecular Microbiol. and Immunology, Baltimore, MD

Abstract Body:

Barth Syndrome (BTHS) is a rare, X-linked disorder caused by defects in the gene TFAZZIN, resulting in defective cardiolipin remodeling. BTHS is clinically characterized by cardiomyopathy, neutropenia, growth defects, and myopathy and in contrast to many mitochondrial disorders, there are limited central nervous system effects. There is a high risk of morbidity and mortality, with no approved disease-specific treatments. Our lab previously identified defects in Complex I of the mitochondrial respiratory chain, and dysregulation of PARL and PGAM5 mediated mitochondrial quality control in a TFAZZIN-KO HEK293 model. However, it is not known why the effects of TFAZZIN, which is ubiquitously expressed, have a tissue-specific pattern of disease expressivity. To understand tissue specific Complex I and mitochondrial quality control defects in BTHS we differentiated TFAZZIN-KO iPSCs into cardiomyocytes (an affected tissue in BTHS) and neurons (a spared tissue in BTHS). To broadly characterize each cell type in an unbiased manner, we first performed RNAseq. This analysis identified cell-type specific dysregulation in pathways related to mitochondrial quality control and metabolism including Wnt signaling, apoptosis, autophagy, and cardiomyocyte-specific dysregulation in glucose metabolism. To illustrate the effects of defective mitochondrial quality control, we next characterized the morphology of each cell type with confocal microscopy. TFAZZIN-KO cardiomyocytes revealed significant morphological abnormalities compared to wild type including decreases in mitochondrial number, area, perimeter, and branching. Importantly, these mitochondrial abnormalities were not found in TFAZZIN-KO undifferentiated cells or neural progenitors. We next investigated bioenergetic phenotypes in each cell type. Oxygen consumption studies showed impaired maximal respiratory capacity in TFAZZIN-KO cardiomyocytes and neurons but less so in undifferentiated cells. The respiratory capacity of wild-type cardiomyocytes was improved by fatty acid supplementation mimicking the cardiac physiological environment; however, that of TFAZZIN-KO cardiomyocytes was further severely impaired. We next plan to study substrate utilization in TFAZZIN-KO neural progenitors to determine if these effects are recapitulated. To understand new directions for potential therapeutic intervention in BTHS, we are currently targeting specific fatty acid pools and early steps in cardiolipin metabolism in cardiomyocytes and neurons with the goal of remediating cellular lipids, mitochondrial morphological defects, and oxygen consumption abnormalities.

Mendelian Phenotypes Posters - Wednesday

PB1893. Investigating variants of uncertain significance in orofacial cleft trios.

Authors:

K. Diaz Perez¹, S. Curtis¹, A. Sanchis-Juan², T. Head¹, S. Ho¹, B. Carter³, T. McHenry⁴, M. Bishop¹, L. Valencia-Ramirez⁵, C. Restrepo⁵, L. Moreno⁶, G. Wehby⁶, T. Beaty⁷, J. Murray⁸, E. Feingold⁴, M. Marazita⁹, D. Cutler¹, M. Epstein¹, H. Brand¹⁰, E. Leslie¹; ¹Emory Univ., Atlanta, GA, ²Broad Inst. of Harvard and MIT, Somerville, MA, ³Agnes Scott Coll., Decatur, GA, ⁴Univ of Pittsburgh, Pittsburgh, PA, ⁵Fundacion Clinica Noel, Medellin, Colombia, ⁶Univ Iowa, Iowa City, IA, ⁷Johns Hopkins Univ, Sch PubHlth, Cockeysville, MD, ⁸U of Iowa, Iowa City, IA, ⁹Univ Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA, ¹⁰Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

Orofacial clefts (OFCs) are common craniofacial birth defects, including cleft lip (CL), cleft lip and palate (CLP), and cleft palate (CP). OFCs are etiologically heterogeneous, and many of the genetic risk factors are not well understood. Recently, we investigated likely pathogenic variants in 503 genes associated with primarily Mendelian OFCs using whole-genome sequencing of 841 OFC cases and 294 controls. We found likely pathogenic variants in 9.16% of cases and only 1.36% of controls ($p < 0.0001$). These likely pathogenic variants in cases were distributed across 40 genes, underscoring the substantial genetic heterogeneity in OFCs. However, most reviewed variants (59.5%) were classified as “variants of uncertain significance” (VUSs) and were more frequent in cases than controls ($p = 0.002$) even though cases and controls had similar distributions of VUSs per person. Overall, VUSs were overwhelmingly missense variants, highlighting the difficulty of assessing pathogenicity for missense variation. Furthermore, we noted a high number of VUSs among cases were in genes highly constrained to missense variation (Genome Aggregation Database (gnomAD) missense Z score ≥ 3.1) ($p = 0.10$). To assess whether individual genes exist with a significant excess of VUSs in cases, we performed gene-based association tests using the Optimal Sequence Kernel Association Test. Overall, we found no single gene with a significant excess of VUSs in cases that could explain the genome-wide burden of VUSs in cases versus controls. However, we found that VUSs in *PRICKLE1* were nominally associated with the risk of OFCs ($p = 0.04$). All other nominally associated genes (e.g., *PHYH*, *SOX9*) showed a higher frequency of VUSs in controls (*PHYH*, $p = 0.03$; *SOX9*, $p = 0.03$). These results remained consistent after removing 81 “solved” individuals with likely pathogenic variants. Stratifying by cleft types identified additional nominal associations, including VUS variants in *GLI3* with CL ($p = 0.009$). *GLI3* VUSs were also identified with individuals with CLP but were not significantly associated. Overall, these data suggest a collective effect of risk variants in various genes among cases, including variants that are not easily classified as pathogenic, and variants in these genes should be subjected to functional validation to determine their true effect. Cumulatively, these results further highlight the etiological heterogeneity of OFCs and provide insight into the role of VUSs in candidate genes for risk to OFCs.

Mendelian Phenotypes Posters - Thursday

PB1894. Isolated acanthosis nigricans in a Mexican girl with a nonsense variant in *FGFR3* gene.

Authors:

M. Abreu González¹, R. Santillán-Martínez¹, L. Hernández Ancheyta¹, Y. Centeno Navarrete², S. Contreras-Capetillo³; ¹Genos Medica, Mexico City, Mexico, ²Hosp. Gen. Agustín O'Horán, Secretaría de Salud de Yucatán,, Yucatán, Mexico, Mexico, ³Centro de Investigaciones Regionales Dr. Hideyo Noguchi, Univ. Autónoma de Yucatán,, Yucatán, Mexico

Abstract Body:

Acanthosis nigricans (AN) is characterized by velvety thickening and hyperpigmented hyperkeratosis of the skin, usually affecting the folds of the neck, armpits, forehead and groin. Syndromic AN has been described with gain of function variants in *FGFR3* (MIM *134934) as part of severe achondroplasia with developmental delay, Crouzon syndrome and hypochondroplasia. The constitutive activation of the *FGFR3*, which are only caused by missense mutations, led to inhibition of chondrocytes growth that produces skeletal anomalies. Meanwhile, the isolated familial acanthosis nigricans (IAN) is originated by mutations with a weak dominant activity in the receptor. Recently, were described two cases of IAN with a nonsense variant in *FGFR3* [Tahara, et al 2021].

In this study we present a female Mexican patient, the only affected in the family, with IAN. On examination at 10 years old she had height (p.75-90) and weight (p.25). The skin was thick and hyperpigmented specially in the jaw, neck and armpits. No skeletal anomalies or neurological defects were detected. Insulin, HOMA, ACTH and 17-hydroxiprogesterone were within normal limits. The skin biopsy was reported as papillomatosis, hyperkeratotic and AN.

A next generation sequencing targeted panel was ordered (Sophia Genetics, CES V2 Panel). The bioinformatics pipeline analysis of the 29 genes associated with AN identified the c.2302G>T (p.Glu768*) pathogenic variant in *FGFR3*, that is predicted to introduce a premature termination codon in the last exon, and it is been suggested that generates a weak constitutive active receptor that correlates with the clinical phenotype of our patient.

Here we described the third case of IAN, the mildest form of the *FGFR3*-related disorders, due to a c.2302G>T (p.Glu768*) in *FGFR3* that expands the phenotype of the type of mutations that could be found in this gene. Patients with IAN, regardless of the presence skeletal alterations, should have a *FGFR3* analysis performed [Hirai, et al 2017].

Mendelian Phenotypes Posters - Wednesday

PB1895. *ITGB8* is a candidate disease gene for autosomal dominant and recessive trait forms of muscular dystrophy and neurological disease

Authors:

S. Barish¹, H. Du², Y. Li¹, J. Fatih¹, D. Calame¹, T. Mitani³, D. Pehlivan¹, J. Lupski¹, J. Posey¹; ¹Baylor Coll. of Med., Houston, TX, ²BCM, Houston, TX, ³Kanagawa Children's Med. Cente, Yokohama, Kanagawa, Japan

Abstract Body:

Integrins are a family of cell adhesion molecules that facilitate cell-cell and cell-extracellular matrix (ECM) contact. Integrins form heterodimers of an alpha and a beta subunit. There are 26 integrins in the human genome (18 alpha, 8 betas), nine of which have been associated with Mendelian conditions. Integrins have primarily associated with skin diseases such as epidermolysis bullosa (MIM #226730) and none have been associated with neurological disease to date. In this study, we identified three unrelated probands with putatively pathogenic variants in *ITGB8*. Two of the probands carry a homozygous missense variant allele, NM_002214.3 c.1846G>A (p.Asp616Asn), that is ultrarare (minor allele frequency < 0.1%) in both control and mixed rare disease populations. These probands present with neurological phenotypes including developmental delay, intellectual disability, microcephaly, and dysmorphic features. The third proband carries a heterozygous putative loss-of-function variant, NM_002214.3 c.1635C>A (p.Y545*) in exon 10/14 that is absent from control populations. This proband presents with facioscapulohumeral muscular dystrophy (FSHD) but analysis of *SMCHD1* and *DNMT3B*, as well as D4Z4 repeat number, did not yield any potential damaging variants to support an FSHD type 1 or type 2 molecular diagnosis. Variants in *ITGB8* were verified in each family via Sanger dideoxy sequencing and segregated in accordance with Mendelian expectations for an autosomal recessive rare disease trait in the first two families. Parental samples for the third family were not available, but the variant was not detected in either the unaffected sibling or child of the proband. *In silico* tools predict *ITGB8* to be intolerant to loss-of-function (pLOEUF = 0.26) and the identified variants are predicted to damage protein function. Together, our data suggest that variants in *ITGB8* are associated with dominant and recessive forms of neurogenetic rare disease trait.

Mendelian Phenotypes Posters - Thursday

PB1896. Joint analysis of de novo mutations in multiple traits and gene-expression data improves statistical power for the prioritization of genes associated with disease

Authors:

T-H. Nguyen, K. Kendler, B. Riley, S-A. Bacanu; Psychiatry Dept., Virginia Commonwealth Univ., Richmond, VA

Abstract Body:

Motivation: De novo mutations from parent-offspring trio data have helped identify risk genes for neuropsychiatric and other disorders. However, for many disorders, only a small number of genes have been discovered. Fortunately, identified genes of multiple disorders are expressed mostly in specific tissues or cell types. Therefore, integrating the gene expression information of the enriched tissues or cell types can increase statistical power. We previously showed that mTADA (multi-trait transmission and de novo association test), which performs a joint analysis of de novo mutations from multiple traits, helps identify additional risk genes. We also showed that integrating gene-sets from gene expression data into this method could improve statistical power. However, the previous approach is not able to jointly analyze multiple tissues/cell types. Here, we propose a flexible approach which integrates both 1) de novo mutations from two traits and 2) the gene-expression information from multiple tissues/cell-types for the prioritization of risk genes.

Results: We tested the new method on simulated data. The new approach performs better than mTADA in the identification of risk genes, especially when enriched tissues are used. We applied the new method to a de novo mutation dataset of 10,012 parent-offspring trios from five disorders. We identified more risk genes than its previous version. For example, by using a threshold of $FDR < 0.05$ in the joint analysis of autism spectrum disorder (ASD) and epileptic encephalopathies (EE): 51 genes were identified by the new approach while only 46 genes were reported by mTADA for ASD.

Availability: The new method is incorporated into our active mTADA R package, which contains various functions for parent-offspring trio analyses.

Mendelian Phenotypes Posters - Wednesday

PB1897. *KIF1A* novel variant p.(Ser887Profs*64) exhibits clinical heterogeneity in a Pakistani family individuals with HSANIIC

Authors:

M. Azam¹, S. Ghafoor¹, M. Rafiq¹, S. Abbas Shah¹, M. Ajmal¹, R. Qamar²; ¹COMSATS Univ. Islamabad, Islamabad, Pakistan, ²Sci. and Technology Sector, ICESCO, Rabat, Morocco, Rabat, Morocco

Abstract Body:

Hereditary sensory and autonomic neuropathies (HSANs) are rare heterogeneous group of neurological disorders caused by peripheral nerve deterioration. The HSANs sub-clinical classes have clinical and genetic overlap which often lead to misdiagnosis. In the present study a Pakistani family with five affected members suffering from neuropathy were genetically analyzed to identify the disease causative element in the family. Genome wide high-density single nucleotide polymorphism (SNP) microarray analysis of DNA samples of five affected and nine healthy members was carried out followed by whole exome sequencing of the proband and an affected sibling. Shared homozygous regions in all affected were identified through homozygosity mapping approach. The largest homozygous region of 14.1Mb shared by the five affected members of the family was identified on chromosome 2. Subsequent exome sequencing identified a novel single nucleotide deletion c.2658del; p.(Ser887Profs*64) in *KIF1A*. Segregation analysis revealed that this mutation was homozygous in all five affected individuals of the family with severe clinical manifestation, while members of the family that were heterozygous carriers shared only abnormal skin features (scaly skin) with the homozygous affected members. A novel frameshift mutation p.(Ser887Profs*64) in *KIF1A* is the potential cause of severe HSANIIC in a Pakistani family along with incomplete penetrance in mutation carriers. We demonstrate that using a combination of different techniques strengthens the gene finding approach along with exact clinical sub categorization in HSAN group of inherited diseases.

Mendelian Phenotypes Posters - Thursday

PB1898. Large-scale genomic analyses of 150 consanguineous kindreds from the Middle East and North Africa identify novel neurodevelopmental disease mechanisms

Authors:

R. Duan¹, D. Marafi^{2,1}, D. Calame¹, I. Herman¹, T. Mitani¹, M. Eldomery³, A. Saad^{1,4}, J. Fatih¹, H. Du¹, C. Grochowski¹, Z. Dardas¹, A. Jolly¹, M. Dawood¹, M. Zaki⁴, G. Abdel-Salam⁴, N. Saadi⁵, h. elbendary⁴, H. Moussa⁶, K. Rafat⁴, I. Hegazy⁴, M. Issa⁴, R. Alshamari⁷, M. Ashkanani⁸, M. Arabi⁹, L. Bastaki⁷, M. Zakariah¹⁰, M. Djuric¹¹, D. Sadik⁷, H. Alrohaif⁷, S. Gouda⁷, E. El-Anany¹⁰, BCM-GREGoR Consortium, S. Jhangiani¹, Z. Coban-Akdemir^{1,12}, C. Carvalho¹³, R. Gibbs¹, D. Pehlivan¹, J. Posey¹, J. Lupski¹; ¹Baylor Coll. of Med., Houston, TX, ²Kuwait Univ., Kuwait City, Kuwait, ³Indiana Univ Sch. of Med., Indianapolis, IN, ⁴Natl. Res. Ctr., Cairo, Egypt, ⁵Univ. of Baghdad, Baghdad, Iraq, ⁶Specialties Hosp. Ctr., Nouakchott, Mauritania, ⁷Kuwait Med. Genetics Ctr., Kuwait City, Kuwait, ⁸Amiri Hosp., Ministry of Hlth., Kuwait City, Kuwait, ⁹Ibn Sina Hosp., Kuwait City, Kuwait, ¹⁰Adan Hosp., Ministry of Hlth., Kuwait City, Kuwait, ¹¹Farwaniya Hosp., Ministry of Hlth., Kuwait City, Kuwait, ¹²The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ¹³Pacific Northwest Res. Inst., Seattle, WA

Abstract Body:

Neurodevelopmental disorders (NDDs) affect >3% of newborns worldwide due to nervous system dysfunction. Driven by the Clan Genomics hypothesis, family-based genomics, and rare variant analyses, the discovery of novel NDD-traits has reached an ‘exponential phase.’ Nevertheless, the mutational mechanisms and perturbations of biological balance underlying this clinically heterogeneous class of disorders remain largely unknown. Here, we performed genomic analyses of a cohort of 150 mostly consanguineous families with NDDs from the Middle East and North Africa (MENA) region. Preliminary data analyses reveal a molecular solved rate of 103/150 (69%) families via exome sequencing and deep clinical phenotyping, including 87/150 (58%) solved by known disease genes/loci and 16/150 (11%) cases solved by proposed candidate genes. Sixty-five distinct genes/loci linked to known NDD traits, including 72 (94%) SNVs/indels and 5 (6%) CNVs. Among 87 families solved with known disease genes/loci, recessive NDD traits driven by autozygosity were found in 73/87 (84%) of families, while 10/87 (12%) families revealed *de novo* dominant NDD traits despite reported parental consanguinity; inherited X-linked hemizygous and compound heterozygous alleles were also observed in 2/87 (2%) families, respectively. The average total absence of heterozygosity, a surrogate measure of runs of homozygosity (ROH) calculated from 30 unrelated families in this cohort, was 313 Mb (range: 118 to 647Mb), and causal autozygous variants fall within ROH blocks in an average of 12.04 Mb (range: 1.2 to 60.5 Mb). For families with a single molecular diagnosis, the total ROH in the MENA cohort is significantly higher than in the previously published TBM2-NDD and Arthrogryposis cohorts (one-way ANOVA, $p < 0.0001$), however, there are no statistical differences in the total ROH among the three cohorts for the families with multi-locus pathogenic variants. Moreover, identifications of candidate genes led to discoveries of multiple novel neurobiological mechanisms such as i) biallelic loss-of-function (LoF) of *LGI3* causing mislocalization of Kv1 channel complexes in juxtaparanode of myelinated peripheral axons, resulting in a novel peripheral nerve hyperexcitability syndrome; ii) LoF mutations in *SLC38A3*, a sodium-coupled neutral amino acid transporter required for the glutamate/GABA homeostasis, cause a novel AR-type of developmental epileptic encephalopathy. Together, family-based genomic analysis on a diverse population accelerates the discoveries of novel NDD genes/loci, serving as an “entry-point” to elucidate unknown neurobiology of disease.

Mendelian Phenotypes Posters - Wednesday

PB1899. Large-scale zebrafish-based functional analysis of genes associated with neurodevelopmental disorders

Authors:

S. Thyme, C. Conklin, W. Gannaway, E. Torija, V. Martina, A. Ginsparg, C. Calhoun, M. Vivian, M. Capps, Thyme Lab Undergraduate Assistants (>5); Univ. of Alabama at Birmingham Med. Sch., Birmingham, AL

Abstract Body:

The goal of the Thyme lab is to uncover the molecular basis for complex neurodevelopmental disorders and discover drug treatments. We use the larval zebrafish model, taking advantage of new high-resolution and large-scale approaches for neural phenotyping and genetic analysis, which we pioneered for analyzing mutants generated with CRISPR/Cas9. Previously, we assessed whole-brain activity, brain morphology, and behavior of zebrafish loss-of-function mutants for orthologs of 132 human schizophrenia-associated genes. This work set the stage for in-depth molecular studies of several mutants with compelling phenotypes and additional screens of genes involved in other disorders. One example of an in-depth study is of the autism- and schizophrenia- associated transcription factor ZNF536, which we have found regulates a completely unstudied putative protease through RNA-seq and genetic interaction data. We have ongoing, unpublished screens for other genes linked to schizophrenia by exome sequencing (7 genes), rare variants linked to childhood-onset schizophrenia (8 genes and 4 patient mutations), and genes directly linked to autism (17 genes and 3 patient mutations) and indirectly through the implicated program of developmentally regulated microexon splicing (43 genes). We have collected whole-brain activity maps and behavioral data for more than 50 mutant lines from these screens. Comparing the phenotypes of truncating mutations to patient-specific mutations has provided insight into the function of these mutations. In parallel to our zebrafish work, we have built a new computational method for in silico drug discovery. This method integrates a knowledge-based approach with the physical energy potential of the Rosetta macromolecular modeling program. Testing several hundred predicted compounds in larval zebrafish is feasible, and many proteins are highly conserved, making it an ideal model for compound screening and testing new tools and methods. Our proposed pipeline of integrating computation, whole-organism screening, and in-depth mechanistic studies promises to yield potentially therapeutic molecules that impact diverse mechanisms underlying neural development and function.

Mendelian Phenotypes Posters - Thursday

PB1900. Leveraging orthogonal sequencing and optical mapping technologies for the precision diagnosis of neurodevelopmental disorders in a Middle Eastern family based cohort.

Authors:

Z. Siddig¹, D. Hammadi¹, A. Nour¹, S. Hassan¹, M. Ghorbani¹, H. Naeem¹, I. Diboun¹, S. Moosa¹, R. Abouelhasan¹, M. Tauro¹, S. Nassar¹, R. Farhad¹, R. Razali², I. Alazwani¹, M. Hashmi¹, R. Matthew¹, L. Liu¹, G. Wang¹, S. Poolat¹, A. El Khouly¹, S. Tomei¹, S. Lorenz¹, D. Love¹, B. Seim¹, K. Fakhro^{3,4}, A. Satti¹, R. Benini¹, Y. Mokrab^{1,4,5}; ¹Sidra Med., Doha, Qatar, ²Dept. of BioMed. Sci., Coll. of Hlth.Sci., Qatar Univ., Doha, Qatar, ³Sidra Med., Doha, Qatar, Qatar, ⁴Dept. of Genetic Med., Weill Cornell Med., New York, NY, ⁵Coll. of Hlth.and Life Sci., Hamad Bin Khalifa Univ., Doha, Qatar

Abstract Body:

Establishing genotype-phenotype correlations for neurodevelopmental disorders (NDDs) is a substantial challenge, especially amongst Middle Eastern populations enriched for these conditions due to high level of consanguinity. We implemented a platform integrating data from short and long read WGS, optical mapping and EHRs to diagnose and characterize unresolved pediatric cases with neurodevelopmental disorders in Qatar. In a pilot project, we recruited a cohort with intractable epilepsy or developmental delay with no genetic diagnosis, consisting of 102 families of patients. For these we generated Illumina WGS and PacBio long-read sequencing, from which we obtained high quality SNVs and Indels as well as SV and CNVs respectively. EHRs were collected and mapped to HPO from which the terms global developmental delay, delayed speech, and seizures were found to be most common. Family segregation genomic analysis was done using Congenica© using gene panels from the G2P consortium. Variants were annotated with allele frequency from Qatar Genome Project and other public datasets, as well as pathogenicity predictors including Exomiser and CADD. Candidate disease variants were then prioritized following ACMG guidelines. Using this approach, we analyzed a set of 14 families leading to the shortlisting of 23 variants on 22 genes. Nine genes were known to be implicated in NDDs. Amongst the variants, 9 variants were previously reported including 3 with phenotypes reported as matching the patients'. We report a *de novo* missense variant (chr15:44881558C>A) in *SPG11* in a female patient with spastic tetraplegia, cerebral palsy, global developmental delay, and seizures. *SPG11* is strongly linked to spastic paraplegia in the Decipher Developmental Disorders Genes database. Furthermore, we identified 9 novel variants on known pathogenic genes potential link to patient phenotypes. For instance, we identified a novel *de novo* pathogenic splice donor variant (chr19:13136367G>A) in *NFIX* in a male patient who exhibits symptoms of global developmental delay and tall stature. Known syndrome caused by *NFIX* mutations is Soto's syndrome, characterized by developmental delay and overgrowth (PMID: 25736188).Based on PacBio long read data we identified CNVs across genes and genomic regions known to be linked to NDDs as well as novel regions. For example, we identified an 80.9kbp homozygous deletion at 21q22.12 in a patient who presented with microcephaly and global developmental delay. As we continue to analyze this patient cohort, the results will have significance impact on to the precision diagnosis and understanding of developmental disorders in the Middle East and worldwide.

Mendelian Phenotypes Posters - Wednesday

PB1901*. LIM Homeobox 1 gene variants contribute to Mayer-Rokitansky-Küster-Hauser syndrome

Authors:

D. Kira¹, J. Butler², L. P. Chorich¹, Z. A. Hawkins¹, J. A. Phillips III³, L. C. Layman⁴; ¹Augusta Univ., Augusta, GA, ²Morehouse Sch. of Med., Atlanta, GA, ³Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁴Med Coll GA at Augusta Univ., Augusta, GA

Abstract Body:

Introduction: Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome consists of congenital absence of the uterus and vagina and affects 1/4,500 women. Affected individuals present with normal breast development, primary amenorrhea, and a 46,XX karyotype. The Mullerian aplasia may be isolated (type I) or associated with renal, skeletal, cardiac, and auditory anomalies (type II). The etiology of MRKH largely remains unknown, but multiple reports of familial clustering suggest possible genetic causes. Copy number variants (CNVs) have been reported, including a recurring 1.2 Mb chromosome 17q12 deletion. Interestingly, the LIM Homeobox 1 gene (*LHX1*) locus resides on 17q12 and it encodes the LHX1 transcription factor containing two LIM Zinc finger motifs and one DNA binding domain. LHX1 is important in renal and uterine development in rodents, and is essential for development of Müllerian duct epithelial progenitor cells. **Hypothesis:** Heterozygous deleterious *LHX1* variants including CNVs may cause MRKH. **Method:** gDNAs from 134 individuals with MRKH were analyzed by quantitative PCR (qPCR) to identify potential *LHX1* deletions since DNA sequencing failed to reveal single nucleotide variants. Heterozygous *LHX1* exon 4 deletions were found in gDNAs from 3/134 unrelated individuals. Total RNA was obtained from lymphoblasts from these 3 individuals with MRKH cultured with or without puromycin (a translation inhibitor that suppresses nonsense-mediated decay or NMD). RNAs were subjected to RT-PCR, followed by PCR for exons 3-5 of *LHX1*. PCR products were confirmed, and subjected to colony PCR and Sanger sequencing. **Results:** qPCR of gDNA demonstrated a de novo heterozygous 166 bp *LHX1* deletion (c.676_841del;pV226 Ifs*120) deleting the entire exon 4 sequence in all three patients, which was confirmed by RT-PCR. An intense exon 3-5 band was seen on gel electrophoresis by RT-PCR at differing concentrations of cDNA only with puromycin treatment. Without puromycin, no band was seen with 500ng or 1ug, but demonstrated a faint band with 1.5 and 2ug of cDNA. By colony PCR and sequencing, both the WT and variant were observed. Without puromycin most colonies were variant; with puromycin most colonies were WT. **Conclusion:** Our data indicate that gDNAs of 3/134 persons with MRKH have an *LHX1* exon 4 deletion, which is predicted to disrupt the 3' portion of the DNA binding domain. We hypothesize that these exonic deletions, extend into flanking introns to create, eliminate or activate cryptic splice sites that trigger NMD that leads to their degradation. Our findings provide additional evidence that LHX1 CNVs along with previously reported WNT4 and HNF1B variants contribute to MRKH.

Mendelian Phenotypes Posters - Thursday

PB1902. Lisch epithelial corneal dystrophy is caused by heterozygous loss-of-function variants in *MCOLN1*

Authors:

K. Patterson¹, J. X. Chong¹, D. D. Chung², W. Lisch³, H. Møller⁴, D. Garczarczyk-Asim⁵, E. Dreisler⁶, University of Washington Center for Rare Disease Research, A. J. Aldave⁷, M. Bamshad¹, A. R. Janecke⁸; ¹Univ. of Washington, Seattle, WA, ²Univ. of California, Los Angeles, Los Angeles, CA, ³Univ. Med. Ctr. of the Johannes Gutenberg- Univ. Mainz, Mainz, Germany, ⁴Dept. of Pediatric Ophthalmology, Aarhus Univ. Hosp., Aarhus, Denmark, ⁵Dept. of Pediatrics I, Med. Univ. of Innsbruck, Innsbruck, Austria, ⁶N.Jespersensvej 3, DK-2000 Copenhagen, Frederiksberg, Denmark, ⁷Jules Stein Eye Inst/UCLA, Los Angeles, CA, ⁸Innsbruck Med Univ, Innsbruck, Austria

Abstract Body:

Lisch epithelial corneal dystrophy (LECD) was delineated in 1992 in a large family with diffuse grayish epithelial opacities of the cornea in distinctive radial, feathery or club-shaped patterns. The opacities consisted of crowded, transparent microcysts in retroillumination of the cornea. Light microscopy shows vacuolization of the epithelial cells, particularly in outer layers of the cornea and electron microscopy discloses coalescent intracellular vacuolization of the corneal wing cells. In vivo confocal microscopy shows intraepithelial hyperreflective dystrophic areas containing hyporefective round structures, sharply demarcated from normal epithelial areas. LECD often causes no clinical complaints or causes blurred vision if the pupillary axis is involved. Using a combination of linkage analysis, exome, and genome sequencing in the original family, a second multiplex family, and a cohort of seven unrelated cases, we identified five heterozygous, rare (MAF < 6x10⁻⁵ in gnomAD) variants in *MCOLN1*, explaining 6/9 families. Three variants, c.514C>T (p.Arg172*), c.576C>A (p.Cys192*), and c.406-2A>G are predicted to be loss-of-function and two were missense, c.776T>C (p. Leu259Pro) in a proband and affected father and c.694A>C (p.Thr232Pro), in a proband.

In the homozygous and compound-heterozygous state, three of these variants have been previously reported to cause the rare lysosomal storage disorder mucopolipidosis IV (MLIV) and one meets ACMG Likely Pathogenic criteria. MLIV is characterized by severe psychomotor delay, progressive visual impairment, and achlorhydria, and corneal opacity is considered a hallmark finding. Staining of epithelial cells shows autofluorescent inclusions containing phospholipids, mucopolysaccharides, and gangliosides. *MCOLN1* encodes mucolipin-1, a member of the transient receptor potential (TRP) cation channel family, which is expressed in intracellular membranes. Most molecules trafficked through the lysosomal compartment are delayed in MLIV cells, either due to a primary transport defect, or secondarily because they are prevented from delivery by the excess storage. MLIV impairs vision by a combination of corneal clouding and retinal degeneration leading to severe visual impairment by adolescence.

Based on the overlapping clinical observations of diffuse epithelial inclusions and increased autofluorescence of such cells and material in LECD and MLIV, we conclude that at least some, if not most, carriers of *MCOLN1* loss-of-function mutations present with LECD.

Mendelian Phenotypes Posters - Wednesday

PB1903. Lissencephaly spectrum disorders: Clinical, radiological molecular results of 70 Egyptian families

Authors:

M. Zaki¹, M. Issa¹, H. elbendary¹, K. Rafat¹, J. Gleeson², M. Abdel-Hamid³; ¹Natl. Res. Ctr., Cairo, Egypt, ²Univ California, La Jolla, CA, ³Natl. Res. Ctr., Giza, Egypt

Abstract Body:

Lissencephaly (LIS) encompasses a spectrum of malformations of cortical development due to defective neuronal migration with diverse genetic contextual. Brain magnetic resonance allows the characterization of various anatomic and morphological abnormalities and further radiological-genotype correlations. We studied a cohort of 92 Egyptian patients derived from 70 families with LIS spectrum and microlissencephaly. Based on the radiological classification, target gene sequencing was done for LIS1, DCX, RELN, VLDLR, and LAMB1 genes. For patients with nonspecific neuroradiological patterns, whole-exome sequencing (WES) was performed. The causative mutations were identified in 60 families. LIS1 gene constitute 11.7 % of families followed by RELN and VLDLR, both 11.4% (8F), DCX (7.6%-5F) DYNC1H1(7.6%-5F), KATNB1 (7.6%- 5F), TUBA1A (5.7%-4F), APC2 (5.7%-4F), PIDD1 (4.2%-3F), LAMB1(2.8%-2F), TMTC3 (2.8% 2F), CDK5 (1.4%-1F), ACTG1(1.4%-1F) and MACF1 (1.4%-1F). Ten families (14.2%) remained unsolved. Because of the high consanguinity rate in our community, autosomal recessive linked LIS was prevalent. Novel mutations dominated our study and were significantly present in all families with RLN and VLDLR genes. Interestingly, the novel mutation identified in the MACF1 gene was homozygous. Herein, we present the phenotype and radiological findings of a large series of lissencephaly from the same ethnic group. We expand the mutational spectrum of lissencephaly and point to the predominance of autosomal recessive LIS.

Mendelian Phenotypes Posters - Thursday

PB1904. Local versus systemic control of bone and skeletal muscle mass by targeting components of the transforming growth factor- β signaling pathway using pharmacological and transcriptional approaches.

Authors:

H. Chandok¹, S-J. Lee¹, E. L. Germain-Lee², A. Lehar¹, Y. Liu¹, M. Michaud¹, J. George¹; ¹The Jackson Lab., Farmington, CT, ²Dept. of Pediatrics, Univ. of Connecticut Sch. of Med., Farmington, CT

Abstract Body:

Among the major health challenges for astronauts during prolonged space travel are loss of muscle mass and loss of bone mass. Here, we investigated the effects of targeting the signaling pathway mediated by the secreted signaling molecules, myostatin and activin A, in mice sent to the International Space Station. We show that targeting this signaling pathway has significant beneficial effects in protecting against both muscle and bone loss in microgravity, suggesting that this strategy may be effective in preventing or treating muscle and bone loss not only in astronauts on prolonged missions but also in people with disuse atrophy on Earth in individuals who are bedridden or wheelchair-bound from illness. Here, we used both genetic and pharmacological approaches to investigate the effect of targeting MSTN/activin A signaling in mice that were sent to the International Space Station. *Wild type* mice lost significant muscle and bone mass during the 33 d spent in microgravity. Muscle weights of *Mstn*^{-/-} mice, which are about twice those of *wild type* mice, were largely maintained during spaceflight. Systemic inhibition of MSTN/activin A signaling using a soluble form of the activin type IIB receptor (ACVR2B), which can bind each of these ligands, led to dramatic increases in both muscle and bone mass, with effects being comparable in ground and flight mice. To deepen the understanding, we performed local versus systemic control of bone and skeletal muscle mass by targeting components of the transforming growth factor- β signaling pathway. Using an allelic series corresponding to varying expression levels of endogenous *Fst*, we show that FST acts in an exquisitely dose-dependent manner to regulate both muscle mass and bone density. Moreover, by employing a genetic strategy to target *Fst* expression only in the posterior (caudal) region of the animal, we show that the effects of *Fst* loss are mostly restricted to the posterior region, implying that locally produced FST plays a much more important role than circulating FST with respect to regulation of muscle and bone. Finally, we show that targeting receptors for these ligands specifically in osteoblasts leads to dramatic increases in bone mass, with trabecular bone volume fraction being increased by 12- to 13-fold and bone mineral density being increased by 8- to 9-fold in humeri, femurs, and lumbar vertebrae. These findings demonstrate that bone, like muscle, has an enormous inherent capacity for growth that is normally kept in check by this signaling system and suggest that the extent to which this regulatory mechanism may be used throughout the body to regulate tissue mass may be more significant than previously appreciated.

Mendelian Phenotypes Posters - Wednesday

PB1905*. Long Read Sequencing and Expression Studies of *AHDC1* Deletions in Xia-Gibbs Syndrome Reveal a Novel Genetic Regulatory Mechanism

Authors:

V. Chander^{1,2}, M. Mahmoud^{1,2}, J. Hu^{1,2}, Z. Dardas², C. M. Grochowski², M. Dawood^{1,2}, M. Khayat^{1,2}, H. Li¹, S. Li¹, S. Jhangiani¹, V. Korchina¹, H. Shen¹, G. Weissenberger¹, Q. Meng¹, M-C. Gingras¹, D. M. Muzny¹, H. Doddapaneni¹, J. E. Posey², J. R. Lupski^{1,2,3,4}, A. Sabo¹, D. Murdock¹, F. J. Sedlazeck^{1,5}, R. A. Gibbs^{1,2}; ¹Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, ²Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, ³Texas Children's Hosp., Houston, TX, ⁴Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, ⁵Dept. of Computer Sci., Rice Univ., Houston, TX

Abstract Body:

Xia-Gibbs syndrome (XGS; OMIM: 615829) is a neurodevelopmental disorder (NDD) characterized by delayed motor milestones, speech delay, intellectual disability (ID) and hypotonia. Most individuals molecularly diagnosed with XGS through whole-exome sequencing harbor *de novo* truncating mutations in the AT-Hook DNA Binding Motif Containing 1 gene, *AHDC1*, that are predicted to prematurely terminate translation leading to a truncated protein. Some missense mutations in critical regions of the *AHDC1* protein also lead to XGS. In addition, recent reports of large *de novo* heterozygous deletions that encompass the *AHDC1* gene in individuals with symptoms that overlap those of XGS individuals have been ascribed as diagnostic for the disorder. The evidence that *AHDC1* haploinsufficiency is pathogenic is incomplete; however, primarily since there are usually many flanking genes involved in the contiguous deletion intervals. We analyzed 19 individuals with large (up to 20 Mb) deletions that involve *AHDC1* along with other genes in the region, to dissect genotype-phenotype relationships. We focused on one informative proband, with the smallest known contiguous *AHDC1* deletion of ~350Kb, encompassing eight other genes within chr1p36.11 (*FGR*, *IFI6*, *FAM76A*, *STX12*, *PPP1R8*, *THEMIS2*, *RPA2*, *SMPDL3B*) and terminating within the first intron of *AHDC1*. We identified the breakpoint junctions and resolved the phase of the deletion using a combination of short and long-read sequencing, via the Oxford Nanopore sequencing platform. Molecular characterization and quantification of RNA expression patterns in whole blood revealed that *AHDC1* exhibited a mono-allelic expression pattern, but we observed no deficiency in overall *AHDC1* expression levels, in contrast to the other deleted genes that showed a 50% reduction in mRNA expression. These results suggest that *AHDC1* expression in this individual is being compensated potentially by a novel regulatory mechanism and that some *AHDC1* deletions may be benign, depending on the breakpoints, phasing and molecular status. This knowledge informs how the *AHDC1* gene is regulated and advances understanding of regulatory mechanisms in NDDs.

Mendelian Phenotypes Posters - Thursday

PB1906. Long term prognosis in two adult patients with malonyl CoA decarboxylase deficiency

Authors:

S. Yano, R. McGowan, K. Moseley; LAC+USC Med Ctr, Keck Sch. of Med, USC, Los Angeles, CA

Abstract Body:

Background: Malonyl-CoA decarboxylase deficiency (Malonic aciduria) is a rare autosomal recessive inborn error of metabolism. Approximately 40 patients, among whom only a few patients were diagnosed by newborn screening (NBS), have been reported. Developmental delay, seizures, hypoglycemia, metabolic acidosis, abnormal brain MRI findings, and cardiomyopathy are frequently associated. Long-term clinical outcomes into adulthood are not well known. *Case reports:* We report on a pair of adult siblings, aged 32y and 24y, with Malonic Aciduria. These patients were diagnosed at age 4y and birth, respectively, based on plasma/urine metabolic analytes with enzymatic confirmation with skin fibroblasts (older sibling: 7% of normal). Treatment includes dietary interventions with a low-fat diet with MCT supplementation and medications with carnitine and an ACE inhibitor. The older sibling had episodes of respiratory distress and an enlarged heart at 4 days of age, hypoglycemia, metabolic acidosis, and seizures (cerebral infarction) at 2 months. This patient has microcephaly and a global developmental delay. Younger sibling has been treated with dietary interventions and medication for cardiomyopathy since early infancy. He has never developed a metabolic decompensation, i.e., hypoglycemia and metabolic acidosis. He does not have microcephaly or seizures but has a mild intellectual disability. He completed college and was employed until recently. Both elder and younger siblings developed hypertriglyceridemia and later hypercholesterolemia at age 25y and 10y, respectively. *Discussion:* Long-term prognosis of patients with Malonic aciduria remains unclear. Since NBS has been introduced, patients with Malonic aciduria can be diagnosed at birth. Malonyl-CoA decarboxylase plays a role in the regulation of fatty acid synthesis and oxidation. Hypertriglyceridemia has not been documented in this condition. The elder and younger siblings developed hypertriglyceridemia and later hypercholesterolemia approximately at age 25y and 10y, respectively. Treatment for hyperlipidemia with an HMG-CoA reductase inhibitor alone may have a potential risk to cause a depletion of free CoA. Further research may look at the benefits of treatment with C7 molecule compared to MCT oil (C8) for the prevention of hyperlipidemia and intellectual disability as well as improvement in cardiomyopathy.

Mendelian Phenotypes Posters - Wednesday

PB1907. Long-term clinical course of Heyn-Sproul-Jackson syndrome.

Authors:

H. Futagawa¹, K. Fukuda¹, H. Yamanaka¹, M. Kuroda¹, S. Ito¹, M. Honda¹, H. Suzuki², M. Yamada², T. Takenouchi², K. Kosaki², H. Yoshihashi¹; ¹Tokyo Metropolitan Children's Med. Ctr., Fuchu, Tokyo, Japan, ²Keio Univ. Sch. of Med., Tokyo, Japan

Abstract Body:

<Introduction> In *DNMT3A*(2q23.3), loss-of-function variant caused Tatton-Brown-Rahman syndrome (MIM#615879) which characterized by overgrowth, intellectual disability and dysmorphic facial features, while gain-of-function variant cause Heyn-Sproul-Jackson syndrome (MIM#618724:HESJAS) which characterized with microcephalic dwarfism and developmental delay (Heyns et al, 2019). Here, we report the oldest patient of HESJAS and add the detailed clinical description. <Patient> 24-year-old woman. She was born at 41 weeks of gestation. Birth weight 2492g (-2.3SD) and head circumference (OFC) 32.5cm (-0.9SD). Developmental milestones achieved hold head up at 3-month, roll over at 5-month, sit without support at 9-month and independent walk at 18-month. At 3-year of age, she was diagnosed developmental delay and included special education support. Visual acuity declined rapidly between the ages of 20 to 22 due to progressive cataract. She does not have other visceral complication. Now, her height, body weight and OFC were 138.2cm (-4.0SD), 45.0kg (-1.8SD) and 49.3cm (-4.8SD), respectively. She employed easy work at welfare facility because of mild intellectual disability. Her dysmorphic features included microcephaly, sparse scalp hair, blepharophimosis, thin upper lip vermilion and brachydactyly. WES (IRUD) revealed de novo pathogenic variant in *DNMT3A* (NM_022552.4:c.916T>C p.Trp306Arg) that was assumed to be gain of function, and a diagnosis of HESJAS was made. <Discussion> Previous reports of HESJAS have been limited to infants (1-4 years of age), with no reports in adults. In previous reports, developmental delay, severe short stature and microcephaly exceeding -4.0 SD were common findings, and this patient had similar features. While the developmental status of present patient was mild, reported patients have moderate to severe. There seemed to be the wide spectrum of the development. The characteristic feature of present patient was the rapid loss of vision acuity in adulthood. According to the research of *Drosophila* which does not have endogenous methylation, induced the expression of *Dnmt3a* into *Drosophila* resulted to microphthalmia or anophthalmia. It is possible that a gain-of-function mutation in the *DNMT3A* may have caused ocular symptoms associated with HESJAS. Further investigation is warranted in future case series. Somatic variants in the *DNMT3A* are known to be associated with hematopoietic disease. In general, the age of onset is middle-aged, and preventive health care regarding tumor development was considered for this patient. <Conclusion> Management of HESJAS in adulthood should anticipate progressive ocular disease.

Mendelian Phenotypes Posters - Thursday

PB1908*. Loss of adipocyte phospholipase gene *PLAAT3* causes lipodystrophy with neurological features due to inactivated arachidonic acid-mediated PPAR γ ; signaling

Authors:

N. Schuermans¹, S. El Chehadeh², D. Hemelsoet³, M-C. Vantyghem⁴, J. Gautheron⁵, E. Bogaert¹, H. Morishita⁶, T. Eguchi⁶, P. Hilbert⁷, N. Van Doninck⁸, M-C. Taquet⁹, T. Rosseel¹, G. De Clercq¹, E. Debackere¹, C. Van Haverbeke¹⁰, J-B. Chanson¹¹, B. Funalot¹², F-J. Authier¹³, S. Kaya¹⁴, W. Terryn¹⁵, S. Callens¹⁶, B. Depypere¹⁷, J. Van Dorpe¹⁰, B. Poppe¹, F. Impens¹⁸, N. Mizushima⁶, I. Jéru⁵, C. Depienne¹⁹, B. Dermaut²⁰; ¹Ctr. for Med. Genetics, Ghent Univ. Hosp., Ghent, Belgium, ²Service de Génétique Médicale, Hôpital de Hautepierre, Strasbourg, France, ³Dept. of Neurology, Ghent Univ. Hosp., Ghent, Belgium, ⁴Endocrinology, diabetology, metabolism Dept., PRISIS competence center, Lille Univ. Hosp., Lille, France, ⁵Sorbonne Université-Inserm UMRS_938, Ctr. de Recherche Saint-Antoine (CRSA), Paris, France, ⁶Dept. of Biochemistry and Molecular Biology, Graduate Sch. and Faculty of Med., The Univ. of Tokyo, Tokyo, Japan, ⁷Inst Pathologie et Genetique, Gosselies, Belgium, ⁸Gen. Hosp. AZ Nikolaas, Dept. of Endocrinology and Diabetology, Sint-Niklaas, Belgium, ⁹Dept. of Internal Med. and Nutrition, Hopitaux Univ.ires Strasbourg, Strasbourg, France, ¹⁰Dept. of Pathology, Ghent Univ. Hosp., Ghent, Belgium, ¹¹Service de neurologie Hôpital de Hautepierre, Strasbourg, Strasbourg, France, ¹²CHU Dupuytren, Limoges, France, ¹³Inserm UMR955, Créteil, Créteil, France, ¹⁴Insitut für Humangenetik, Univ.skllinikum Essen, Essen, Germany, ¹⁵Dept. of Nephrology, Jan Yperman Ziekenhuis, Ieper, Belgium, ¹⁶Dept. of Gen. Internal Med., Ghent Univ. Hosp., Ghent, Belgium, ¹⁷Dept. of Plastic and Reconstructive Surgery, Ghent Univ. Hosp., Ghent, Belgium, ¹⁸VIB Proteomics Core, VIB, Ghent, Belgium, ¹⁹Inst. für Humangenetik, UK Essen, Essen, Germany, ²⁰Ghent Univ., Ghent, Belgium

Abstract Body:

PLAAT3 is a phospholipid modifying enzyme predominantly expressed in white adipose tissue (WAT). It is a candidate drug target as *Plaat3* deficiency in mice protects against picornavirus infection and diet-induced obesity. We identified five patients from three independent consanguineous families, with homozygous loss-of-function mutations in *PLAAT3*, presenting with severe lipodystrophy and neurological features including intellectual disability and a demyelinating peripheral neuropathy. PLAAT3-deficient mice and human WAT showed a failure to liberate arachidonic acid (AA) from membrane phospholipids resulting in an inactive gene network downstream of adipogenesis master regulator and anti-diabetic drug target PPAR γ . CRISPR/Cas9-mediated *PLAAT3*^{-/-} human adipose stem cells displayed insulin resistance and showed a disturbed differentiation characterized by a significant decrease in lipid droplet formation and a downregulation of PPAR γ ; and perilipin. These findings establish PLAAT3 deficiency in humans as a novel type of hereditary lipodystrophy due to an AA- and PPAR γ -dependent defect in WAT differentiation and function.

Mendelian Phenotypes Posters - Wednesday

PB1909*. Loss of C-terminal Mediator Complex subunit-11 impairs fetal brain development and cause severely progressive neurodegeneration

Authors:

E. Cali¹, S-J. Lin², C. Rocca¹, Y. Sahin³, A. Al Shamsi⁴, S. El Chehadeh⁵, M. Chaabouni⁶, K. Mankad⁷, S. Eftymiou⁸, D. Murphy¹, SYNaPS Study Group, C. Petree², K. Huang⁹, K. Monastiri¹⁰, M. Ruggieri¹¹, Y. Şahin¹², L. Al Gazali¹³, M. Abi Warde¹⁴, B. Gerard¹⁵, G. Zifarelli¹⁶, C. Beetz¹⁷, S. Fortuna¹⁸, M. Soler¹⁹, E. Valente²⁰, G. Varshney², R. Maroofian¹, V. Salpietro¹, H. Houlden¹; ¹Univ. Coll. London, London, United Kingdom, ²Oklahoma Med. Res. Fndn., Oklahoma City, OK, ³Genoks Genetic Lab., Ankara, Turkey, ⁴Tawam Hosp., Abu Dhabi, United Arab Emirates, ⁵Hôpitaux Univ.ires de Strasbourg, Strasbourg, France, ⁶LGC genetics, Tunis, Tunisia, ⁷Great Ormond Street Hosp. for Children, Dept. of Radiology, London, United Kingdom, ⁸UCL Inst. of Neurology, London, United Kingdom, ⁹Oklahoma Med. Res. Fndn., OK, OK, ¹⁰Fattouma Bourguiba Hosp., Monastir, Tunisia, ¹¹AOU "Policlinico", PO "G. Rodolico", Catania, Italy, ¹²Biruni Univ.si, Biruni, Turkey, ¹³Coll. of Med. & Hlth.Sci., UAE, United Arab Emirates, ¹⁴CHU Strasbourg, Strasbourg, France, ¹⁵Laboratoires de Diagnostic Génétique, Unité de génétique moléculaire, Nouvel Hôpital Civil, Strasbourg, France, ¹⁶Centogene AG, Rostock, Germany, ¹⁷Centogene US LLC, Rostock, Germany, ¹⁸Univ. of Trieste, Trieste, Italy, ¹⁹Italian Inst. of Technology (IIT), Genova, Italy, ²⁰Univ. of Pavia: Univ. degli Studi di Pavia, Pavia, Italy

Abstract Body:

The Mediator (MED) is an evolutionarily conserved multi-subunit protein complex that modulates the activity of several transcription factors and different critical components of the overall transcriptional machinery. Individual deficits in subunits part of the MED complex have been implicated in different neurological disorders. Exome or genome sequencing were performed in five unrelated families identified via different research networks and Matchmaker Exchanges. Deep clinical and brain imaging evaluations were performed by clinical pediatric neurologists and neuroradiologists. The functional impact of the candidate variant on both MED11 RNA and protein was assessed by RT-PCR and Western Blotting using fibroblast cell lines derived from one affected individual and controls and by computational approaches. Knockouts in zebrafish were generated by CRISPR/Cas9.

A recurrent, segregating homozygous variant in MED11 (c.325 C>T; p.Arg109Ter) was identified in all seven affected children. The variant results in a premature stop codon and a putative protein lacking the last nine residues of MED11 C-terminal.

The disease is characterized by microcephaly, profound neurodevelopmental impairment, exaggerated startle response, myoclonic seizures, progressive widespread neurodegeneration, and premature death. Functional studies on patient-derived fibroblasts did not show a loss-of-protein-function but rather disruption of the C-terminal of MED11, likely impairing binding to other MED-subunits. A zebrafish knockout model recapitulates key clinical phenotypes (microcephaly, premature death, exaggerated response to light stimuli).

Our results establish MED11 as a novel gene causing a severe neurodegenerative disorder potentially associated with premature death in humans. Loss of the C-terminal of MED subunit 11 may affect its binding efficiency to other MED subunits, thus implicating the MED-complex stability in brain development and 'MEDopathies' in neurodegeneration.

Mendelian Phenotypes Posters - Thursday

PB1910*. Loss of FOCAD, operating in the SKI mRNA surveillance pathway, causes a pediatric syndrome with liver cirrhosis.

Authors:

R. Moreno Traspas¹, T. Teoh¹, P. Wong¹, M. Maier¹, K. Lay¹, N. Ali¹, A. Larson², F. Al Mutairi³, N. Abbas Al-Sannaa⁴, E. Ali Faqeih⁵, H. Arshad Cheema⁶, J. Dupont⁷, J. Dupont⁷, R. Sandoval⁸, L. Ramos⁹, P. Bauer¹⁰, B. Cogne¹¹, A. Bertoli-Avella¹², M. Vincent¹¹, K. Girisha¹³, B. REVERSADE¹⁴; ¹A*STAR - Genome Inst. of Singapore, Singapore, Singapore, ²Children's Hosp. Colorado, Aurora, CO, ³King Abdulaziz Med. City Riyadh, Riyadh, Saudi Arabia, ⁴John Hopkins Aramco Hlth.care, Dhahran, Saudi Arabia, ⁵Children's Specialist Hosp., King Fahad Med. City, Riyadh, Saudi Arabia, ⁶The Children's Hosp., Lahore, Pakistan, ⁷Lisbon North Univ. Hosp. Ctr., Lisbon, Portugal, ⁸Hosp. Sírío-Libanês, Brasilia, Brazil, ⁹Mendelics Genomic Analysis, Sao Paulo, Brazil, ¹⁰Centogene GmbH, Rostock, Germany, ¹¹Nantes Univ. Hosp. Ctr., Nantes, France, ¹²Centogene AG, Rostock, Germany, ¹³Kasturba Med. Coll., Manipal, Udupi, India, ¹⁴Genome Inst. of Singapore, Singapore, Singapore

Abstract Body:

Cirrhosis is a chronic, life-threatening disease characterized by fibrotic scarring and inflammation that disrupts liver architecture and function. It is typically the result of alcoholism or hepatitis viral infection. In this study we report 14 children from 10 unrelated families presenting with a syndromic form of pediatric liver cirrhosis. By genome/exome sequencing, recessive variants in *FOCAD* were found to segregate with the disease. Zebrafish lacking *focad* phenocopied the human disease, revealing a signature of altered mRNA degradation processes in the liver. Using patient's primary cells and CRISPR/Cas9-mediated inactivation in human hepatic cell lines, we find that FOCAD deficiency compromises the SKI mRNA surveillance pathway by reducing the levels of the RNA helicase SKIC2 and its cofactor SKIC3. FOCAD knockout hepatocytes exhibited lowered albumin expression and signs of persistent injury accompanied by CCL2 overproduction. Our results reveal the importance of FOCAD in maintaining liver homeostasis and disclose a possible therapeutic intervention point via inhibition of the CCL2/CCR2 signalling axis.

Mendelian Phenotypes Posters - Wednesday

PB1911. *LYRM7* mutations causing mitochondrial complex 3 deficiency nuclear type 8 (MC3DN8) in a Five-year-old boy without cavitating leukoencephalopathy

Authors:

R. Alfattal¹, **H. A. Alsharhan**², A. Algaith³, B. Albash⁴, R. M. Elshafie⁵, A. Alshammari⁴, A. Alahmad⁴, F. Dashti⁶, R. Alsafi⁷; ¹Dept. of Pediatrics, Amiri Hosp., Ministry of Hlth., Kuwait City, Kuwait, ²Dept. of Pediatrics, Farwania Hosp., Kuwait Med. Genetic Ctr., Ministry of Hlth., Faculty of Med., Kuwait Univ., Sabah Al-Nasser, Kuwait, ³Faculty of Med., Szeged Univ., Szeged, Hungary, ⁴Kuwait Med. Genetic Ctr., Ministry of Hlth., Sulaibikhat, Kuwait, ⁵Dept. of Pediatrics, Hlth.Sci. Ctr., Faculty of Med., Kuwait Univ., Kuwait City, Kuwait, ⁶Dept. of Radiology, Ibn Sina Hosp., Ministry of Hlth., Shuwaikh, Kuwait, ⁷Dept. of Pediatrics, Adan Hosp., Ministry of Hlth., Hadiya, 52700, Kuwait

Abstract Body:

Abstract: Defects of complex III (CIII) respiratory chain, result in characteristic but rare mitochondrial disorders associated with distinct neuroradiological findings. The underlying molecular defects affecting mitochondrial CIII assembly factors are few and yet to be identified. *LYRM7* assembly factor is required for proper CIII assembly where it acts as a chaperone for the Rieske iron-sulfur (UQCRFS1) protein in the mitochondrial matrix and stabilizing it. We present here the fifteenth individual with *LYRM7*-associated mitochondrial leukoencephalopathy secondary to a previously reported pathogenic homozygous *LYRM7* variant, c.2T>C, p.(p.Met1?). Similar to previously reported individuals, we report a 4-year-old male proband, who presented with recurrent metabolic and lactic acidosis, encephalopathy and myopathy. Further, he has additional, previously unreported features, including acute stroke like episode with bilateral central blindness and optic neuropathy, recurrent hyperglycemia and hypertension associated with metabolic crisis. However, he has no signs of psychomotor regression. He has been stable clinically with residual left-sided blindness, with no more metabolic crises for 2-year-period on mitochondrial cocktail. Although the reported brain MRI findings in other affected individuals are homogenous, it is slightly different in our index, revealing evidence of bilateral almost symmetric multifocal periventricular T2 hyperintensities with hyperintensities of the optic nerves and the optic chiasm but with no cavitation or cystic changes. This report describes new phenotypic and radiological spectrum of *LYRM7*-associated CIII deficiency. It also summarizes the clinical and molecular data of the previously reported individuals to describe the full phenotypic spectrum.

Mendelian Phenotypes Posters - Thursday

PB1912. *MED13L* knockout in cerebral organoids leads to a shifted developmental program through abnormal cis-regulatory element activation

Authors:

J. Ghomid^{1,2}, J. Sige¹, M. Balerdi¹, D. Laboy³, R. Ziffra³, N. Ahituv³; ¹Univ. of Lille, Lille, France, ²Lille Univ. Hosp., Lille, France, ³UCSF, San Francisco, CA

Abstract Body:

The Mediator is a large coregulator complex conserved from yeast to humans. It has emerged as a master coordinator of development and cell lineage determination through interactions with various transcription factors. Pathogenic variants of the gene cause the *MED13L*-syndrome, a neurodevelopmental disorder including intellectual disability. The role of the gene in neurodevelopment is unknown, and the consequences of *MED13L* variants remain to be deciphered.

We developed a *MED13L* ko cerebral-organoid model from hPSc and analyzed, at the single-cell level, the transcriptome (sc-RNAseq) and the chromatin accessibility (scATAC-seq). Sc-RNAseq and ScATAC-seq data analysis show that wt organoids produce cortical neurons, and *MED13L* ko organoids produce only neuroretinal cells. In the *MED13L* ko model, we observed coordinated activation of O5-O7 and O9 *OTX2* CREs as reported in neuroretinal development, while early specific forebrain *OTX2* CREs, AN1, and AN2, were not activated. At *PAX6* locus, we also observed differential expression of the neuroretinal specific CREs HS5, HS3, and HS2, in the ko model.

OTX2 and *PAX6* are early markers of both cerebral cortex and neuroretinal development. We suppose that activation of neuroretinal specific CREs driving expression of both latter genes would explain the downstream upregulation of genes involved in neuroretinal differentiation, including *CRX*, *VSX*, and *USH2A*. These results partially explain the shifted developmental program in ko organoids. Based on these data, *MED13L* is probably critical for proper enhancer activation during the neural induction. Then, neurodevelopmental disorder observed in patients with *MED13L* pathogenic variants is possibly linked to a global neural gene expression misregulation.

Mendelian Phenotypes Posters - Wednesday

PB1913. Metatranscriptomics detects emerging multidrug resistant *Candida auris* in a family with a mild TP63 associated ectodermal dysplasia

Authors:

P. Fortina¹, L. Youssefian¹, J. Park¹, A. Saeidian², F. Palizban³, S. Khodavaisy⁴, J. Uitto¹, H. Vahidnezhad⁵; ¹Thomas Jefferson Univ, Philadelphia, PA, ²Children's Hosp. of Philadelphia, Philadelphia, PA, ³Univ. of Tehran, Tehran, Iran, Tehran, Iran, Islamic Republic of, ⁴Tehran Univ. of Med. Sci., Tehran, Tehran, Iran, Tehran, Iran, Islamic Republic of, ⁵Thomas Jefferson Univ., Philadelphia, PA

Abstract Body:

Tumor protein 63, encoded by TP63, has an essential role in epidermal differentiation and ectodermal development. Autosomal dominant mutations in human TP63 result in several overlapping syndromes, with some established genotype-phenotype correlations based on the protein domain affected. Several groups have linked TP63-associated ectodermal dysplasia (ED) syndromes with varying degrees of immune dysfunction. In this study, we followed the clinical course of a daughter and father who presented with mild ED symptoms, chronic mucocutaneous candidiasis (CMC), recurrent bacterial infections, and equivocal immune function status. Whole-exome sequencing (WES) and whole-transcriptome analysis via RNA sequencing (RNA-Seq) disclosed a recurrent monoallelic mutation in TP63. Additionally, RNA-Seq data was used for detection of non-albicans *Candida* species, *Candida auris* and *Candida parapsilosis*, in CMC lesions of the daughter and father, respectively. Although a strong association between ED and TP63 mutations exists, the mechanism of immune dysregulation remains unclear. Our cases attest to the variable expressivity of TP63-associated disorders and the potential role of TP63 in maintaining immune system integrity. Furthermore, this study represents the first case of CMC caused by the multidrug-resistant pathogen *C. auris* in the context of TP63-associated ED and demonstrates the utility of RNA-Seq for concomitant mutation detection and accurate identification of rare and often misidentified pathogens.

Mendelian Phenotypes Posters - Thursday

PB1914. Metatranscriptomics reveals association of α -, β -, and γ -HPVs with typical epidermodysplasiaverruciformis in a large cohort of patients with CIB1, TMC6, or TMC8 mutations

Authors:

L. Youssefian; Thomas Jefferson Univ, Philadelphia, PA

Abstract Body:

Cutaneous human papillomavirus (HPV) infection typically manifests with isolated warts. However, some patients in familial clustering develop extensive and protracted HPV infections, primarily the β -HPV types 5 and 8, with distinct cutaneous findings. This clinical entity, epidermodysplasia verruciformis (EV), with autosomal recessive inheritance, is characterized by numerous cutaneous flat warts in childhood, which progress into squamous cell carcinomas later in life. The “typical” form of EV, not vulnerable to other infections, is caused by mutations in CIB1, TMC6, or TMC8, which impair keratinocyte-intrinsic immunity to β -HPV infection. Mutations in other genes related to T-cell development or function have been associated with “atypical” EV in patients with other infections. We developed a whole-transcriptome sequencing-based method on RNA isolated from skin biopsies for concomitant detection of viral and human genetic determinants of cutaneous wart lesions in a cohort of 50 EV patients. This method, VirPy, can detect 926 viruses, including more than 400 HPVs, and the corresponding human mutations. Nine distinct mutations in TMC8 (n=2), TMC6 (n= 5) and CIB1 (n=2) in 12 distinct families, including 14 patients were detected. The most predominant HPV in this cohort was HPV14. In addition, the RNA-seq data were examined for variant detection and prioritization, pathogenicity confirmation, and RNA expression profiling. Besides, we identified a total of 20 different HPVs including 16 β -, three α -, and one γ -HPV (HPV128) in a patient with TMC8 mutation. In summary, the utilization of RNA-Seq as a first-tier diagnostic method allowed us to simultaneously profile the transcriptome of the host for mutation detection and explore the consequences of variants of unknown significance as well as profile the cutaneous virome of EV patients.

Mendelian Phenotypes Posters - Wednesday

PB1915. Modeling dystroglycanopathy in zebrafish through CRISPR-Cas9-mediated knockout of *pomgnt2*

Authors:

K. Flannery, N. Battula, B. Karas, K. Terez, M. Manzini; Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ

Abstract Body:

Dystroglycanopathies are autosomal recessive neuromuscular disorders characterized by congenital muscular dystrophy (CMD) with brain and/or ocular malformations. These disorders are clinically heterogeneous, ranging from milder limb-girdle muscular dystrophy to Walker Warburg Syndrome, which is the most severe form of CMD leading to childhood lethality. Mutations in 19 genes have been implicated in these diseases all converging on the regulation of glycosylation of the transmembrane glycoprotein alpha-dystroglycan (alpha-DG). Alpha-DG is component of the dystrophin-glycoprotein complex which is critical to connect the actin cytoskeleton to the extracellular matrix in the muscle, eye, brain, and other tissues. One major hurdle in studying the pathogenesis of these genetically heterogeneous disorders is the lack of appropriate models since many of the genes mutated in humans lead to early embryonic lethality in mice. In this study we sought to develop a novel zebrafish model of dystroglycanopathy targeting the glycosyltransferase *Protein O-linked Mannose N-acetylglucosaminyl Transferase 2 (pomgnt2)* which is mutated in severe CMD. We targeted *pomgnt2* through CRISPR-Cas9 microinjections in zebrafish embryos using two parallel approaches: F0 crispants and stable mutant lines. The crispant approach targets the first few cells after fertilization leading to mosaic larvae with a high degree of somatic mutations and is faster than developing stable mutants. Crispant *pomgnt2* larvae injected with two independent guide RNAs exhibited a spectrum of phenotypes over the first 4 days post fertilization (dpf) similar to previously published morpholino oligonucleotide knockdowns. A large portion of the crispants showed shortened, curved tails with impaired mobility and several had severe nonviable lesions, including underdeveloped head and eyes, severe spinal alterations, and pericardial and yolk sac edema. In parallel, several juvenile F0 zebrafish were outbred to establish multiple stable *pomgnt2* mutant zebrafish lines which can be used alongside crispants for further analyses. We intend to characterize phenotypes, muscle and eye integrity, and brain abnormalities caused by *pomgnt2* loss of function in both crispants and mutants, as well as perform as high-throughput drug screenings of compounds with the potential to ameliorate the effects of *pomgnt2* loss of function.

Mendelian Phenotypes Posters - Thursday

PB1916. Molecular and behavioral characterization of a novel mouse model of Snyder-Robinson Syndrome: A path towards therapeutic development

Authors:

D. Kemaladewi, O. Akinyele, M. Johnson, A. Munir, M. Perez; Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA

Abstract Body:

The polyamines putrescine, spermidine, and spermine are ubiquitous polycationic molecules crucial for cellular processes such as gene expression, signal transduction, and cell proliferation. Abnormal accumulation of spermidine due to mutations in *SMS* gene, encoding spermine synthase protein, causes Snyder-Robinson syndrome (SRS), a rare X-linked recessive disorder that manifests as mental retardation, thin habitus, and low muscle tone with no available treatments. The development of effective therapy for SRS requires a suitable disease-specific animal model that recapitulates the abnormalities seen in patients.

A novel SRS mouse model carrying a missense mutation in the *Sms* gene, leading to a glycine-to-serine substitution at position 56 (G56S) in the SMS protein, was generated. We aim to study the molecular mechanisms underlying SRS pathogenesis and develop a therapeutic genome editing strategy using the G56S mice. We hypothesize that *Sms* mutations lead to neurological and musculoskeletal pathophysiology in the G56S mice and mimic the disease presentation in SRS patients.

We measured SMS protein expression by western blot and determined polyamine contents using mass spectrometry. We evaluated brain structure using T2-weighted MRI and bone density by micro-CT scan. We carried out a battery of behavioral assays, such as Morris Water Maze, Fear Conditioning, and open field tests to investigate their learning ability and anxiety level. RNA-Sequencing was performed to seek a novel disease pathomechanism in these mice.

We observed a complete loss of SMS protein in the G56S mouse brain, skeletal muscles, and liver, resulting in a significant increase in the spermidine content. In addition, the G56S mice showed a significant reduction in body weight, length, bone density, and behavioral activity towards the center zone of an open field. The mice also exhibit a significant level of anxiety and impaired learning ability compared to the wild type. Furthermore, there were significant volumetric reductions in the hippocampal and amygdala, which further support the neurological pathophysiology in these mice. Transcriptomic analyses reveal downregulation of genes involved in oxidative phosphorylation and EIF2 signaling in the brain, providing the molecular mechanism that can be further exploited for therapy.

Our data indicate that the G56S mouse recapitulates the molecular and phenotypic abnormalities seen in SRS patients. Current experiments focus on rescuing these defects using genetic and pharmacological approaches, which will open the door to future therapeutic interventions for this condition.

Mendelian Phenotypes Posters - Wednesday

PB1917. Molecular and clinical review of 95 Polish pediatric patients with the clinical suspicion of Alport syndrome.

Authors:

P. Halat-Wolska¹, L. Obrycki², E. Ciara¹, M. Pac², K. Gadomska-Prokop², M. Rydzanicz³, P. Stawinski^{1,3,4}, J. Feber⁵, B. Chalupczynska¹, D. Jurkiewicz¹, M. Pelc¹, D. Piekutowska-Abramczuk¹, D. Siestrzykowska¹, D. Wicher¹, K. Chrzanowska¹, R. Ploski³, M. Litwin²; ¹The Children's Mem. Hlth.Inst., Dept. of Med. Genetics, Warsaw, Poland, ²The Children's Mem. Hlth.Inst., Dept. of Nephrology, Warsaw, Poland, ³Warsaw Med. Univ., Dept. of Med. Genetics, Warsaw, Poland, ⁴Inst. of Physiology and Pathology of Hearing, Dept. of Genetics, Warsaw, Poland, ⁵The Children's Hosp. of Eastern Ontario, Div. of Nephrology, Dept. of Pediatrics, Ontario, ON, Canada

Abstract Body:

Introduction: Alport syndrome (AS) is a genetically heterogeneous nephropathy resulting from *COL4A3*, *COL4A4* and *COL4A5* pathogenic changes. AS can be transmitted as X-linked, autosomal recessive (AR) or dominant (AD) difficult to differentiate from benign familial hematuria (BFH). The symptoms vary widely from mild to severe among patients. Because of the significant lifetime risk for kidney failure, hearing and ocular symptoms, an early diagnosis, detailed genotype-phenotype correlation and personalized therapy is crucial. **Materials and Methods:** We examined 95 Polish pediatric patients with suspicion AS by next-generation sequencing (NGS) of glomerulopathy genes panel. **Results:** Clinical diagnosis was confirmed at the molecular level with 81% compliance (n=77). Overall 62 pathogenic/likely pathogenic *COL4A3-5* variants were revealed (known=36, novel=26). About 97% (n=60) of them were SNVs: frameshift, in-frame, missense, nonsense, splice-site and CNVs: deletions of single and six *COL4A5* exons. In 70% (n=54) the inheritance was X-linked *COL4A5*, 21% (n=16) AD AS/BFH and 6% (n=5) AR *COL4A3-4*. We revealed a recurrent *COL4A5* c.1871G>A p.G624D variant, with allele frequency 30% (17/56). Of these, 2 patients carried additional *COL4A3* heterozygous change, resulted in digenic inheritance (3%). Genotypes were compared with the clinical manifestations: renal function, proteinuria, microalbuminuria, hematuria, arterial hypertension, hearing and ocular abnormalities. Microalbuminuria and proteinuria excretion were significantly different between the genotypes. Variant p.G624D gave milder symptoms (older age hematuria, lower microalbuminuria and proteinuria, rare ocular abnormalities) than other *COL4A5* changes. The phenotype of patients with digenic AS did not differ significantly in the study group. Diagnosis age, hypertension, hypertension with >500mg/24h proteinuria, microalbuminuria, ocular symptoms were the predictors of GFR reduction for all patients. **Conclusions:** The differences in AS phenotype occur depending on the genotype. NGS is an effective approach to obtain genetic information, which help to predict the clinical course, especially for patients with mild or atypical phenotype. Also, we broaden the molecular spectrum of AS, which will facilitate future research on the genotype-phenotype correlations. **Partially supported by CMHI M29/18, MEiN 7071/IB/SN/2020 and MEiN 7088/II-KDM/SN/2020**

Mendelian Phenotypes Posters - Thursday

PB1918. Molecular Genetic Defects of Hereditary Angioedema in Saudi Arabia.

Authors:

H. Alajlan¹, F. Sheikh¹, M. Albanyan¹, H. Alruwaili¹, B. Al-Saud¹, R. Arnaout¹, M. Al-Hamed¹, H. Al-Mousa¹, S. AlShareef¹, A. Alazami²; ¹KING FAISAL SPECIALIST Hosp. AND RESEARCH CENTER, Riyadh, Saudi Arabia, ²King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia

Abstract Body:

Hereditary Angioedema (HAE) is a rare life threatening inherited disorder characterized by swelling of the larynx, subcutaneous tissues, and the lining of the gut and lungs. These symptoms develop as the result of deficiency or dysfunction of proteins that help to maintain the normal flow of fluids through the capillaries. The mode of inheritance is autosomal dominant. HAE affects approximately 1 in 50,000 individuals, and the main gene linked to pathogenesis is *SERPING1*, which encodes C1 esterase inhibitor (C1-INH). Here we present the clinical and genetic profiles of HAE patients from a single Saudi institution, which is a major national referral center. Fifty-one (51) patients with a diagnosis of HAE were recruited, making this one of the largest cohorts ever studied in the world and the largest ever reported from the Arab world. Clinically, the primary HAE subtype in our cohort was type I, accounting for 76% of our cases. Polymerase chain reaction (PCR) analysis detected three (3) novel frameshift mutations in *SERPING1*, in addition to two (2) novel nonsense mutations. All individuals that tested negative for *SERPING1* by Sanger were evaluated by Multiplex Ligation Dependent Probe Amplification (MLPA). This identified a heterozygous ~900bp deletion affecting *SERPING1* exon 4 in a large multi-generational family. This was the sixth novel mutation uncovered in our study. In total, *SERPING1* mutations were uncovered in 49 patients (96%). The remaining unsolved cases (4%) were screened for *F12*, *PLG*, *ANGPT1*, *MYOF*, *KNG1* and *HS3ST6* but no relevant genetic variants were found.

Mendelian Phenotypes Posters - Wednesday

PB1919*. Molecular insights into the pathogenesis of Chediak-Higashi syndrome and the biology of Lysosomal Trafficking Regulator.

Authors:

M. Morimoto, E-R. Nicoli, C. Kuptanon, J. Serra-Vinardell, P. Sharma, C. Ciccone, M. Huizing, C. Toro, W. A. Gahl, W. J. Inrone, M. C. V. Malicdan; NIH, Bethesda, MD

Abstract Body:

Introduction: Lysosome-related organelles (LROs) are cell type-specific cellular compartments, each with a unique morphology, composition, and function; they contribute to diverse processes such as pigmentation, blood coagulation, immunity, and neurological function. Dysfunction in the biogenesis and trafficking of lysosomes and LROs lead to LRO-related disorders including Chediak-Higashi syndrome (CHS), a rare autosomal recessive disorder characterized by partial oculocutaneous albinism, a bleeding diathesis, immunological dysfunction, and neurological impairment. While CHS was first described in the 1940s and the gene Lysosomal Trafficking Regulator (*LYST*) was implicated as the genetic cause in 1996, molecular insights into how *LYST* deficiency leads to CHS have been elusive. The key challenges include the exceptionally large size of the gene (11.4 kb coding sequence) and protein (429 kDa), and the absence of a validated antibody. **Methods:** To further our understanding of the pathogenesis of CHS and the biology of *LYST*, novel pathogenic *LYST* variants were identified by genomic DNA and cDNA Sanger sequencing within our CHS patient cohort, relative *LYST* mRNA expression was assessed in several cell types by quantitative PCR, and *LYST*-interacting candidates were identified in melanocytes by immunoprecipitation (IP)-mass spectrometry (MS). **Results:** 23 novel pathogenic *LYST* variants were identified in patients with a clinical diagnosis confirmed by the presence of pathognomonic giant intracellular granules. *LYST* mRNA expression varied widely among different cell types; there was no detectable expression in iPSCs, low expression in fibroblasts and iPSC-derived neurons, moderate expression in NK cells, and high expression in melanocytes. CHS patient fibroblasts, NK cells, and melanocytes generally had decreased relative *LYST* mRNA expression. To advance the protein studies of *LYST*, rabbit polyclonal antibodies against 2 *LYST* peptides were developed and tested in parallel with 3 commercially available antibodies. Three of the 5 antibodies exhibited a high molecular weight band that was present in control but absent in CHS patient melanocytes by immunoblot analysis. IP-MS uncovered several novel *LYST*-interacting candidates. **Conclusions:** The identification of novel pathogenic *LYST* variants, the analysis of *LYST* expression among different cell types, including those from patients, and the preliminary detection of *LYST*-interacting proteins will contribute to the earlier diagnosis of CHS patients and serve as the groundwork for understanding the mechanism of disease and the identification of therapeutic targets.

Mendelian Phenotypes Posters - Thursday

PB1920. Monoallelic variation in the DExH-box helicase DHX9, a product of the *DHX9* gene paralog, perturbs neurodevelopment & causes peripheral nerve axon degeneration

Authors:

D. Calame¹, L. Garrett², A. Jolly³, M. Dawood¹, A. Kurolap⁴, J. Hunter⁵, R. Spillmann⁶, S. Lalani⁷, A. Revah Politi⁸, S. RONDEAU⁹, O. Clothide⁹, G. Barcia⁹, Q. Tan⁶, I. Thiffault¹⁰, K. Sheikh¹¹, S. Biliciler¹¹, D. Mei¹², F. Melani¹², V. Shashi⁶, Y. Yaron¹³, P. Gawlinski¹⁴, W. Wiszniewski¹⁵, J. Posey¹, R. Gibbs⁷, R. Guerrini¹², S. Hölter², J. Lupski¹; ¹Baylor Coll. of Med., Houston, TX, ²Helmholtz Zentrum München, Neuherberg, Germany, ³1 Baylor Plaza, Houston, TX, ⁴Tel Aviv Sourasky Med. Ctr., Tel Aviv, Israel, ⁵Texas Children's Hosp., Houston, TX, ⁶Duke Univ., Durham, NC, ⁷Baylor Coll. Med., Houston, TX, ⁸Inst. for Genomic Med.- Columbia Univ., New York, NY, ⁹Necker Enfants Malades Hosp, Paris, France, ¹⁰Childrens Mercy, Kansas City, MO, ¹¹UT Hlth.Sci. Ctr. at Houston, Houston, TX, ¹²Univ. of Florence, Florence, Italy, ¹³Tel Aviv Sourasky Med Ctr, Tel Aviv, Israel, ¹⁴Inst. of Mother and Child, Warsaw, Poland, ¹⁵Oregon Hlth.& Sci. Univ, Portland, OR

Abstract Body:

Background: DExD/H-box RNA helicases are a large paralogous gene family implicated in neurodevelopmental disorders (NDD) and cancer. DExH-box helicase 9, encoded by the gene *DHX9*, is highly expressed in the developing and adult nervous system and regulates functions frequently disrupted in NDD including transcription, triple stranded DNA-RNA hybrids called R-loops, and homologous recombination (HR); yet its function in the neuronal development and homeostasis remains enigmatic.

Methods: To explore *DHX9*'s role in the human nervous system, we performed a primary analysis of >30,000 exomes and genomes and searched GeneMatcher and other NDD cohorts for rare and damaging monoallelic or biallelic *DHX9* variants. *Dhx9*^{-/-} mice were generated using International Mouse Phenotyping Consortium 'knockout first' targeting strategy and underwent detailed phenotypic characterization.

Results: The *DDX-DHX* gene paralogs exhibit a strong correlation between missense and loss-of-function (LoF) constraint in gnomAD. *DHX9* has the highest mutational constraint of all *DDX-DHX* genes (missense Z-score 5.84, LOEUF 0.1). Genomic analyses revealed 17 individuals with qualifying *DHX9* variants and NDDs or axonal Charcot-Marie-Tooth disease (CMT2). 16 variants were identified: 5 LoF and 11 missense including one recurrent variant (NM_001357.5):c.3497G>C; p.(Arg1166Pro). All variants were absent from gnomAD except for c.1417G>A; p.(Val473Ile) (1 heterozygote) in an individual with mild NDD. As *de novo* mutations are a major driver of NDDs, we performed family analysis if parental samples were available (12/17) and found *DHX9* variants occurred *de novo* in all cases. Missense variants clustered in domains required for helicase activity, nucleic acid binding, or nuclear localization. Quantitative HPO analysis supported nascent genotype-phenotype correlations, with LoF variants clustering with mild NDD phenotypes, whereas missense variants within the nuclear localization signal (NLS) cluster with severe NDD. *Dhx9*^{-/-} mice demonstrate neurologic abnormalities including hypoactivity in novel environments, tremor, reduced grip strength and body mass, and sensorineural hearing loss.

Conclusions: These data provide compelling evidence for *DHX9* variation as a cause of autosomal dominant (AD) NDD or CMT2, suggests null alleles result in mild neurologic phenotypes, implicates inappropriate cytoplasmic DHX9 localization as a cause of severe NDD, and dissects the role of DHX9, R-loops, HR, and double-strand DNA break repair in neurodevelopment, neurodegeneration, & genomic stability.

Mendelian Phenotypes Posters - Wednesday

PB1921. Monogenic Familial Hypercholesterolemia in the eMERGE Network: Penetrance and Associated Coronary Heart Disease Risk

Authors:

O. Dikilitas¹, A. Sherafati¹, B. A. Satterfield¹, D. Kochan¹, S. Saadatagah¹, L. Jones², E. Perez³, K. Anderson⁴, Z. Salvati², S. Hebring⁵, R. Sharp¹, A. Sturm⁶, R. Gibbs⁷, J. Linder⁴, H. Zouk⁸, W. Chung⁹, J. Peterson¹⁰, G. Jarvik¹¹, H. Rehm¹², D. Roden¹³, M. Williams¹⁴, T. Manolio¹⁵, I. Kullo¹; ¹Mayo Clinic, Rochester, MN, ²Geisinger, Danville, PA, ³Brigham and Women's Hosp., Boston, MA, ⁴Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁵Marshfield Clinic, Marshfield, WI, ⁶Geisinger, Lewis Center, OH, ⁷Baylor Coll. Med., Houston, TX, ⁸Harvard Med. Sch., Cambridge, MA, ⁹Columbia Univ., New York, NY, ¹⁰Vanderbilt Univ Med Ctr., Nashville, TN, ¹¹Univ Washington Med Ctr., Seattle, WA, ¹²Massachusetts Gen. Hosp., Boston, MA, ¹³VUMC, Nashville, TN, ¹⁴Geisinger, Holmen, WI, ¹⁵Natl. Human Genome Res. Inst., Bethesda, MD

Abstract Body:

BACKGROUND: The prevalence and penetrance of actionable pathogenic/likely pathogenic variants detected in a 'genome-first' approach are unclear. We assessed prevalence and penetrance of pathogenic familial hypercholesterolemia (FH) variants and their association with coronary heart disease (CHD) in a multicenter targeted sequencing study. **METHODS:** Adult participants (n=18,544) at 7 sites were enrolled in a prospective cohort study, as part of phase III of the eMERGE network, to study the clinical impact of returning results from targeted sequencing of 68 actionable genes, including *LDLR*, *APOB*, and *PCSK9*. FH variant prevalence and penetrance (defined as LDL-C >155 mg/dl) were estimated after excluding participants enrolled based on hypercholesterolemia. Multivariable logistic regression was used to estimate the odds of CHD compared to age & sex matched controls (5-10 for each case) without FH-associated variants. **RESULTS:** We detected FH associated variants in 160 individuals (mean age 59.81 ± 17.26, 41.3% male). The prevalence of FH-associated pathogenic variants was 1 in 188 (69 of 13,019 unselected participants). The mean LDL-C level was 320 ± 154 mg/dl, and the level was significantly higher (434 ± 133.3 mg/dl) in those with predicted loss of function (pLOF) variants in *LDLR*. Penetrance on average was 87.5%. The presence of an FH variant was associated with CHD (odds ratio [OR] 3.1, 95% confidence interval [CI] 2.1-4.7) and early onset CHD (OR 3.5, 95% CI 2.3-5.4). The magnitude of association was larger for pLOF variants in *LDLR* (OR 9.84, 95% CI 1.8-65.6). **CONCLUSION:** In a multisite cohort of EHR-linked biobanks, monogenic FH was prevalent, penetrant, and associated with the presence of CHD. These results suggest potential utility of sequencing EHR-linked biobanks to detect FH.

Mendelian Phenotypes Posters - Thursday

PB1922. Multi-omic approach identifies a novel non-coding deletion at Xq28 in a patient with X-linked primary immunodeficiency.

Authors:

D. Bonner^{1,2}, M. Frees², L. M. Tulu², A. M. Thomas¹, R. A. Ungar³, P. Goddard³, D. Moslehi¹, J. N. Kohler¹, C. M. Reuter¹, S. Marwaha¹, S. Montgomery³, P. Fisher¹, E. Ashley³, M. T. Wheeler², J. A. Bernstein¹; ¹Stanford Ctr. for Undiagnosed Diseases, Stanford, CA, ²Stanford University, Stanford, CA, ³Stanford Univ., Stanford, CA

Abstract Body:

We evaluated a 2-year-old male with a personal and family history of immunodeficiency at the Stanford Center for Undiagnosed Diseases. He initially presented with low T cell receptor excision circle levels on newborn screening. Further testing revealed near absence of recent thymic emigrants, defects of T cell proliferation and decreased natural killer cell function. He underwent hematopoietic stem cell transplantation at 9 months old. Family history is significant for 7 maternal male relatives who died from infections in childhood in a pattern consistent with X-linked inheritance. Clinical genetic testing including microarray, exome and genome sequencing (GS) and transcriptome sequencing (RNA-seq) were non-diagnostic.

Trio GS data was re-analyzed with a structural variant calling and prioritization pipeline, and results were integrated with expression outliers from trio blood RNA-seq analysis. Re-analysis prioritized a novel hemizygous maternally inherited 2.8kb deletion at Xq28 encompassing a primarily intergenic region upstream of *ARHGAP4* with partial overlap of the 3'UTR of *NAA10*. RNA-seq identified significantly reduced expression of *ARHGAP4* (z-score -8.9) and slightly reduced expression of *NAA10* (z-score -2.0) when compared to 330 controls. We considered this a promising diagnostic candidate as it was the only under-expression outlier that correlated with a rare variant on the X chromosome.

Patients reported with loss-of-function variants in *NAA10* and *ARHGAP4* do not demonstrate susceptibility to infection nor clinical immunodeficiency. A patient with a 34.4kb deletion overlapping with our patient's deletion and a phenotype of immunodeficiency and nephrogenic diabetes insipidus is reported.¹ The authors speculated that this patient's immunodeficiency was caused by loss of the intergenic region between *ARHGAP4* and *NAA10* since it is highly conserved across species and involved in cellular response to hypoxemia. Recently, this intergenic region has also been described to have a regulatory role in chromatin accessibility and transcriptional regulation.

Our patient is the second reported case with loss of the 2.8kb intergenic region between *ARHGAP4* and *NAA10* at Xq28 and primary immunodeficiency, providing further evidence for the locus-disease relationship. Research into the functional and epigenetic consequences of this deletion will clarify the mechanism of disease. Ongoing studies include segregation analysis, methylome sequencing and ATAC-sequencing.

¹Broides A, et al. Severe combined immunodeficiency associated with nephrogenic diabetes insipidus and a deletion in the Xq28 region. Clin Immunol. 2006;120(2):147-155

Mendelian Phenotypes Posters - Wednesday

PB1923. Mutational analysis of a Romanian Bardet-Biedl syndrome cohort revealed an overabundance of causal *BBS12* variants

Authors:

S. Khan^{1,2,3}, I. Focșă^{4,5}, M. Budișteanu^{6,7,8}, C. Stoica^{4,9}, F. Nedelea^{4,10}, L. Bohîlțea⁴, L. Caba¹¹, L. Butnariu^{11,12}, M. Pânzaru^{11,12}, C. Rusu^{11,12}, C. Jurcă^{13,14}, C. Bănescu¹⁵, M. Bălgrădean^{4,16}, E. Davis¹; ¹Stanley Manne Children's Res. Inst., Ann & Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, ²Human Molecular Genetics Lab., Hlth.Biotechnology Div., Natl. Inst. for Biotechnology and Genetic Engineering (NIBGE-C), Faisalabad, Pakistan, ³Pakistan Inst. of Engineering and Applied Sci. (PIEAS), Islamabad, Pakistan, ⁴Univ. of Med. and Pharmacy "Carol Davila", Bucharest, Romania, ⁵Cytogenomic Med. Lab., Bucharest, Romania, ⁶Psychiatry Res. Lab., "Prof. Dr. Alexandru Obregia" Clinical Hosp. of Psychiatry, Bucharest, Romania, ⁷Med. Genetic Lab., "Victor Babeș "Natl. Inst. of Pathology, Bucharest, Romania, ⁸Dept. of Med. Genetics, Faculty of Med., "Titu Maiorescu" Univ., Bucharest, Romania, ⁹Dept. of Pediatrics, Clinical Inst. Fundeni, Bucharest, Romania, ¹⁰Genetics Dept., Clinical Hosp. Filantropia, Bucharest, Romania, ¹¹Dept. of Med. Genetics, "Grigore T. Popa" Univ. of Med. and Pharmacy, Iași, Romania, ¹²Regional Med. Genetics Ctr., "Sf. Maria" Children's Hosp., Iași, Romania, ¹³Dept. of Genetics, Faculty of Med. and Pharmacy, Univ. of Oradea, Oradea, Romania, ¹⁴Dept. of Pediatrics, "Dr. Gavril Curteanu" Municipal Clinical Hosp., Oradea, Romania, ¹⁵"George Emil Palade" Univ. of Med., Pharmacy, Sci. and Technology, Târgu Mureș, Romania, ¹⁶Dept. of Pediatrics and Pediatric Nephrology, Emergency Clinical Hosp. for Children "Maria Skłodowska Curie", Bucharest, Romania

Abstract Body:

Bardet-Biedl syndrome (BBS), is an emblematic ciliopathy hallmarked by pleiotropy, inter- and intra-familial phenotype variability, and extensive genetic heterogeneity. BBS is a rare (~1/140,000 to ~1/160,000 in Europe) autosomal recessive pediatric disorder characterized by progressive retinal degeneration, truncal obesity, polydactyly, cognitive impairment, renal dysfunction, and hypogonadism. Twenty-seven genes implicated in the structure or function of the primary cilium have been implicated in BBS, and explain the molecular basis for ~75-80% of individuals. To investigate the mutational spectrum of BBS in Romania, we ascertained a cohort of twenty-three individuals (17 Eastern European, 5 Romani, and 1 Arab). Following informed consent, we performed proband whole exome sequencing (WES). We prioritized variants using bioinformatic filtering and ranked them according to ACMG guidelines for variant pathogenicity. We detected 17 different putative disease-causing single nucleotide variants or small insertion-deletions and two pathogenic exon disruptive copy number variants in known BBS genes in 17 pedigrees. Segregation analysis was performed in all available family members. Eleven different BBS genes harbored causal lesions, confirming the heterogeneous mutational basis of BBS. The most frequently impacted genes were *BBS12* (35%), followed by *BBS4*, *BBS7*, and *BBS10* (10% each) and *BBS1*, *BBS2*, and *BBS5* (4% each). Notably, homozygous *BBS12* p.Arg355* variants were present in seven pedigrees of both Eastern European and Romani origin (n=3 and n=4, respectively). Among the remaining families, two harbored heterozygous BBS gene changes which did not explain the phenotype, and four were bereft of a likely cause among known BBS genes, or the remainder of the exome. Three families harbored second-site variants in BBS genes. In conclusion, the diagnostic rate in our cohort was 74%, suggesting that the remaining families harbor deep intronic or large structural variants that are intractable to WES. Our data show that although the diagnostic rate of BBS in Romania is consistent with other worldwide cohorts, we observed a unique distribution of causal BBS genes, including an overrepresentation of *BBS12* due to a recurrent nonsense variant, that has important implications for regional diagnostics.

Mendelian Phenotypes Posters - Thursday

PB1924. Mutation-specific pathophysiological mechanisms of *AFF3*

Authors:

S. Bassani¹, N. Voisin¹, J. Chrast¹, A. Brusco², F. Sirchia³, L. Turban⁴, S. Schubert⁴, J. Schlump⁵, A. Rami⁴, D. DeMille⁶, P. Bayrak-toydemir⁷, G. R. Nelson⁸, K. Wong⁸, L. Duncan⁹, C. Gilissen¹⁰, L. Vissers¹¹, R. Pfundt¹², R. Kersseboom¹³, H. Hilde¹⁴, G. Å. M. Hansen¹⁴, C. Jonsrud¹⁴, M. Falkenberg Smeland¹⁵, M. Lyons¹⁶, C. M. B. Carvalho Fonseca^{17,18}, C. Zhang¹⁹, J. Lupski^{19,20,21,22}, L. Potocki^{19,22}, F. Petit²³, B. Yalcin²⁴, N. Guex²⁵, A. Reymond¹; ¹Ctr. for Integrative Genomics, Univ. of Lausanne, Lausanne, Switzerland, ²Univ. of Torino, Torino, Italy, ³Univ. of Torino, Italy, Torino, Italy, ⁴Inst. of Human Genetics, Univ. of Leipzig Med. Ctr., Leipzig, Germany, ⁵Dept. of Pediatrics, Ctr. for Neuromedicine, Gemeinschaftskrankenhaus Herdecke Gerhard-Kienle-Weg, Herdecke, Germany, ⁶Genomics Analysis 396, ARUP Lab., Salt Lake City, UT, ⁷ARUP Lab., Salt Lake City, UT, ⁸Pediatric Neurology, Univ. of Utah Sch. of Med., Salt Lake City, UT, ⁹Dept. of Pediatrics, Med. Ctr. North, Vanderbilt Univ. Med. Ctr., Nashville, TN, ¹⁰Radboud Univ. Med. Ctr., Nijmegen, Nederland, Netherlands, ¹¹Radboudumc, Nijmegen, Netherlands, ¹²Radboud Univ. Nijmegen Med. Ctr., Nijmegen, Gelderland, Netherlands, ¹³Ctr. for genetic developmental disorders, southwest Zuidwester, Middelharnis, Netherlands, ¹⁴Dept. of Med. Genetics, Univ. Hosp. of North Norway, Tromsø, Norway, ¹⁵Dept. of Paediatric Rehabilitation, Univ. Hosp. of North Norway, Tromsø, Norway, ¹⁶Greenwood Gen Ctr, N Charleston, SC, ¹⁷Pacific Northwest Res. Inst. (PNRI), Broadway, Seattle, WA, ¹⁸Dept. of Molecular and Human Genetics, BCM, Houston, TX, ¹⁹Dept. of Molecular and Human Genetics, BCM, Houston, TX, ²⁰Human Genome Sequencing Ctr., BCM, Houston, TX, ²¹Dept. of Pediatrics, BCM, Houston, TX, ²²Texas Children's Hosp., Houston, TX, ²³CHRU Lille, Lille, Nord, France, ²⁴NeuroGEMM (NeuroGenetics of Mouse Models Res. Group), INSERM U1231 GAD Lab., Tassigny, Dijon, France, ²⁵Bioinformatics Competence Ctr., Univ. of Lausanne, Lausanne, Switzerland

Abstract Body:

We previously described the KINSSHIP syndrome, an autosomal dominant disorder associated with *de novo* missense variants in the degnon of *AFF3*, a sequence important for its binding to ubiquitin ligase. The eighteen affected individuals shared a recognizable pattern of anomalies that included intellectual disability (ID), epileptic encephalopathy, mesomelic dysplasia and horseshoe kidney. Mouse knock-ins and overexpression in zebrafish suggested a dominant-negative mode of action and a pathological effect of increasing amount of *AFF3*. Additional screening of ID cohorts identified twelve individuals from eight families who carried truncation or deletions rather than missense. They presented with milder phenotypes suggesting an alternative haploinsufficiency mode of action. Corroboratingly, mouse knockouts display brain malformations with enlarged lateral ventricles and decreased corpus callosum size, neurological and skeletal anomalies including vertebrae fusion and transformation and abnormal skull shape, whereas zebrafish knockdowns presented with circling behaviour and decreased velocity indicative of a broad neurological defect. About a tenth of the fish showed also skeletal malformation, yolk sac enlargement and pericardium edema. Finally, we identified four unrelated ID patients with homozygous or compound heterozygous predicted-to-be deleterious missense variants in *AFF3*. Consistent with causatives, 3D protein modelling of these variants suggested that it cannot be accommodated without affecting the local geometry. *AFF3* encodes a component of the transcriptional super elongation complex that regulates the expression of genes involved in neurogenesis and development. Our data suggest that *AFF3* mutations result in either recessive, dominant-negative or haploinsufficiency phenotypes. Consistent with this hypothesis we found that the three types of variants differently impact cellular transcriptome.

Mendelian Phenotypes Posters - Wednesday

PB1925. Myelomeningocele: targeted search for novel variants in Mexican families.

Authors:

Q. Medina^{1,2}, L. Muñoz¹, Y. Svyryd¹, R. Rebollar², L. Luna¹, A. Aguayo¹, O. Mutchinick¹; ¹Inst. Natl. de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, ²UNAM, Mexico City, Mexico

Abstract Body:

Introduction. Neural tube closure defects are severe congenital malformations of the central nervous system, the most frequent variety being spina bifida and of this myelomeningocele (MMC). In Mexico, the prevalence of MMC is ~1/1000 live births. They have multifactorial etiology resulting from a polygenic genetic predisposition. Although some risk gene variants (RGV) have been proposed, these are not the same across populations, and the known variants have only explained part of the etiologic fraction of the defect. Next-generation sequencing (NGS) has driven rapid progress in genomic research, allowing the identification of new RGV characteristics of each population.

Objective. To investigate novel RGV by sequencing 11 Candidate Genes (CG) involved in the neural tube development in Mexican nuclear families with MMC.

Materials and Method. The 300 studied trios were recruited from 16 Teleton (CRIT) centers in Mexico. *BMP4*, *FUZ*, *GLI3*, *GPR61*, *PTCHI*, *ITGB1*, *PLXNB2*, *NCAM1*, *ALX1*, *PAX1*, and *TBXT* sequencing was performed at the Research Support Network on a NextSeq500 using an Illumina® panel designed for these CG exome sequencing. The variant calling was performed using the DNA Amplicon platform, and the variant annotation with Variant Effect Predictor. The “novel” variant was defined as a variant not registered in dbSNP, genomeAD v2, or COSV databases. In addition, Sanger sequencing was used to confirm the novel variants. Finally, if inherited from one of the parents, the confirmed ones were looked at in a healthy sibling when available.

Results. Of the 983 variants, 855 had a depth greater than 30. Of these, 299 were present only in the parents, and 556 in cases too. Of these, 78 were novel variants not registered in the reviewed databases, 15/78 were identified in coding or regulation regions, but only seven were confirmed as novel variants by Sanger sequencing. Four missense variants were found in *GLI3* (2), *ITGB1* (1), and *PLXNB2* (1). One donor splicing site variant was also found in the *ITGB1* gene. In addition, two synonymous variants in *FUZ* (1) and *PTCHI* (1) were identified. It is worth mentioning that each confirmed variant was found in 1 family out of the 300 studied; all were inherited but not transmitted to the healthy sibling.

Conclusions. Seven novel not previously reported variants were identified in the myelomeningocele cases, of which 5 have potentially pathogenic consequences in the coding regions of proteins essential in developing the neural tube: *IGBT1*, *GLI3*, and *PLXNB2*. However, functional studies are needed to confirm the risk of these variants for MMC in the Mexican population.

Mendelian Phenotypes Posters - Thursday

PB1926. New *trans*-modifier genes in the most common monogenic eye disorder, ABCA4/Stargardt disease.

Authors:

J. Zernant¹, T. Nagasaki¹, W. Lee¹, P-Y. Su¹, S. H. Tsang¹, G. A. Fishman², R. Allikmets¹; ¹Columbia Univ., New York, NY, ²The Pangere Ctr. for Inherited Retinal Diseases, The Chicago Lighthouse, Chicago, IL

Abstract Body:

Over 1,500 variants in the *ABCA4* locus cause phenotypes ranging from severe, early-onset retinal degeneration to very late-onset maculopathies. The resulting ABCA4/Stargardt disease is the most prevalent Mendelian retinal disorder, although its underlying clinical heterogeneity, including penetrance of many alleles, is not completely understood. We hypothesized that a share of this complexity is explained by variants in other, unlinked loci, i.e. *trans*-modifying genes. We sought to identify these by performing exome sequencing in a large cohort, for a rare disease, of 583 patients and compared variation in 285 genes that have been associated with simple or syndromic retinopathies, to cohorts of ethnically matched controls. We also compared the allele frequencies between three ABCA4 disease subgroups. The strongest association was found with the frequent variant p.Ala96Ser in *RGS9BP*, minor allele frequency (MAF) 0.41 in the patient cohort vs. 0.32 in the control group, $p=2.404E-07$. In the late-onset ABCA4 disease subgroup, defined by the hypomorphic p.Asn1868Ile allele and including the c.4253+43G>A variant, the minor allele frequency for the *PRPH2* haplotype-tagging p.Asp338Gly SNP was 0.14 vs. 0.25 in the remaining cohort, $p=0.0027$. In the disease subgroup, defined by the most frequent pathogenic *ABCA4* variant in patients of European descent, p.Gly1961Glu, *PITPNM3* variant p.Pro17Ser presented with increased MAF=0.17, compared to other patients, MAF=0.11; $p=0.008$. Moderately rare, likely functional, variants, with MAF<0.01 and CADD>20, were enriched in *CEP290* in p.Gly1961Glu patient group, where 10.6% of 151 patients harbored a *CEP290* variant compared to 3.5% in 342 patients with other *ABCA4* pathogenic variants ($p=0.0017$). We identified three additional genes with possible modifying effect on ABCA4 disease. These findings should be replicated in other large ABCA4 disease cohorts and the specific *trans*-modifying effects of these variants/genes on the penetrance, expression and progression of ABCA4 disease have to be elucidated.

Mendelian Phenotypes Posters - Wednesday

PB1927. Novel clinical entity: Autosomal recessive missense *GPC6*-related skeletal dysplasia (ARM-GSD)

Authors:

M. Crenshaw, M. L. Meyers, J. Duis, M. Saenz, E. Elias, K. Brown, V. Slegesky, N. J. L. Meeks; Univ. of Colorado, Aurora, CO

Abstract Body:

Background: Skeletal dysplasias encompass disorders of cartilage or bone with a reported incidence of 1/5,000 (Geister, 2015). *GPC6*, located at chromosomal position 13q32, codes for protein glypican 6, which is part of the proteoglycan family that is linked to glycosylphosphatidylinositol (GPI)-anchored heparan sulfate (OMIM 604404). It assists with membrane translocation, membrane anchoring, and growth factor signal regulation (Capurro, 2017). The only described phenotype associated with this gene is autosomal recessive omodysplasia-1 (OMIM 258315), caused by loss of function variants in *GPC6* (Campos-Xavier, 2009). Patients with omodysplasia-1 have severe short stature (standard deviation range: -3 to -5) with shortening and tapering of the proximal long bones of the upper and lower extremities, creating a club-like appearance; they can also have facial hemangiomas, frontal bossing, low-set ears, flat nasal bridge, anteverted nares, motor delays, and cryptorchidism (Bayat, 2020; Elcioglu, 2004). **Case series:** We report a family with five affected out of nine total siblings who presented with an unspecified skeletal dysplasia with features distinct from omodysplasia-1. The patients' height for age ranged from the 0.18 to 9.54%ile (*Z* scores -2.91 to -1.31). Radiographically, the patients had subtle platyspondyly, bilateral coxa breva, humeri and femoral foreshortening, and ulnae minus deformity (in those with closed epiphyses). One sibling required surgical intervention for bilateral talocalcaneal fibrous coalition, four had ear abnormalities, and two had dental crowding. None of the siblings had cognitive deficits, developmental delays, facial hemangiomas, nor cryptorchidism. All affected individuals were found to have homozygous missense variants in *GPC6*, which segregated with affected individuals in the family. Carrier status was confirmed in both parents. **Conclusion:** The patients presented represent a new clinical entity that we are naming autosomal recessive missense *GPC6*-related skeletal dysplasia (ARM-GSD). The siblings differ from the omodysplasia-1 patients in that their variants are missense versus truncating or partial/full gene deletions in patients with omodysplasia-1. The patients presented do not have as profound short stature nor club-like bone appearance, and they have subtle radiographic findings of the spine and all extremities. Given the molecular and phenotypic differences in patients with omodysplasia-1 and the patients presented, we propose a new skeletal dysplasia syndrome associated with biallelic missense variants in *GPC6*.

Mendelian Phenotypes Posters - Thursday

PB1928. Novel homozygous nonsense mutation of *MLIP* and compensatory alternative splicing in a late-onset distal myopathy.

Authors:

J. Mezreani^{1,2}, S. Audet^{1,2}, F. Martin¹, J. Charbonneau¹, V. Triassi^{1,3}, E. Bareke¹, A. Laplante¹, J. Karamchandani⁴, R. Massie⁴, C. H. Chalk⁴, E. K. O'Ferrall⁴, M. Tetreault^{1,2,3}; ¹CRCHUM, Montreal, QC, Canada, ²Dept. of NeuroSci.s, Univ. of Montreal, Montreal, QC, Canada, ³Dept. of Bioinformatics, Univ. of Montreal, Montreal, QC, Canada, ⁴Montreal Neurological Inst., Montreal, QC, Canada

Abstract Body:

Efficiently establishing diagnoses for myopathy-related patients is a complex task due to the important clinical heterogeneity surrounding these pathologies, and concertedly, their multiple overlapping traits. This highlights the need for implementation of more robust diagnostic methods, such as clinical sequencing, as well as the advancement of research, furthering our understanding and the literature encompassing this group of disease. Despite the growing accessibility of clinical sequencing, functional interpretation of candidate variants remains a major hurdle to molecular diagnostics.

In this study, we aim to describe a new adult-onset myopathy with muscle weakness and hyperCKemia, for which we uncovered a novel homozygous nonsense variant in Muscular *LMNA*-Interacting Protein (*MLIP*). Following inconclusive gene panel testing, RNA-sequencing served as a diagnostic tool for this complex case. On top of variant identification, differential expression analysis revealed a significant downregulation of this gene, which had a surprisingly mild effect on *MLIP* protein expression, as observed through Western Blot. RT-PCR and long-read sequencing (LRS) both support an important transcriptome shift in the patient, where decreased *MLIP* levels are seemingly due to nonsense-mediated decay of transcripts containing the exon 5 mutation. Moreover, a compensatory mechanism upregulates the functionally lacking isoforms, and generates novel transcripts. We suspect this irregular molecular event to play a role in the phenotypes of the patient, for whom an earlier-onset of pathology would have been expected.

To complement this study, an *in vitro* myoblast model, designed with CRISPR/Cas9 genome editing, has been developed by our team. The disease-relevant model will allow further assessment of the functional impact of the variant on a cellular level. In order to better define the molecular role as well as the pathogenic effect of *MLIP*, we aim to investigate multiple processes, such as the proliferation rate, differentiation aptitude, protein localization and partner interactions.

These findings support the recently discovered clinical implications of *MLIP* variants in a pediatric myopathy, highlighting for the first time its relevance in adult-onset cases. These results also underline the power of LRS as a tool for the functional assessment of variants of unknown significance, as well as the definition of accurate isoform profile annotations in a tissue specific manner. Finally, we hope our work will greatly benefit *MLIP* literature in a broader fashion.

Mendelian Phenotypes Posters - Wednesday

PB1929. Novel homozygous variant in *TFAP2B* in a colombian family with Char syndrome and distinctive neurobehavioral features.

Authors:

I. Bernal¹, F. Sir-Mendoza², S. MARADEI ANAYA³; ¹Biotecgen S.A.S, Bogota, Colombia, ²Biotecgen S.A.S, Bogotá, Colombia, ³Biotecgen S.A.S, BOGOTÁ, Colombia

Abstract Body:

Char syndrome is a rare genetic condition with a widely variable phenotype, characterized by unusual faces, patent ductus arteriosus (PDA) and absent or hypoplastic phalanges of the fifth finger. There is no evidence of association with neurodevelopmental involvement such as intellectual disability or autism spectrum disorders. Char syndrome is caused by heterozygous pathogenic variants in the *TFAP2B* gene in an autosomal dominant manner. We described a 13-year-old boy with sleep disorder, multiple craniofacial and limb anomalies, patent ductus arteriosus and Asperger syndrome. He has 11 siblings, three of them with suspicion of an autism spectrum disorder, five with PDA, and six with minor facial anomalies. Additionally, craniofacial anomalies, PDA, polythelia and behavioral disorders were identified in the mother. Trio exome sequencing was performed for the proband, identifying a heterozygous pathogenic variant (NM_003221.4):c.218C>G p.(Pro73Arg), in the *TFAP2B* gene, inherited from the mother. Char syndrome diagnosis was considered for the proband, so we performed an analysis of the variant in other family members, using Sanger sequencing. Four maternal generations were tested, identifying the pathogenic variant in a heterozygous state in five siblings of the proband, his grandfather, one aunt and her two sons, a great aunt and her daughter, and interestingly, we found it to be present in a homozygous state in one sibling of the proband. In conclusion, pathogenic variant c.218C>G p.(Pro73Arg), in the *TFAP2B* gene, was identified in 13 individuals of the same colombian family with a wide variety of signs and symptoms. Neurological or behavioral impairment was observed in all affected patients, ranging from oppositional defiant disorder to Asperger syndrome, findings not previously associated with Char syndrome. The homozygous individual presents a more severe craniofacial deformity and neurological involvement, compared to those heterozygotes, but cardiac and skeletal involvement is similar in both scenarios. To our knowledge, c.218C>G p.(Pro73Arg) variant has not been previously reported in the literature. Although it would be necessary to perform functional studies to better understand if this variant accounts for the observed neurocognitive and behavioral phenotypic landscapes, having found a significant number of affected individuals in the same family manifesting these signs/symptoms is highly suggestive of a possible independent phenotype-genotype correlation, whether it could be part of an alternate phenotype of Char Syndrome or even a different unreported medical condition.

Mendelian Phenotypes Posters - Wednesday

PB1931. Novel locus identification for primary congenital glaucoma in a Pakistani family

Authors:

H. Khan¹, M. Azam², F. Akhtar³, S. Ali², R. Qamar², H. Ayub¹; ¹Pak Austria Fachhochschule: Inst. of Applied Sci. and Technology, Haripur, Pakistan, ²COMSATS Univ. Islamabad, Islamabad, Pakistan, ³Pakistan Inst. of Ophthalmology, Al-Shifa Trust Eye Hosp., Rawalpindi, Pakistan

Abstract Body:

Primary congenital glaucoma (PCG) is an autosomal recessive disorder of the trabecular meshwork & anterior chamber angle of the eye, accounting for up to 18% of childhood blindness around the globe. The prevalence of PCG is continuously rising in Pakistani population, which is attributed to the higher rate of consanguineous marriages. A PCG family of Pakistani origin with two affected & five healthy individuals was studied using different molecular genetic techniques & bioinformatic tools to identify the genetic basis of the disorder in the current family. A 250K array analysis, CA marker analysis & whole exome sequencing was done for the family to identify the disease causative mutation. Selected identified exome variants were subjected to segregation analysis. Though none of the variants segregated with the disease phenotype among the family members but homozygosity mapping revealed a 2 Mb homozygous region common among the affected individuals on chromosome 7 (7q34). CA marker analysis confirmed the segregation of the novel locus with the disease in the family. Thus, in the current study we report the identification of a 2 Mb plausible locus 7q34 to be associated with congenital glaucoma in a Pakistani family.

Mendelian Phenotypes Posters - Thursday

PB1932. Novel variants expand understanding of *PPP1R13L* - related cardio-cutaneous syndrome

Authors:

C. Close¹, K. Anyane-Yeboa²; ¹Div. of Clinical Genetics, Dept. of Pediatrics, Columbia Univ. Med. Ctr., New York City, NY, ²Div. of Clinical Genetics, Dept. of Pediatrics, Columbia Univ. Med. Ctr., New York, NY

Abstract Body:

Pathogenic variants in *PPP1R13L*, encoding iASPP protein, are responsible for an autosomal-recessive cardio-cutaneous syndrome. Previous studies have shown that iASPP function is essential to desmosome regulation and loss of iASPP contributes to a cardio-cutaneous phenotype.

To date, there are 10 families affected by biallelic variants in *PPP1R13L* reported in the literature. Affected individuals have dilated cardiomyopathy (DCM), ichthyosis-like skin findings, woolly hair, dystrophic nails, and wedge-shaped teeth.

Here, we describe a family with atypical features of *PPP1R13L* - related cardio-cutaneous syndrome caused by novel variants. Our patient is an 18-year-old male twin who was born with anal atresia, diagnosed with DCM at age 2, received heart transplant at age 3, and developed steatocystoma papules (SP) at age 12. His twin brother was also born with anal atresia and is identically affected by DCM and SP. They are assumed to be monozygotic. They were born to non-consanguineous parents of Caribbean descent had a sister who died at 21 months from idiopathic cardiac fibrosis.

To date, variants in *KRT17* are the only known genetic cause of steatocystoma multiplex. Sequencing and deletion/duplication analysis of *KRT17* did not reveal any pathogenic variants in our patient. SNP microarray was negative but exome sequencing (ES) with samples from our patient and his mother revealed compound heterozygosity of the variants p.D325Gfs*69 (c.974_1008del) and p.W799* (c.2397 G > A) in *PPP1R13L*. ES did not identify any variants related to anal atresia or cardiomyopathy. The frameshift variant, p.D325Gfs*69, was maternally inherited and is classified as a pathogenic variant as per ACMG recommendations. It is predicted to result in protein truncation or nonsense-mediated decay causing loss of function. The nonsense variant, p.W799*, was not present in the maternal sample and is likely in trans. This variant is interpreted as a likely pathogenic variant as per ACMG recommendations as it is predicted to result in protein truncation. Neither of these variants have been previously reported to our knowledge. Upon receiving these results, genetic testing was recommended for our patient's twin brother and results are pending.

This case marks the first report of SP and anal atresia in a patient with *PPP1R13L* variants. It expands our understanding of *PPP1R13L* function as it suggests a broader phenotypic spectrum than previously appreciated. *PPP1R13L* should be considered when evaluating individuals with SP and conditions known to be caused by abnormal intercellular structures such as DCM and atresia.

Mendelian Phenotypes Posters - Wednesday

PB1933. Novel variants in the *PKD1* and *PKD2* genes in German patients with autosomal dominant polycystic kidney disease

Authors:

R. Zarbock, K. Mayer, M. Scholz, J. Philippou-Massier, K. Hörtnagel, I. Rost; Med.ver Genetics, Ctr. for Human Genetics and Lab. Med. Dr. Klein, Dr. Rost and collea, Martinsried, Germany

Abstract Body:

Background: Autosomal dominant polycystic kidney disease (ADPKD) is a potentially fatal genetic kidney disease with a prevalence of 1:1000. ADPKD accounts for a significant proportion of end-stage renal disease. Most ADPKD patients carry a causative sequence variant in the *PKD1* or *PKD2* gene. The molecular diagnosis of ADPKD is hampered by high allelic heterogeneity and the presence of six highly homologous *PKD1* pseudogenes. To date, the ADPKD variant database lists 1225 different pathogenic variants in *PKD1* and 196 in *PKD2*. In addition, causative variants in *GANAB* and *DNAJB11* have been discovered in rare cases.

Methods: Over a 36-month period, DNA from 119 individuals (61 men, 58 women) with suspected ADPKD was analyzed as part of routine diagnostic testing. Detection of variants was performed by targeted next-generation sequencing of ADPKD genes and, in case of *PKD1*, also by long-range PCR followed by Sanger sequencing. Copy number variations were detected using NGS data and, for *PKD1* and *PKD2*, also by multiplex ligation-dependent probe amplification (MLPA).

Results: A total of 73 variants were identified in 119 patients, corresponding to a detection rate of 61.3%. 65 variants were identified in *PKD1* (89.0%) and 8 in *PKD2* (11.0%). These include 39 novel variants detected in *PKD1* and one novel variant in *PKD2*. No variants were detected in the other ADPKD-associated genes. 40 variants (1 missense, 12 nonsense, 19 frameshift, 5 splice site, and 3 large deletions) were classified as pathogenic according to ACMG guidelines. An additional 12 variants were rated as likely pathogenic (9 missense, 2 splice site, and 1 in-frame deletion). The remaining 21 variants (20 missense and 1 in-frame deletion) were classified as variants of uncertain significance. Truncating variants thus accounted for 52% of all sequence changes found. While the presence of pseudogenes may interfere with NGS, all *PKD1* variants identified by Sanger sequencing were also detected by NGS. However, Sanger sequencing was essential because especially exons 1 and 42 of *PKD1* had poor coverage in some samples.

Conclusions: Our results highlight the high allelic heterogeneity of variants, which is complicated by the presence of variants of uncertain significance. The discovery of 40 previously undescribed variants has significantly expanded the spectrum of known ADPKD causative variants.

Mendelian Phenotypes Posters - Thursday

PB1934. One family, two disorders, three affected: a diagnostic odyssey highlighting the importance of WES as first-tier testing in patients with a specific phenotype but a consanguineous background.

Authors:

N. Alsaleh^{1,2}, M. Alfadhel^{1,3,2}, M. Abothnain^{1,2}, W. Altuwaijri^{4,3,2}, W. Eyaid^{1,3,2}; ¹Genetics and Precision Med. Dept., King Abdullah Specialized Children's Hosp., King Abdulaziz Med. City, Ministry of Natl. Guard-Hlth.Affairs (NGHA), Riyadh, Saudi Arabia, ²King Abdullah Intl. Med. Res. Ctr. (KAIMRC), Riyadh, Saudi Arabia, ³King Saud bin Abdulaziz Univ. for Hlth.Sci., Ministry of Natl. Guard-Hlth.Affairs (NGHA), Riyadh, Saudi Arabia, ⁴Div. of Neurology, Dept. of Pediatrics, King Abdullah Specialized Children Hosp., King Abdulaziz Med. City, Ministry of Natl. Guard-Hlth.Affairs (NGHA), Riyadh, Saudi Arabia

Abstract Body:

Introduction: studies have shown that the diagnostic yield of whole-exome sequencing (WES) compared to multi-gene panels and cytogenetic studies are higher, especially in a consanguineous population. Alstrom syndrome is an autosomal recessive disorder caused by a homozygous variant in *ALMS1*. Characterized by cone-rod dystrophy, obesity, and dilated cardiomyopathy. Ullrich congenital muscular dystrophy is an autosomal recessive disorder caused by a homozygous variant in *COL6A2*. Characterized by congenital hypotonia, hyperlaxity of distal joints and contracture of proximal joints, kyphoscoliosis, and gross motor delay. Here we present a consanguineous family with three affected children and a diagnostic odyssey ended by WES. **Method:** a retrospective chart review of the three affected children was done. Parents are first cousins. Mother had a total of 7 pregnancies, two stillbirth females at term. The first affected child, a male, presented antenatally with decreased fetal movement. At birth, he was hypotonic with abnormal joint extension of the hand and feet. Later he developed gross motor delay, kyphoscoliosis, tracheostomy dependent, failure to thrive, ichthyosis, and dysmorphic features. The second affected child was a female with congenital hypotonia. She had nystagmus, dilated cardiomyopathy, obesity, global developmental delay, joint hypermobility, scoliosis, and bronchial asthma with recurrent pneumothorax. The third affected child, a female, presented with congenital hypotonia. She had global developmental delay, nystagmus, dilated cardiomyopathy, dysmorphic features, obesity, and recurrent pneumothorax. All three children had unremarkable metabolic screening labs, radiological studies, and chromosomal analysis. They all had elevated CK levels. **Results:** Chromosomal microarray for all three affected children was negative except for long stretches of homozygosity. WES for the third affected child reported a homozygous likely pathogenic novel variant in *ALMS1*. Later WES was requested for the first affected child and it reported a homozygous pathogenic variant in *COL6A2*. Targeted gene testing for the second affected child is pending for both *COL6A2* and *ALMS1* variants. **Conclusion:** WES's current availability and somewhat quick turnaround time significantly impacted reaching a diagnosis in a timely manner. Variable expressivity is known in several genetic disorders; however, the presence of a more severe phenotype, especially in patients with consanguineous backgrounds, could be due to a second genetic disorder. Therefore, WES is the most proper first-tier testing, especially in a consanguineous population.

Mendelian Phenotypes Posters - Wednesday

PB1935*. Osteogenically differentiated human fibroblasts - an alternative model to study bone diseases

Authors:

M. Pekkinen^{1,2,3}, S. Pihlström^{4,2}, K. Määttä^{5,2}, R. Mäkitie^{5,2,6}, M. Aronen², O. Mäkitie^{5,2,3,7}; ¹Res. Program for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, University of Helsinki, Finland, ²Folkhälsan Res. Ctr., Helsinki, Finland, ³Children's Hosp., Helsinki Univ. Hosp. and Univ. of Helsinki, Helsinki, Finland, ⁴Res. gram for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, ⁵Res. Program for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, ⁶Dept. of Otorhinolaryngology – Head and Neck Surgery, Helsinki Univ. Hosp. and Univ. of Helsinki, Helsinki, Finland, ⁷Dept. of Molecular Med. and Surgery and Ctr. for Molecular Med., Karolinska Inst.t, Stockholm, Sweden

Abstract Body:

Several skeletal disorders exhibit abnormal osteoblast development and function along with mineralization defects, creating a demand for an osteoblastic *in vitro* system. However, native osteoblasts are difficult to isolate from affected patients and problematic to expand *in vitro*. Similar issue concerns bone marrow mesenchymal stem cells (MSCs), the original progenitors of osteoblasts. One potential alternative source of cells that are non-immunogenic, easily expandable and readily available through a minimal invasive harvesting procedure are human dermal fibroblasts. Therefore, we developed an *in vitro* culturing technique to transdifferentiate fibroblasts into osteoblast-like cells. We obtained human fibroblasts from forearm skin biopsy, approved by an ethical committee, and differentiated them into osteoblast-like cells by treating them with β -glycerophosphate, ascorbic acid and dexamethasone. The concentrations and duration of the components varied during the osteogenic treatment. Additionally, we differentiated human commercial MSCs, derived from bone marrow, into osteoblasts and used as a positive control. The osteoblastic phenotype was verified by staining alkaline phosphatase (ALP) and calcium deposits (Alizarin red and von kossa staining), and measuring mRNA and protein levels for specific osteoblastic markers using RNA sequencing and proteomic analysis. After 14 days of treatment, both fibroblasts and MSCs stained positive for ALP together with a significant increase in bone specific ALP compared to untreated cells. At a later time point, both cell types deposited minerals, as confirmed by staining, indicating mineralization. Ingenuity Pathways Analysis of RNA sequencing data from fibroblasts and MSCs showed that the osteoarthritis pathway was activated in both cell lines. Since osteoblasts have an essential role in osteoarthritis, by e.g. producing numerous transcription factors, growth factors, and other molecules involved in osteoarthritis pathogenesis, these data are supportive of osteogenic differentiation. Together, this data indicate that our *in vitro* treatment induces osteoblast-like differentiation in fibroblasts and MSCs, producing an *in vitro* osteoblastic cell system.

Mendelian Phenotypes Posters - Thursday

PB1936. Patients with biallelic GGC repeat expansions in *NOTCH2NLC* exhibiting a typical neuronal intranuclear inclusion disease phenotype

Authors:

S. Kameyama^{1,2}, T. Mizuguchi¹, H. Doi³, S. Koyano⁴, M. Okubo³, M. Tada³, H. Shimizu⁵, H. Fukuda^{1,3}, N. Tsuchida^{1,6}, Y. Uchiyama^{1,6}, E. Koshimizu¹, K. Hamanaka¹, A. Fujita¹, K. Misawa¹, S. Miyatake^{1,7}, K. Kanai⁸, F. Tanaka³, N. Matsumoto¹; ¹Dept. of Human Genetics, Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, ²Dept. of Pathology, Keio Univ. Sch. of Med., Tokyo, Japan, ³Dept. of Neurology and Stroke Med., Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, ⁴Dept. of Neurology, Yokohama Minami Kyosai Hosp., Yokohama, Japan, ⁵Dept. of Pathology, Brain Res. Inst., Niigata Univ., Niigata, Japan, ⁶Dept. of Rare Disease Genomics, Yokohama City Univ. Hosp., Yokohama, Japan, ⁷Clinical Genetics Dept., Yokohama City Univ. Hosp., Yokohama, Japan, ⁸Dept. of Neurology, Fukushima Med. Univ. Sch. of Med., Fukushima, Japan

Abstract Body:

The genotype-phenotype correlation of *NOTCH2NLC*-related diseases may be attributed to multiple factors such as repeat lengths, epigenetic modification, and repeat motifs. However, the relationship between repeat expansion zygosity and clinical phenotype has not yet been fully elucidated. Here we report two patients with neuronal intranuclear inclusion disease (NIID) harboring the biallelic GGC repeat expansion in *NOTCH2NLC* to uncover the impact of repeat expansion zygosity on the clinical phenotype of *NOTCH2NLC*-related diseases. We recruited two Japanese patients clinically diagnosed with NIID. The genomic DNA of these patients was used for repeat-primed PCR and fluorescent amplicon length PCR (AL-PCR) to screen for the *NOTCH2NLC* repeat expansion. The zygosity of the entire *NOTCH2NLC* GGC repeat expansion and DNA methylation were comprehensively evaluated using Southern blotting and targeted long-read sequencing. As a result, in the AL-PCR, we observed amplicons suggestive of only repeat expansion alleles but not those of the wild-type allele, which indicated the presence of the biallelic repeat expansion in *NOTCH2NLC*. PacBio targeted long-read sequencing revealed that one patient harbored almost the same expanded allele containing GGC repeats in *NOTCH2NLC* ([GGC]₉₅ in the consensus sequence) without a non-expanded allele, which suggested homozygous repeat expansion. The other patient harbored two expanded alleles with different lengths of GGC repeats ([GGC]₆₅[GGA]₂[GGC]₂ or [GGC]₉₉[GGA]₂[GGC]₂ in the consensus sequences), which suggested compound heterozygous repeat expansion. The GGC repeats and the nearest CpG island were hypomethylated in all expanded alleles in both patients. Both patients harboring biallelic GGC repeat expansion showed a typical dementia-dominant NIID phenotype. In conclusion, the biallelic GGC repeat expansion in two typical NIID patients indicated that *NOTCH2NLC*-related diseases can be completely dominant.

Mendelian Phenotypes Posters - Wednesday

PB1937. Patients with primary familial brain calcification due to variants in SLC20A2 are not commonly diagnosed with Parkinson's disease

Authors:

J. B. Chang¹, C. Barnhill¹, J. L. Wenke², G. Nicolas³, H. Shen¹, S-G. Ji¹; ¹BridgeBio Pharma Inc., San Francisco, CA, ²Nashville BioSci.s, Nashville, TN, ³CHU Rouen, Univ. of Rouen Normandie, and Inserm U1245, Rouen, France

Abstract Body:

Primary familial brain calcification (PFBC) is a rare genetic disorder with incomplete penetrance. Clinical phenotypes overlap with that of Parkinson's disease (PD) in some patients. Because diagnoses for the PFBC are extremely uncommon, this raises the question of whether PFBC patients are diagnosed with PD. We sought to gain further clarity on the PFBC phenotype by analyzing patients in Genomics England, the UK Biobank, and BioVU with pathogenic variants in SLC20A2, the most common gene that causes autosomal dominant PFBC (PFBC-SLC20A2). Within a small cohort of PD patients in Genomics England, we were able to identify an enrichment of pathogenic SLC20A2 variants. However, in the UK Biobank, we did not detect a similar association with PD phenotypes. Instead, we did detect an association between patients with pathogenic SLC20A2 variants ($n = 49$) and giant cell arteritis (GCA, $p = 1.2e-5$), a disease that has been linked to calcification. This association was corroborated by a separate cohort of 102 patients with PFBC-SLC20A2, in which 1 patient also was diagnosed with GCA. Finally, within BioVU, we identified 21 PD patients who also presented with calcification in the basal ganglia. Whole exome sequencing of these patients revealed that none of them had pathogenic variants in SLC20A2 or other PFBC genes, whereas sequencing of 8 putative PFBC patients revealed 2 with pathogenic variants in SLC20A2 and another 4 in other PFBC genes. Taken together, our results show that PFBC patients are unlikely to be commonly diagnosed with PD.

Mendelian Phenotypes Posters - Thursday

PB1938. PHACE syndrome: New strategies to identify the disease mechanism for this unsolved recognizable condition

Authors:

A. Cuillerier¹, L. McDonell², T. Hartley^{3,4}, L. de Kock¹, G. Del Gobbo¹, A. White-Brown¹, S. Goobie², T. Balci⁵, L. Chad⁶, S. White⁷, M. Lines⁸, R. Mendoza-Londono⁶, C. Armour⁹, P. Au⁸, K. Kernohan¹⁰, Care4Rare Canada, D. Dymant⁴, K. Boycott⁴; ¹Children's Hosp. of Eastern Ontario Res. Inst., Ottawa, ON, Canada, ²Dalhousie Univ., Halifax, NS, Canada, ³Univ. of Ottawa, Ottawa, ON, Canada, ⁴Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, ⁵London Hlth.Sci. Ctr., London, ON, Canada, ⁶The Hosp. for Sick Children/ Univ. of Toronto, Toronto, ON, Canada, ⁷Royal's Children Hosp., Victoria, Australia, ⁸Cumming Sch. of Med. / Univ. of Calgary, Calgary, AB, Canada, ⁹Children's Hosp. of Eastern Ontario/Univ. of Ottawa, Ottawa, ON, Canada, ¹⁰Newborn Screening of Ontario / CHEO Res. Inst. (Care4Rare), Ottawa, ON, Canada

Abstract Body:

PHACE syndrome is characterized by **P**osterior fossa brain malformations, segmental facial **H**emangioma, **A**rterial anomalies, **C**ardiac defects, **E**ye anomalies, and sternal clefting or supraumbilical raphe. The hallmark of the disease is the presence of a hemangioma greater than 5cm in diameter of the head, including scalp, which usually leads physicians to a PHACE syndrome diagnosis. Although the exact incidence is unknown, over 250 cases are reported in the literature, and it is suspected that PHACE syndrome is underdiagnosed. Despite tremendous efforts by multiple research groups, the underlying etiology of PHACE syndrome has not been identified. It is hypothesized, however, that the syndrome is due to a somatic mutation, likely in a cancer genetic pathway. In order to elucidate the underlying molecular cause of PHACE syndrome, Care4Rare Canada has gathered a cohort of nine families recruited in Canada, Australia and New Zealand, including a set of discordant monozygotic twins in which one twin presents with PHACE syndrome. We have collected DNA from blood for probands and unaffected relatives, and DNA from saliva, buccal, unaffected and affected lesions from probands when possible. We have one family for which post-mortem samples of arteries, carotid, aorta, cerebellum, heart and skin were collected from the proband and another where we have fresh cardiovascular tissue. We are using multiple technologies and laboratory approaches to investigate multiple types of genetic alterations. Chromosomal microarray yielded no compelling large CNVs. Similarly, exome sequencing (ES) and short-read genome sequencing (GS) identified no compelling SNVs, CNVs, SVs or non-coding DNA variants. Some of the ES and GS data have been shared with the Centers for Mendelian Genomics and combined with hundreds of other existing datasets from PHACE patients in the hopes that a combined reanalysis will be fruitful. Functional studies on patient cell lines, including a targeted expression array and growth assays have similarly yielded no compelling candidates or pathways. Given the unrevealing nature of these previous data, we have moved on to broader approaches. Currently we are analyzing methylation, deep ES data, RNASeq, and long-read GS in multiple families using multiple tissue types, with a particular focus on the analysis of the discordant monozygotic twins. We are hopeful that our multi-omics, multi-tissue, multi-family approach will identify clues to the genetic etiology of PHACE syndrome.

Mendelian Phenotypes Posters - Wednesday

PB1939. Phenotypic Characterization of 127 Individuals with *MECP2* Duplication Syndrome.

Authors:

J. Fatih¹, M. Gandhi², C. M. Grochowski¹, H. K. Harris³, B. Suter^{4,5,6}, D. G. Glaze^{4,5,6}, J. R. Lupski^{1,5,6,7}, H. Y. Zoghbi^{1,4,5,6,8}, C. M. B. Carvalho^{1,2}, D. Pehlivan^{1,4,5,6}; ¹Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, ²Pacific Northwest Res. Inst., Seattle, WA, ³Dept. of Pediatrics, Baylor Coll. of Med. and Meyer Ctr. for Dev.al Pediatrics, Houston, TX, ⁴Section of Neurology and Dev.al NeuroSci., Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, ⁵Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, ⁶Texas Children's Hosp., Houston, TX, ⁷Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, ⁸Howard Hughes Med. Inst. and Jan and Dan Duncan Neurological Res. Inst., Houston, TX

Abstract Body:

MECP2 duplication syndrome (MDS) is a rare neurologic condition found exclusively in males. MDS arises due to increased copies of the Xq28 region including *MECP2*. To date, there are several MDS cohort studies (~50 individuals or less in each) that have reported varying frequency of clinical features but their severity was not fully investigated. In this study, we have enrolled 127 individuals with MDS whose molecular diagnoses were confirmed by routine clinical genetic studies. We performed a comprehensive phenotypic characterization of all clinical features. Fifty-six of 113 cases (50%) experienced complications during pregnancy and delivery. Postnatally, 28/109 (26%) were admitted to the NICU and their stay ranged from two days to 104 days. Congenital hypotonia, one of the defining features of MDS, was present in 64% of the individuals (70/109). Severe to profound neurodevelopmental delay was present in all subjects, with most functioning at a 12 to 18 month developmental level. The gastrointestinal system was the second most common system involved. Chewing and swallowing problems were present in 90% (101/112) and 42% of those were G-tube dependent. Constipation, often requiring suppositories and enemas, was reported in 97/110 (88%) individuals. Recurrent infections were prevalent in majority of the subjects. Pneumonia (73%, N=85/116), upper respiratory infection (79%, N=84/107) and urinary tract infections (38%, N=30/79) were the most reported conditions. Mild dysmorphic features were widely observed (90%, N=65/72). Growth parameters including head circumference were mostly within normal range, with 21.3% microcephalic, 21.3% macrocephalic and the remaining 57% normocephalic. Epilepsy, with an average age of onset of 8.4 years, was a dynamic feature and universally present in all individuals after age 15 years. Most patients' epilepsy became refractory within the first two years after onset, and it is the main cause of regression in MDS individuals. Other relatively less common and less severe clinical features include scoliosis (58%, N=45/77), insomnia (53%, N=52/98), sleep apnea (57%, N=54/94), cryptorchidism (43%, N=45/104), hearing deficit (14%, N=13/94) and eye problems, including refraction errors and strabismus (65%, N=63/97). Our study provides a deeper understanding and characterization of the clinical features in MDS subjects. These results will inform clinicians and families regarding the management of individuals with MDS and will provide an important foundation for basic and clinical research.

Mendelian Phenotypes Posters - Thursday

PB1940. Phenotypic variation of outcomes for tail-domain *KIF5A* related variants.

Authors:

M. Jose¹, M. Sikes², T. Wenger², N. Natarajan², I. Glass³; ¹Univ. of Washington, Seattle, WA, ²Seattle Children's Hosp., Seattle, WA, ³Univ of Washington, Seattle, WA

Abstract Body:

Kinesin family member 5A (*KIF5A*) encodes neural kinesin heavy chain, part of the kinesin super family proteins (KIFs), that are expressed in neurons and play a role in intracellular transport. *KIF5A* motor domain-variants are associated with hereditary spastic paraplegia (HSP) and Charcot-Marie-Tooth 2 (CMT2). *KIF5A* tail-variants predispose to amyotrophic lateral sclerosis (ALS) and neonatal intractable myoclonus (NIM). Thus far, just 3 patients with NIM have been reported. Herein, we report a neonate harboring a tail-domain pathogenic variant demonstrating phenotypic divergence for this subtype of *KIF5A* mutations.

A 6-day-old male infant born to healthy, non-consanguineous parents at 39 +weeks gestation with normal birth indices [OFC 35.5 cm (55th %ile), length of 50.5 cm (63rd %ile), birth weight of 3200g (38th %ile)] was transferred because of hypotonia and poor feeding. Initially stable in room air, intermittent hypothermia and by day 7 nasal cannular oxygen support was required, escalated to HFNC, and NGT feeding. Axial > appendicular hypotonia, upper and lower limb girdle weakness, preserved DTRs, absent startle, symmetric paucity of facial motion/eye opening, absence of fixation, bilateral ptosis, staccato unvarying vocalizations were present. Normal metabolic investigations and MR brain imaging were present. EEG showed non-specific mildly encephalopathic features. He failed the newborn hearing screen on the right ear. Prader-Willi DNA methylation confirmed biparental representation and SNP-microarray did not identify a pathogenic CNV. Emergent trio-exome sequencing revealed a heterozygous de novo tail-domain pathogenic *KIF5A* variant c.2819_2903del (p.S940Ifs*80) and a heterozygous paternally inherited pathogenic variant in *CHRNBI*.

This variant in *KIF5A* results in protein truncation and presumed LOF has not previously been observed. It is located in the C-terminal region of *KIF5A*, consistent with 3 other infants reported to have severe impacts from mutations in this domain of the protein. The 3 previously reported infants with tail-domain heterozygous variants demonstrated intractable myoclonus from birth, hypotonia, hearing loss (1/3), and 2/3 with optic nerve atrophy and developmental delay, with early demise at 3 months and 1 year in 2/3. Our case demonstrates a similar devastating neurological outcome with high dependency needs and minimal developmental progress. However, his phenotype diverges thus far as he lacks myoclonus indicating interpretation of variants in the C-terminal encoding region of *KIF5A* need not necessitate myoclonus to occur to assign pathogenicity.

Mendelian Phenotypes Posters - Wednesday

PB1941. PhenylEc: Novel genotypes and nutritional management in phenylketonuria patients from Ecuador.

Authors:

V. Romero¹, A. Campodonico¹, E. Haro¹, A. Mendoza¹, B. Bahamonde¹, J. Pozo²; ¹Univ. San Francisco de Quito, Quito, Ecuador, ²Univ. Tecnica de Cuenca, Cuenca, Ecuador

Abstract Body:

Phenylketonuria (PKU) is an innate error of metabolism due to an enzymatic deficit to convert the phenylalanine into tyrosine. Without nutritional restriction of phenylalanine foods, it results in irreversible intellectual disability. PKU is part of the neonatal screening in Ecuador, however, genetic tests and nutritional counseling are not publicly available. We characterized the molecular variants of affected individual, compared the frequency to worldwide and to other Latin American countries and modified current Guidelines from other regions to rural areas. We reviewed the medical history of Ecuadorian patients with PKU, correlated genotype-phenotype, allele frequencies from the BioPKU database and modelled the protein for the new detected genotypes. In addition, we performed a systematic literary review in English and Spanish of the nutritional databases available and use the nutritional information to develop a Latin American guideline with products from our region. The most frequent worldwide PAH variants and genotypes were not found in Ecuador. The most frequent alleles in Ecuador included p.Arg252Trp, p.Ser349Pro and c.441+5G>T. p.Arg252Trp/ c.441+5G>T genotype and p.Ser349Pro homozygote were the most frequent genotypes. We reported 5 new genotypes not reported in BioPKU database. The American and European recommendations that cannot be used in developing countries. Therefore, we obtained the nutritional information and phenylalanine and tyrosine concentrations from local products and modified the reported recommendations for developing countries. Ecuador is a country with a mixed population - African, European, and Native American. The low frequency of global variants and high frequency of rare suggest indigenous variants and unique genotype-phenotype in eastern part of South America with low frequency of European variants. This highlights the importance of developing new molecular panels for metabolic disorders to analyze and detect variants found specifically in regions of Latin America. Finally, we developed a Nutritional Guideline for Latin America with local products and more accessible recommendations demonstrating the necessity of additional projects to be done in the region.

Mendelian Phenotypes Posters - Wednesday

PB1942. Polygenic risk scores and protein language models predict phenotype severity in individuals with rare clinical pathogenic variants

Authors:

A. Wei¹, R. Border², M. Udler³, V. Ntranos⁴, N. Brandes⁵, N. Zaitlen¹, V. Arboleda¹; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²Univ. of California Los Angeles, Los Angeles, CA, ³Broad Inst., Boston, MA, ⁴Univ. of California, San Francisco, San Francisco, CA, ⁵Univ. of California, San Francisco, San Francisco, CA

Abstract Body:

Background: As more individuals have their genomes sequenced, pathogenic mutations that were previously believed to be 100% penetrant are being identified in individuals with no apparent clinical disease. For example, familial mutations can have differential clinical manifestations due to a combination of differences in genetic background and environment. The combination of how individual amino acid changes and polygenic background influences penetrance and expressivity of a clinical phenotype are unknown. This project improves upon previous studies by utilizing a cohort with ~3x as many subjects to increase power, calculating polygenic risk scores (PRSs) to study common genetic variation, and applying protein language models to improve annotation of pathogenic variants.

Methods: We analyzed whole exome-sequencing and clinical phenotypes from 200,643 individuals in UK Biobank (UKB) to identify carriers of rare, monogenic variants known to cause three metabolic phenotypes: monogenic maturity onset diabetes of the young (MODY), obesity, and hypercholesterolemia. We calculated PRSs for the associated complex traits of type 2 diabetes, body mass index (BMI), and low-density lipoprotein (LDL) levels for all UKB participants with exome sequencing. Scores predicting how single amino acid changes affected protein function were generated by the ESM1-b protein language model.

Results: The penetrance of these rare pathogenic disease variants in the UKB ranged from 20-53%.

Carriers are at higher risk of developing associated diseases than non-carriers (Bonferroni-corrected, $p < 9e-09$). Compared to non-carriers in the top 1% of common-variant polygenic risk, carriers are at similar or higher risk of developing associated diseases. Interestingly, even within the carriers of familial hypobetalipoproteinemia variants, there is a positive association between carrier LDL levels and LDL PRS and (p -value=1.36e-05) suggesting that genetic background influences the expressivity of the monogenic disease variant. Finally, we observe that the BMI of obesity variant carriers varies depending on which monogenic obesity variant is observed; gain-of-function alleles have more positive ESM1-b scores, while loss-of-function variants have more negative scores.

Conclusion: Carriers of monogenic disease rare variants have been treated uniformly as one group; however, we show that PRSs modify the penetrance and expressivity of monogenic disease, highlighting the potential integration of PRSs into genomic medicine.

Mendelian Phenotypes Posters - Wednesday

PB1943. *PPP2R5D*-related Neurodevelopmental Disorder can be inherited from a mildly affected parent.

Authors:

S. Jougheh Doust^{1,2}, E. Fox¹, R. Raja²; ¹Peterborough Regional Hlth.Ctr., Peterborough, ON, Canada, ²Lakeridge Hlth.Oshawa, Oshawa, ON, Canada

Abstract Body:

PPP2R5D-related intellectual disability (MIM#616355; ID39) also known as Jordan's syndrome is a neurodevelopmental disorder characterized by hypotonia, intellectual disability and macrocephaly. The presentation is highly variable. Mild to severe intellectual disability is a major finding. Additional features include seizures and autism. Dysmorphic features may include frontal bossing, down-slanting palpebral fissures and hypertelorism. This is an autosomal dominant condition and almost all reported cases are caused by "de novo" pathogenic mutations in this gene. Here, we report a child with *PPP2RD5*-related ID disorder who inherited the *PPP2RD5*:c.598G>A, p.Glu200Lys, from their mildly affected father. This is a recurrent pathogenic mutation which has always been reported as a de novo mutation in the affected individuals. The child has global developmental delay, macrocephaly(OFC>97th percentile), and hypotonia. Their father also has macrocephaly, mild developmental delay and learning disability. The parent and child share characteristic features including frontal bossing, hypertelorism, down-slanted palpebral fissures, low-set ears, prominent lower jaw, long philtrum, and high arched palate. This case expands our current knowledge about inheritability of this rare neurodevelopmental disorder which previously reported to be caused by de novo mutations in *PPP2RD5* gene.

Mendelian Phenotypes Posters - Thursday

PB1944*. Precision diagnostics for GLUT1 disorder using deep mutational scanning

Authors:

N. Tayebi¹, K. McCall¹, **B. Ricardo**¹, G. Haller², C. Gurnett¹; ¹Washington Univ. in St. Louis, St. Louis, MO, ²Washington Univ. Sch. of Med. in St. Louis, St. Louis, MO

Abstract Body:

Accurate genetic diagnosis is critical, particularly as effective therapies become available for neurological disorders. The accumulation of variants of uncertain significance (VUS) presents a growing crisis that negatively impact the ability to implement precision therapies. The goal of this research is to demonstrate the utility of a growth assay to quantify the functional impact of variants in *SLC2A1*, the gene responsible for GLUT1 deficiency syndrome. The GLUT1 deficiency syndrome encompasses a spectrum of neurological disorders, including early onset seizures with acquired microcephaly and cognitive impairment (classic type), paroxysmal choreoathetosis and dyskinesia, atypical childhood absence epilepsy, alternating hemiplegia, along with many related phenotypes. Because *SLC2A1* is required for growth of HAP1 cells, variants that are deleterious would drop out of the cell populations over time. By sequencing pools of cells at different time points, we can quantitatively determine the impact of variants on cell viability. As proof of principle, we evaluated 58 variants across 3 exons using a growth assay in HAP1 cells. We created a donor library consisting of 22 variants for exon 2, 18 variants for exon 3, and 18 variants for exon 7 in pathogenic, benign, and VUS categories. Variant libraries were introduced into the HAP1-lig4KO cell line using exon specific CRISPR/Cas9. Cell populations were harvested and sequenced at days 5 and 11. Nonsense and known pathogenic variants in all exons dropped out of the population, while known benign variants were not depleted. In addition, *SLC2A1* missense variants identified in the patients with childhood onset epilepsy and mild phenotype had intermediate functional effects in this assay. Our results demonstrate the utility of this assay for generating quantitative functional data for 58 variants in *SLC2A1*. Furthermore, our quantitative functional data correlated with clinical phenotypes, suggesting that it can also be clinically useful for prognosis. Future work is needed to scale the assay such that every possible variant in *SLC2A1* can be quantitatively determined. Functional assays are necessary to improve variant interpretation for neurological disorders and provide important insight into protein structure and function.

Mendelian Phenotypes Posters - Wednesday

PB1945. Pre-clinical studies in induced Pluripotent Stem Cell (iPSC) lines with *SORD* mutations linked to a recessive neuropathy

Authors:

C. Yanick¹, R. Maciel², J. Medina¹, A. Rebelo³, M. Shy⁴, S. Zuchner³, M. Saporta¹; ¹Univ. of Miami Miller Sch. of Med., Miami, FL, ²Univ. of Miami, Miami, FL, ³Univ Miami, Miami, FL, ⁴Univ. of Iowa, Iowa City, IA

Abstract Body:

Introduction: Recently, mutations in the gene coding for sorbitol dehydrogenase (*SORD*) were associated with a new form of recessive inherited neuropathy. To model this CMT type, we took fibroblast samples from patients and reprogrammed them into induced-Pluripotent Stem Cells (iPSCs). We are using these cells to perform assays to assess disease phenotypes and mechanisms as well as evaluate different therapeutic strategies for *SORD* related neuropathy. **Methods:** Fibroblasts were transduced using a Sendai Virus Reprogramming kit to express reprogramming factors. Immunocytochemistry for Oct4, NANOG, and SSEA3 was performed to confirm pluripotency of cells. Immunocytochemistry for HB9, Islet 1 and Beta-Tubulin 3 was performed to confirm differentiation into lower motor neurons. Intracellular sorbitol levels were measured via colorimetric sorbitol kit and mass spectrometry. Flow Cytometry was performed on a FACS-Canto II recording at 20,000 events per sample. **Results:** Previous experiments demonstrated that we successfully generated induced Pluripotent Stem Cell (iPSC) lines from patient fibroblasts with the *SORD* mutations and had been able to reproducibly differentiate them into *SORD* motor neurons. To ensure that we were replicating an important disease phenotype, we measured intracellular sorbitol levels of the *SORD* motor neurons and saw that the intracellular sorbitol levels were elevated compared to control. Preliminary data shows that the motor neurons can accumulate sorbitol from the extracellular environment following incubation in a 200mM sorbitol containing media, as well as release sorbitol following a 48hr withdrawal period. We assessed potential osmotic dysregulation by measuring the size of the neurons using Flow Cytometry and observed no size differences between the *SORD* and control motor neurons. Ongoing and future experiments will look at mechanisms of sorbitol accumulation and release, as well as potential neurodegenerative and mitochondrial phenotypes. **Conclusions:** Our results show that while iPSC-derived *SORD* motor neurons do not display changes in cell size, preliminary data indicates that they can either passively or actively accumulate additional sorbitol when exposed to a hypotonic environment as well as release sorbitol following return to an isotonic media and sorbitol withdrawal. Current experiments are investigating potential mechanisms of sorbitol accumulation and release using small molecule inhibitors. We are also investigating mitochondrial and neurodegenerative cellular phenotypes associated with the *SORD* neuropathy.

Mendelian Phenotypes Posters - Thursday

PB1946. Predictive modeling to define the locus heterogeneity of tRNA synthetase-related peripheral neuropathy.

Authors:

A. Cale, A. Antonellis; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Aminoacyl-tRNA synthetases (ARSs) are ubiquitously expressed, essential enzymes that ligate amino acids to cognate tRNAs in the cytoplasm and mitochondria. Mutations in five ARSs cause autosomal dominant, axonal peripheral neuropathy (i.e., Charcot-Marie-Tooth [CMT] disease). While translation and the integrated stress response have been implicated downstream of CMT-associated ARS variants, a unifying pathological mechanism that explains the locus and allelic heterogeneity has not been identified. Many findings argue against haploinsufficiency as the underlying mechanism, leaving dominant-negative and neomorphic gain-of-function mechanisms as non-mutually exclusive explanations. The majority of CMT-associated ARS alleles cause a loss-of-function effect and all CMT-associated ARSs act as cytoplasmic homodimers; this is consistent with a unifying dominant-negative mechanism. If ARS-associated CMT is caused by a dominant-negative effect, it would be expected that certain variants in any cytoplasmic, dimeric ARS could lead to dominant peripheral neuropathy. To test this, we are employing a predictive modeling strategy that begins with engineering variants in threonyl-tRNA synthetase (*TARSI*), which is a cytoplasmic dimeric ARS that has not been implicated in CMT. We are testing the following panel of *TARSI* missense variants: (a) four reported to be loss-of-function and dominantly lethal in bacteria; (b) one reported to be aminoacylation and editing defective in yeast; and (c) two variants identified in human populations that reside near residues being studied in (a). We tested these variants in a yeast complementation assay, which revealed that all are loss-of-function or hypomorphic alleles. We next tested these variants in an established yeast dominant toxicity assay, where we co-express mutant *TARSI* along with wild-type *TARSI* and assess yeast growth. This experiment revealed at least one dominantly toxic *TARSI* allele that reduces yeast cell growth in the presence of wild-type *TARSI*. Next, we will: (1) develop *C. elegans* and mouse models to assess for motor deficits associated with dominantly-toxic alleles; and (2) identify and introduce dimer-reducing mutations to determine if dimerization is required for any observed neurotoxicity, which would implicate a dominant-negative effect. Successful completion of this study will inform clinicians to screen *TARSI* for pathogenic variants in patients with peripheral neuropathy, and provide insight into the mechanism underlying ARS-associated neuropathy.

Mendelian Phenotypes Posters - Thursday

PB1947*. Probing variants associated with neurodevelopmental disorders by studying social behaviors in *C. elegans*

Authors:

J. Pierce¹, S. Sanchez²; ¹Univ. of Texas, Austin, TX, ²Univ. of Texas at Austin, Austin, TX

Abstract Body:

Neurodevelopmental disorders (ND), including Autism spectrum disorder (ASD), are often typified by defects in social behaviors and sensory integration. The wide variety and severity of ND symptoms is likely explained by mutations in hundreds of genes suspected of causing ND either singly or in combination. So far, only a handful of the genes associated with ASD have been confirmed as causing phenotypes related to ASD in animal models. Even fewer genes have been confirmed as modifying ASD phenotypes in polygenic fashion in these models. Thus, a major goal is to find a convenient way to study which of the hundreds of combinations of mutations in ASD risk genes cause or modify ASD phenotypes singly or in combination. --- To investigate the large numbers of genes and variants associated with ND, we turned to leverage the nematode *C. elegans* as a minimum *in vivo* model to quickly discover genes that cause problems in neurological conditions. We recently discovered that *C. elegans* has conserved versions for 80% of ASD risk genes on the SPARK list. *C. elegans* displays robust social behaviors when feeding and integrates diverse sensory input to make decisions. We found that genetically distinct *C. elegans* strains isolated from around the world display different degrees of social behaviors. Surprisingly, we found that the number and severity of mutations in over 100 orthologs of ASD risk genes significantly correlate with differences in social behaviors across these populations. Conversely, we found that mutations in certain genes implicated in the hypersociality in Williams syndrome correlated with increased social behavior in *C. elegans*. Moreover, we found that mutations in certain ASD risk genes cause social deficits, because we could boost social behaviors by genetically repairing defective ASD risk genes (e.g. Neuroligin, SYNGAP1, and DYRK1A). Together, our results suggest that *C. elegans* can be exploited to study the complex polygenic basis for how risk genes cause and modify phenotypes relevant to ASD and other ND.

Mendelian Phenotypes Posters - Thursday

PB1948. Proposal of intervention of the IL-10 rs1800872 (-592 C>A) as a modulator in the cardiovascular variant of Fabry disease

Authors:

A. Ruiz Ramirez^{1,2}, E. Prado Montes de Oca³, L. E. Figuera^{2,1}; ¹Doctorado en Genética Humana, CUCS, UDG, Guadalajara, Mexico, ²División de Genética, CIBO, IMSS, Guadalajara, Mexico, ³Laboratorio Natl. de Med. Personalizada (LAMPER), Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco (CIATEJ), Guadalajara, Mexico

Abstract Body:

Fabry disease (FD) is an X-linked lysosomal storage disorder, caused by mutations in the α -galactosidase A (*GLA*) gene. However, the neuro-cardiac variant of FD is not well characterized at the molecular level, and modifier genes could act as enhancers or suppressors of the severity of symptoms. A profile of increased IL-10 production is frequently seen in many FD patients, and even correlations between *IL10* variant (-819C/T) and the neurological score ($p = 0.03$) of the Mainz Severity Score Index (MSSI) have been reported. Methodology: 6 male and 7 female patients ≥ 18 years with molecularly confirmed FD were included, according to clinical characteristics and cabinet studies, they were classified in severity according to the MSSI. Patients signed informed consent before whole venous blood was withdrawn. DNA isolated was subsequently amplified, with designed primers for amplification for the region near position -592 of the promoter of *IL10* gene. DNA sequencing was performed with Applied Biosystems SeqStudio and analyzed with 4 Peaks software. Results: Of the 13 patients, 61.53% were classified as moderate (score 20-40), 31% as mild (< 20), and 7% as severe (> 40). The main cardiac alterations were short PR (23%), left ventricular hypertrophy LVH (23%), valvular insufficiency (23%), and without cardiac alterations (30%). The main neurological alteration was the presence of acroparesthesia (100%), in second place tinnitus in (76%), and third the presence of fatigue (53%). The promoter for the region near position -592 was analyzed, the presence of the rs1800872 (-592 C/A) was found in six patients. Correlations between *IL10* (-592 C/A) and both total and cardiac MSSI sub-scores ($p = 0.034$ and $p = 0.035$ respectively) were found. Preliminary conclusions: The role that variations in inflammatory and anti-inflammatory genes may be modifying clinical characteristics, has been considered in different diseases, in the case of lysosomal diseases, specifically Gaucher disease is an example where it has been continuously analyzed how modifying genes, epigenetics, and other factors can lose the clear limits between simple and complex inheritance. In FD, *IL10* had previously been associated with the neurological components of the disease, this gene could be influential in the neuro-cardiovascular phenotype and be more determinant of the severity of the disease than previously thought. IL-10 has been studied for its protective effect in periods of ischemia, where it has the ability to eliminate $[Ca^{2+}]$ induced by ischemia. This type of studies will allow a therapeutic approach depending on the genetic variants that are associated with the *GLA* mutation in the future.

Mendelian Phenotypes Posters - Wednesday

PB1949. Pulmonary function and structure in young adults with types III, IV and VI osteogenesis imperfecta.

Authors:

B. Gochuico¹, M. Hossain¹, S. Talvacchio², M. G. Zuo¹, M. Barton³, A. Dang Do², J. C. Marini²; ¹Natl. Human Genome Res. Inst., Bethesda, MD, ²Eunice Kennedy Shriver Natl. Inst. of Child Hlth. and Human Dev., Bethesda, MD, ³Natl. Heart, Lung, and Blood Inst., Bethesda, MD

Abstract Body:

Background: Osteogenesis imperfecta (OI), an inherited disorder characterized by skeletal deformities and bone fragility, is associated with lung disease due to intrinsic and extrinsic pulmonary defects. We aimed to expand the phenotype of lung disease in a cohort of children and young adults with different OI types.

Methods: Subjects with OI types III (n=8), IV (n=21), or VI (n=5) underwent pulmonary function tests (PFTs) and chest computed tomography (CT) scans at the NIH Clinical Center. For PFTs, arm span and ulnar length measurements were used as height surrogates.

Results: Mean age of subjects with OI types III, IV, or VI was 28.1, 25.2, or 14.5 years, respectively. PFT results were similar using arm spans or ulnar lengths as height surrogates. Air flow parameters were significantly lower in OI type III compared to OI type IV or VI. Regression analysis demonstrated that forced expiratory volume in 1-second (FEV1) correlated negatively with age in OI type IV. Lung volume and gas exchange (diffusion capacity) were significantly lower in OI type III compared to OI type IV or VI. Restrictive lung disease was found in 7 of 8 OI type III subjects, and half of OI type IV subjects. Most OI subjects had reduced diffusion capacity. Regression analyses showed that PFTs correlated negatively with Cobb angle in OI subjects. CT scans revealed findings in lung airways, parenchyma, and extrapulmonary tissue. The most common pulmonary finding was bronchial thickening at the level of small bronchi. CT findings including bronchial thickening (100%, 86%, 100%), atelectasis (88%, 43%, 40%), reticulations (50%, 29%, 20%), ground glass opacities (75%, 5%, 0%), pleural thickening (63%, 48%, 20%), and focal regions of emphysema (13%, 19%, 20%) were detected in OI types III, IV, or VI, respectively. Consistent with these diverse lung parenchymal findings, whole lung tissue density was variable in OI subjects.

Conclusions: Pulmonary function or lung structure abnormalities are common in young adults with OI types III, IV, or VI. Utilizing arm spans or ulnar lengths as height surrogates yields similar PFT results. In OI, air flow is reduced, and FEV1 declines with age in OI type IV. Restrictive lung disease is found in young adults with OI type III and half with OI type IV. Severity of lung function abnormalities correlates with severity of scoliosis. Bronchial thickening in small airways is highly prevalent in OI and may be directly related to abnormal collagen or a secondary inflammatory response.

Mendelian Phenotypes Posters - Thursday

PB1950. Rapid Turnover Skeletal Disease from In-frame Tandem Duplication of Exons 4-9 in *TNFRSF11A* Encoding RANK

Authors:

S. Mumm¹, F. Collins², N. Pokhrel³, S. Duan³, M. Huskey³, W. H. McAlister³, D. Sillence⁴, D. Veis¹, M. P. Whyte³; ¹Washington Univ. Sch. of Med.; Shriners Children's, St Louis, MO, ²Royal Prince Alfred Hosp. and Univ. of Sydney Med. Sch., Sydney, Australia, ³Washington Univ. Sch. of Med., St Louis, MO, ⁴Sydney Children's Hosp, Westmead, Westmead, Australia

Abstract Body:

Mendelian disorders featuring rapid bone turnover result from: i) constitutive RANK activation due to mono-allelic in-frame 12, 15, 18, or 27-bp duplication in exon 1 of *TNFRSF11A* encoding RANK's signal peptide and ii) juvenile Paget's disease (JPD) due to bi-allelic deactivating mutations in *TNFRSF11B* encoding osteoprotegerin (OPG) or a single heterozygous missense mutation in *SP7* encoding osterix (OSX). We report rapid turnover skeletal disease in a mother and fetus due to a novel duplication of *TNFRSF11A*. A 38-year-old nonconsanguineous woman wore hearing aids at age 15 months. Brittle teeth were noted at age 3 years. Loss of adult dentition began at age 9 years. At age 10 years, she was not dysmorphic and without bone pain or fractures. Bone scintigraphy showed increased uptake in the long bones, and CT revealed hypoplastic ossicles and semicircular canals, and absent cochleas. Radiographic survey revealed a normal skull, but thickened cortices of tubular bones. Serum alkaline and acid phosphatase were elevated. Tibia and ileum biopsy reportedly showed unremarkable lamellar bone, slightly increased vascularity, numerous osteoblasts, and normal marrow. At age 14 years, she had normal stature, slight knock-knee deformity, and had fractured major limb bones. Calcitonin and bisphosphonate therapy were given. At age 16 years, biopsy of her maxilla and jaw showed irregular woven bone, and marked osteoblastic and osteoclastic activity in a highly cellular fibrovascular stroma. Teeth were eroded and replaced by osteocementum and fibrous tissue. By age 27 years, she was edentulous. At age 30 years, a prominent forehead and nasal bridge, deep-set eyes, and anterior tibial bowing were apparent. Family history was negative for deafness or fractures. Then, ultrasound of her first pregnancy at 18 weeks revealed abnormal fetal bones including angulated femurs. Sanger sequencing was negative for exons and mRNA splice sites of both *TNFRSF11A/B*. However, microarray-based copy number analysis and confirmatory qPCR showed 3 copies of RANK exons 4-9 in the mother and greater than 3 in the fetus. Long-range sequencing delineated an in-frame tandem duplication predicting a RANK fusion protein of one extracellular RANKL-binding domain combined with double intracellular activation domains. Expression of the mother's variant in 293T cells confirmed larger size of RANK protein, and studies are underway to report the mechanism for this unique disorder featuring constitutive RANK activation.

Mendelian Phenotypes Posters - Wednesday

PB1951. Rare *MECP2* variant associated with the development of Rett syndrome: a case report.

Authors:

S. Crescenti, J. Black, N. Dosa, R. Lebel, A. Sakonju; SUNY Upstate Med. Univ., Syracuse, NY

Abstract Body:

Rett syndrome is a rare neurological X-linked genetic disorder involving mutations in the *MECP2* gene that typically manifest as delayed developmental progress, loss of speech and motor skills, repetitive hand movements, and seizures. We describe an eight-year-old female born by vaginal route at 41 weeks to a 21-year-old G2P1 mother and 24-year-old father. The pregnancy was uncomplicated, and the union is not known to be consanguineous. Birth weight was 3.6kg (55th percentile) and length was 53.3cm (75th percentile). Neonatal course was uncomplicated. She smiled at 3 months, rolled over at 6 months, sat alone at 7-8 months, crawled before 12 months, walked at 12 months, and first spoke words at 3.5-4 years. The patient presented with repetitive stereotypic midline hand-wringing movements, seizures, Chiari malformation type I, bilateral intermittent esotropia, ataxia, and segmental syringohydromyelia. She was referred for genetic sequencing and deletion/duplication testing which revealed a heterozygous in-frame deletion (c.918-926del) in the *MECP2* gene resulting in a three-amino acid deletion (p.Lys307-Arg309del). This specific genetic variant has not been reported in the literature but interrupts the Arg309 residue, and other variants with a disruption of this residue have been shown to be pathogenic. This case reveals a novel in-frame pathogenic *MECP2* variant and further suggests that the disruption of Arg309 plays a significant role in the pathogenesis of Rett syndrome. This patient also displays a unique set of clinical characteristics including Chiari malformation type I and syringohydromyelia in addition to typical clinical manifestations of the condition. The combination of these novelties in this case contribute to its genetic and clinical significance.

Mendelian Phenotypes Posters - Thursday

PB1952. Rare variants in syndromic ciliopathy genes as novel causes of isolated renal disease in adults

Authors:

Z. Sentell¹, L. Mougharbel², Z. Nurcombe¹, S. Babayeva², N. Anastasio³, L. Chu², M. Akpa⁴, P. R. Goodyer⁵, E. Torban⁵, T. M. Kitzler⁵; ¹McGill Univ., Montreal, QC, Canada, ²Res. Inst. of the McGill Univ. Hlth.Ctr., Montreal, QC, Canada, ³McGill Univ. Hlth.Ctr., Montreal, QC, Canada, ⁴McGill Univ, Montreal, QC, Canada, ⁵McGill Univ. Hlth.Ctr., Montreal, QC, Canada

Abstract Body:

Background: It is estimated that at least 10% of adult chronic kidney disease (CKD) is genetic. Renal ciliopathies are a frequent genetic cause of end-stage renal disease in the first three decades of life, often as part of a syndromic phenotype. However, their true genetic contribution to adult CKD may be underestimated, as some individuals present with a late-onset non-syndromic phenotype. This phenotypic variability may be due to hypomorphic variants in established ciliopathy genes that are not routinely identified on clinical testing.

Methods: By use of next-generation sequencing technologies we identified rare variants in ciliopathy genes in two adults with isolated CKD. Using functional analysis and patient-derived cell models, we demonstrate a potential causal genotype-phenotype relationship.

Results: Case 1: Compound heterozygous missense variants (p.P168L; p.T2079M) in *C2CD3* were identified in a proband with isolated CKD. Pathogenic variants in *C2CD3* cause oral-facial-digital syndrome XIV (OFD14; OMIM# 615948), but no cases of isolated renal disease were reported. Using patient-derived skin fibroblasts and urinary renal epithelial cells, we show reduced cilia length, impaired sonic hedgehog signalling, and reduced localization of C2CD3 to the basal body, when compared to healthy control cells. Transcriptional profiling suggests a splice site effect for one of the two variants. Notably, ciliogenesis appeared more severely affected in proband renal cells compared to fibroblasts, suggesting a renal-specific defect. **Case 2:** Pathogenic variants in *CC2D2A* cause at least four different syndromic ciliopathy disorders, but with no reported cases of isolated renal disease. We identified a homozygous nonsense variant (p.R34*) in *CC2D2A* in an adult patient with isolated CKD. Crucially, this variant does not affect all *CC2D2A* protein-coding transcripts. Using public data, we show that transcripts harbouring this variant are predominantly expressed in the kidney, but not in other tissues typically involved in syndromic phenotypes.

Conclusion: Rare variants in syndromic ciliopathy genes are a novel cause of isolated renal disease in adults, potentially due to hypomorphic and renal specific effects. To assess the true genetic contribution to adult CKD detailed functional analysis is required.

Mendelian Phenotypes Posters - Wednesday

PB1953. Rate of Deleterious Copy Number Variants Similar in Early Onset Psychosis and Autism Spectrum Disorders: Implications for Clinical Practice

Authors:

C. Brownstein¹, E. Douard², J. Mollon¹, R. Smith³, M. Hojlo¹, A. Das¹, M. Goldman¹, E. Garvey¹, K. Cabral⁴, J. Li¹, J. Bowen¹, A. Rao¹, C. Genetti⁴, D. Carroll¹, E. Knowles¹, E. Deaso¹, P. Agrawal⁵, A. Beggs⁶, E. D'Angelo¹, L. Almasy⁷, A. Alexander-Bloch⁸, Z. Saci⁹, C. Moreau¹⁰, G. HUGUET¹¹, A. Deo¹, D. Glahn⁶, S. Jacquemont¹², J. Gonzalez-Heydrich¹; ¹Boston Childrens Hosp., Boston, MA, ²Univ. of Montreal, Montreal, QC, Canada, ³Boston Childrens Hosp./ Harvard Med. Sch., Boston, MA, ⁴Boston Children's Hosp, Boston, MA, ⁵Boston Children s Hosp., Boston, MA, ⁶Boston Children's Hosp., Boston, MA, ⁷Children`s Hosp. of Philadelphia, Philadelphia, PA, ⁸Children`s Hosp. of Philadelphia, Philadelphia, PA, ⁹UHC Sainte-Justine Res. Ctr., Montreal, QC, Canada, ¹⁰Université de Montréal, Montreal, QC, Canada, ¹¹CHU Ste justine, Montreal, QC, Canada, ¹²CHU Sainte justine Res. Ctr., Montreal, QC, Canada

Abstract Body:

Objective: Copy number variants (CNVs) are strongly associated with neurodevelopmental and psychotic disorders. Early onset psychotic (EOP) illnesses, where symptoms appear before 18 years of age, are thought to be more strongly influenced by genetic factors than adult-onset psychotic disorders. However, the prevalence and effect of CNVs in EOP is unclear. **Methods.** We documented the prevalence of recurrent CNVs and the functional impact of deletions and duplications genome-wide in 137 children and adolescents with EOP compared to 5,540 individuals with autism spectrum disorders (ASD) and 16,504 population controls. Specifically, we compared the frequency of 47 recurrent CNVs previously associated with neurodevelopmental and neuropsychiatric illnesses in each cohort. Next, CNV risk scores (CRS), indices reflecting the dosage sensitivity for any gene across the genome that is encapsulated in a deletion or duplication separately, were compared between groups. **Results.** Prevalence of recurrent CNVs was higher in EOP than in ASD (OR=2.30, p=0.02) and controls (OR=5.06, p=3x10⁻⁵). However, the difference between the EOP and ASD was attenuated when EOP participants with co-occurring ASD were excluded. CRS was higher in the EOP group when compared to controls for both deletions (OR=1.30, p=9x10⁻⁸) and duplications (OR=1.09, p=0.02). In contrast, the EOP and ASD cohorts did not differ in terms of CRS. **Conclusions.** Given the high frequency of recurrent CNVs in EOP and comparable CRS in the EOP and ASD cohorts, our findings suggest that all children and adolescents with a psychotic diagnosis should undergo genetic screening, as is recommended in ASD.

Mendelian Phenotypes Posters - Thursday

PB1954*. RBM12 Variants Lead to a Novel Neurodevelopmental Syndrome With Atypical Genotype-Phenotype Correlations

Authors:

E. Bhoj, RBM12 Research Consortium; Children`s Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

RBM12 is part of the RNA-binding Motif Protein (RBM) gene family, which are vital for cellular functions such as apoptosis, nonsense mediated decay, and RNA splicing. The related protein RBM10 is the cause for the Mendelian TARP syndrome [OMIM 311900], which results in Talipes equinovarus, Atrial septal defect, Robin sequence, and Persistent left superior vena cava. The function of RBM12 is unknown, but truncating variants in RBM12 have been associated with a polygenic risk for psychiatric conditions and/or mild cognitive impairment. However, there is not a reported Mendelian phenotype associated with variants in *RBM12*. Through gene matchmaking programs, we have connected with 14 individuals from 12 families with heterozygous missense and LOF variants in *RBM12*. We first proposed RBM12 as a novel Mendelian syndrome in 2015, but progress was hampered by the unusual genotype-phenotype correlation that has now been recognized. In review of the patients with RBM12 variants we have connected with, there is a much more severe encephalopathic phenotype in the several patients with missense variants (despite being spread out across the gene beyond on specific domain or active area). The several patients with LOF variants have a much milder phenotype of borderline neurodevelopmental differences, and some are inherited from similarly mildly affected parents. This finding is congruent with what has been previously published on RBM12 LOF variants as a risk factor for psychiatric disease with relatively normal development. Interestingly, the distinctive overlapping facial features (low posterior hairline, epicanthus, deep set eyes, hypertelorism, thick eyebrows, and tubular nose), in both the missense and LOF patients drove the realization that this was a recognizable syndrome. We then recognized there was an enormous degree of variability of the severity of the developmental impairment, but in the opposite direction as expected, with missense variants causing more severe disease than LOF. Therefore, in addition to the description of a novel syndrome, this is an informative example of overcoming unexpected challenges to traditional genetics dogma in collecting patient cohorts through gene matching.

Mendelian Phenotypes Posters - Wednesday

PB1955. Reanalysis of whole exome sequencing data of 50 Iranian families with hereditary hearing loss

Authors:

E. Shokouhian¹, F. Jamshidi¹, M. Mohseni², K. Kahrizi², H. Najmabadi², M. Babanejad¹; ¹Genetics Res. Ctr., Univ. of Social Welfare and Rehabilitation Sci., Tehran, Iran, Tehran, Iran, Islamic Republic of, ²Genetics Res. Ctr., Univ. of Social Welfare and Rehabilitation Sci., Tehran, Iran, Islamic Republic of

Abstract Body:

To date, hearing loss (HL) remains the most common sensory disorder, and in Iran, second to intellectual disability, HL is the most common disability. Over 50% of HL cases are due to genetic factors, of which 70% are non-syndromic. Whole exome sequencing (WES), by covering all the coding regions of the human genome, has facilitated the variant identification process, especially in cases that lack sufficient clinical data. However, the innate inability of WES to read non-coding regions, bioinformatic shortcomings, and human error have limited WES diagnostic yield to 20-60%. Previous studies have shown that because novel disease-gene associations are being discovered, and bioinformatics tools and databases are being updated every year, a periodic 3-5 years follow-up can increase WES diagnostic yield by 15-30%. Here, using a variant calling pipeline with the latest versions of the GATK software package, Picard, and ANNOVAR, we reanalyzed the WES data of 50 Iranian families with hereditary HL. As a result, we were able to identify the mutations that segregated with their respective phenotypes in 10 families (20%), including 8 mutations that were initially missed due to lack of sufficient clinical data (for families with syndromic HL), and the variable expressivity of autosomal dominant hearing loss that had led to misdiagnosing a family with autosomal recessive NSHL. Re-evaluation of their phenotypes led to the identification and segregation of 5 heterozygous mutations (*PAX3*:c.199A>T, *MITF*:c.1198C>T, *MITF*: c.649C>T, *MITF*:c.1189G>T, and *MITF*:c.1069T>C) in families with Waardenburg syndrome, heterozygous *WFS1*: c.2339G>C in a family with Wolfram like syndrome, homozygous *FGF3*: c.45delT in a family with LAMM (Labyrinthine Aplasia, Microtia, and Microdontia) syndrome, and heterozygous *DIAPH1*: c.3610C>T in a family with autosomal dominant NSHL. Compound heterozygous mutations *MYO15A*:c.7788-10A>G/c.10205delC, and *MYO15A*:c.8089-14_8089-1del/c.9572G>A also segregated in two families with autosomal recessive NSHL. Interestingly, the last two families both had splice-altering variants that were initially missed because they were annotated as noncoding. In this study, by including splice prediction tools SpliceAI, MaxEntScan, and dbScSNV in our annotation step we were able to further investigate such variants. Furthermore, this study is still ongoing and more investigations are currently being done. In conclusion, we showed the importance of a comprehensive clinical evaluation for identifying any signs of syndromic deafness, and we pointed out the significance of noncoding and intronic variants that are often neglected.

Mendelian Phenotypes Posters - Thursday

PB1956. Recessive variants in *SLC9A1* cause a syndrome of cerebellar ataxia, amelogenesis imperfecta and variable sensorineural hearing loss.

Authors:

D. Pellerin¹, R. Maroofian¹, L. Fliegel², H. Houlden¹; ¹UCL Queen Square Inst. of Neurology, London, United Kingdom, ²Univ. of Alberta, Edmonton, AB, Canada

Abstract Body:

Background: The *SLC9A1* gene encodes the mammalian Na⁺/H⁺ exchanger isoform 1 (NHE1), a ubiquitously expressed membrane-bound enzyme involved in intracellular pH regulation. Ultra-rare recessive variants in *SLC9A1* have previously been described to cause autosomal recessive spinocerebellar ataxia type 19 (SCAR19) in two families. SCAR19 is characterized by early-onset ataxia and variable hearing loss. *Slc9a1* knockout mice exhibit ataxia, seizure and growth retardation. Here, we report 12 patients with recessive *SLC9A1* variants and a complex syndrome of cerebellar ataxia, amelogenesis imperfecta, developmental delay and variable sensorineural hearing loss.

Methods: Patient phenotyping was performed through serial clinical assessments, dental examinations, audiograms and brain MRI. Candidate variants in *SLC9A1* were first identified by whole-exome sequencing and next characterized *in vitro*. The expression and enzymatic activity of mutant NHE1 proteins were respectively examined by immunoblotting and transient induction with ammonium chloride of transfected NHE1-deficient cells. Intracellular targeting of mutant proteins was assessed by a combination of immunocytochemistry and cell surface biotinylation studies.

Results: We identified 12 patients belonging to eight consanguineous families with homozygous *SLC9A1* variants. Eight novel variants were discovered, including two nonsense, four missense, one frameshift and one splicing variant. Patients presented with moderate to severe cerebellar ataxia from infancy associated with cerebellar atrophy (10/11; 91%) and occasional thinning of the corpus callosum (3/11; 27%) on MRI. In addition to developmental delay, all patients exhibited amelogenesis imperfecta, which had not been previously reported with *SLC9A1* mutations. Sensorineural hearing loss of variable severity was present in nine out of 12 subjects (75%). All identified variants caused lower protein expression, reduced NHE1 enzymatic activity and protein mislocalization.

Conclusion: This study expands the mutational and phenotypic spectrum of SCAR19 and provides functional evidence for the pathogenicity of the newly identified variants. Mutations in *SLC9A1* should be specifically sought for in the presence of early-onset cerebellar ataxia and amelogenesis imperfecta.

Mendelian Phenotypes Posters - Wednesday

PB1957*. Recessive variants in the C-terminal domain of *UFSP2* are associated with severe spondyloepimetaphyseal dysplasia.

Authors:

M. Weisz-Hubshman¹, J. Andrews¹, R. Maroofian², J. Rosenfeld^{1,3}, H. Tajsharghi⁴, E. Karimiani⁵, M. Wangler¹, Y. Bae¹, B. Lee¹; ¹Baylor Coll. of Med., Houston, TX, ²Univ. Coll. London, London, United Kingdom, ³Baylor Genetics, Clinical Diagnostics Lab., Houston, TX, ⁴Sch. of Hlth.Sci., Translational Med., Univ. of Skövde, Skövde, Sweden, ⁵Genetics Res. Ctr., Molecular and Clinical Sci. Inst., St. George's, Univ. of London, London, United Kingdom

Abstract Body:

Chondrodysplasias are rare developmental disorders commonly characterized by cartilage abnormalities with progressive clinical manifestations that significantly impact the patients' quality of life. The genetic causes are variable and can affect extracellular matrix proteins, protein transport, protein processing, and post-translational modifications. Ufmylation is a ubiquitin-like post-translational modification, and its role in chondrogenesis is poorly understood. Variants in *UFSP2* and *DDRGK1*, essential components of the ufmylation process, have been found to cause chondrodysplasia. Autosomal dominant variants in *UFSP2* have been associated with Beukes hip dysplasia with incomplete penetrance and Di Rocco type spondyloepimetaphyseal dysplasia (SEMD). Autosomal recessive variants in the N-terminal domain of *UFSP2* have been recently associated with a neurodevelopmental disorder and epilepsy. Autosomal recessive variants in *DDRGK1* are associated with Shohat type SEMD. We report 3 families with homozygous variants in *UFSP2* (p.D426A, p.N459T) manifesting a severe SEMD phenotype without intellectual disability. Interestingly, all variants associated with skeletal phenotypes reside in the C-terminal domain of the *UFSP2* protein, which harbors the catalytic domain of the cysteine protease. Although the interaction between *DDRGK1* and *UFSP2* is not abrogated by the various disease associated variants *in vitro*, cellular studies suggest protein stability differences between the variants. *Drosophila* studies showed that the newly identified *UFSP2* N459T variant as well as the human wild type *UFSP2* rescue the knock out of the *drosophila* orthologue, suggesting that the variant might be a hypomorph variant. *In vivo*, chondrocyte-specific deletion of *Ddrgk1* in mice showed growth plate disorganization with shortened proliferative zone and enlarged hypertrophic zone that correlated with decreased Sox9 and Col2a1 protein levels and increased Col10a1 expression. Taken together these findings suggest an important role for the ufmylation process in cartilage, and deciphering the mechanism underlying these disorders can contribute significantly to the understanding of post-translational modifications in cartilage development and potentially serve as therapeutic target.

Mendelian Phenotypes Posters - Thursday

PB1958. Reconsidering SIX5 as a monogenic cause for Congenital Anomalies of the Kidney and Urinary Tract (CAKUT) and Branchio-oto-renal syndrome (BOR)

Authors:

J. Chang¹, G. Leone¹, A. Thompson¹, I. Patel¹, L. Woo², C-H. Wu³; ¹Case Western Reserve Univ., Cleveland, OH, ²Univ. Hosp./Case Western Reserve Univ, Cleveland, OH, ³Univ. Hosp./Case Western Reserve Univ, Cleveland Heights, OH

Abstract Body:

Background: Congenital anomalies of the kidneys and urinary tract (CAKUT) are a range of defects that occur during embryonic kidney and genitourinary tract development and are present in half of children born with chronic kidney disease. Branchio-oto-renal syndrome (BOR) is an autosomal dominant condition characterized by branchial defects, otic defects, renal abnormalities, and CAKUT. Both CAKUT and BOR are portrayed with incomplete penetrance and variable expressivity. The EYA1 and SIX5 pathogenic variants have been established as monogenic causes of BOR by Hoskins et al. based on a 2007 study. However, Krug et al. presented their BOR cohort in 2011, and the SIX5 variant was rarely seen and co-existed with EYA1 deletions. Thus, Krug et al. questioned if SIX5 is a causal gene for BOR. We present one case of CAKUT/BOR with a SIX5 variant c.1655C>T (p.T552M) and no EYA1 variant, identified by target resequencing, and lean toward supporting SIX5 pathogenic variant as a monogenic cause of CAKUT/BOR.

Case Presentation:The patient is a 27-year-old female referred to genetics with a past medical history of left solitary kidney, vesicoureteral reflux, VATER syndrome, scoliosis, anal atresia, and chronic kidney disease. Additionally, mild to moderate sensorineural hearing loss was noted in her right ear in 2008. The patient's congenital anomalies are reminiscent of BOR syndrome. Genetic evaluation was performed with target resequencing with a panel of CAKUT genes which identified an HPSE2 variant of c.1768C>A (p.Arg590Ser) heterozygously and a SIX5 variant of c.1655C>T (p.Thr552Met) heterozygously.

Discussion:This SIX5 c.1655C>T (p.T552M) was reported as a causal gene and variant for CAKUT and BOR by Hoskins et al. with a functional study in 2007. Here we report another occurrence of the SIX5 variant c.1655C>T (p.T552M), in conjunction with a wild-type EYA1, occurring in a patient with signs of CAKUT and BOR. We would like to refamiliarize this variant of SIX5 to the medical community again as a potential causal gene and a pathogenic variant of CAKUT and BOR syndrome. Further consideration and establishment of SIX5 variants may be warranted to address the current discrepancy within literature.

Mendelian Phenotypes Posters - Wednesday

PB1959. Report of a novel *LAMB3* variant and an overview of *LAMB3* variants associated with junctional epidermolysis bullosa

Authors:

E. Tan¹, M. J. A. Koh²; ¹KK Women's & Children's Hosp., Singapore, Singapore, ²KK Women's & Children's Hosp., Singapore, Singapore

Abstract Body:

Background: Epidermolysis bullosa (EB) is a group of heterogeneous disorders with considerable overlap in the clinical presentations between the different types. Accurate diagnosis is particularly challenging in younger patients in whom the clinical features have not fully manifested. To date, at least 20 genes have been shown to cause skin fragility. **Case presentation:** The patient is a 26-year-old Chinese female from a non-consanguineous marriage with no family history of blistering skin disorders. She had blisters and erosions at sites of trauma since infancy, involving mainly her trunk and limbs. Her teeth and nails were normal. She was clinically diagnosed as EB Simplex (EBS) but continued to have moderate blistering throughout her childhood, adolescence, and adult life, and was treated with judicious wound care with non-adherent wound dressings. She married and went on to have a normal offspring. **Results:** Sequencing of her germline DNA revealed a heterozygous missense substitution (c.1064G>A:p.Cys355Tyr) affecting a highly conserved amino acid residue, and a single-nucleotide deletion (c.2116del: p.Ile706Ter) that would lead to premature termination of translation. The missense variant is novel while the deletion variant has a low frequency of 0.011 in East Asians. The discovery of two pathogenic variants in *LAMB3* and her mild presentations indicate that she has the generalized intermediate type (non-Herlitz) type of junctional EB (JEB). **Discussion:** JEB is the rarest form of EB caused by biallelic mutations in *LAMB3*. There are 129 unique damaging mutations for *LAMB3* in the Human Gene Mutation Database (HGMD) that are single-nucleotide variants or small insertions/deletions that are associated with JEB. We present the analysis of 157 unrelated reported cases for whom the identity of both alleles and their JEB subtype are known, and correlate their genotypes with the clinical presentations of the different types of JEB.

Mendelian Phenotypes Posters - Thursday

PB1960. Report of *PLP1* Gene Mutation in Pelizaeus-Merzbacher Disease

Authors:

S. Seyedhassani¹, M. Kadkhodazadeh¹, R. Fallah², L. Najafi³; ¹Dr. Seyedhassani Med. Genetic Ctr., Yazd, Yazd, Iran, Islamic Republic of, ²Yazd Shahid Sadughi Med. Univ., Yazd, Yazd, Iran, Islamic Republic of, ³Tehran Azad Univ., Tehran, Iran, Islamic Republic of

Abstract Body:

Introduction: Proteolipid protein I (*PLP1*) gene encodes a transmembrane proteolipid protein is the primary constituent of myelin. This protein may play a role in the compaction, stabilization and development of myelin sheaths, oligodendrocyte and axon. The PLP1-related disorders have included Pelizaeus-Merzbacher disease(PMD) and spastic paraplegia type 2. (PMD) is a progressive and degenerative central nervous system disorder and inherited as X-linked recessive pattern and classify into two types. Their phenotypes can overlap. It is characterized by developmental delay, rotary nystagmus, psychomotor delay, dysarthria, seizures and mental retardation **Case report:** A 2.5 years old boy was referred to genetic center. He was born from non-consanguineous parents and suffering from mental retardation , quadriplegia and developmental delay. He had history of the recurrent seizures since the age of 40 days. Also, he had two brothers with same clinical manifestations that were dead about 3.5 years. **Material and methods:** We did whole exome sequencing and then sanger sequencing for the confirmation of findings. **Result:** A hemizygous c.T137C missense mutation was detected at exon 3 in *PLP1* gene. Confirmation of WES finding showed normal and heterozygous conditions in father and mother respectively. Also, genotype- phenotype correlation was done and the disease confirmed. **Conclusion:** PMD is characterized clinically by nystagmus, spastic quadriplegia, ataxia, and developmental delay. This mutation change the amino acid, leucine to proline and can affect proteolipid protein I structure. Main expression of this gene is in brain. Therefore, cerebral signs and symptoms are significant.

Mendelian Phenotypes Posters - Wednesday

PB1961. Rewiring our understanding of the enteric nervous system: Identification of diverse developmental lineages of enteric neurons and cell-state specific roles in maintenance across aging.

Authors:

J. Slosberg, A. Singh, S. Nagaraj, S. Kulkarni, L. Goff; Johns Hopkins Univ., Baltimore, MD

Abstract Body:

Within the enteric nervous system (ENS), the current understanding of cellular roles and developmental origins has guided the design and interpretation of targeted experiments. However, a comprehensive understanding of the various cell types that make up the postnatal ENS and a consensus about their roles in maintaining proper gut function has not been achieved. This is especially true across conditions and disease-states such as aging. Prior studies in this field have suffered from a bias either due to an assumption of a single developmental lineage for all adult ENS neurons, or a preconceived transcriptional identity to annotate ENS neurons. Thus, by performing single cell transcriptomics on all cells of the murine gut wall, we aimed to conduct an unbiased survey of the diversity of cellular transcriptional states that comprise the murine ENS. During this process, we have identified a previously mischaracterized population of post-natal born enteric neurons derived from the mesoderm. We performed unbiased 10X Genomics-based scRNA-seq analyses of cells from the longitudinal muscle and myenteric plexus (LM-MP) tissue of the gut wall, which contains a large population of ENS cells. We collected tissue from mice of three postnatal ages of Days 10, 20, and 60, to identify three major ENS populations - neural crest-derived neurons (NC-neurons), neural crest-derived glial cells (Neuroglia), and the population of mesoderm-derived neurons (MENS). We validated cell type-specific markers from our scRNA-seq reference by immunohistochemistry and single molecule RNA-FISH in neural crest lineage fate mapping mice. Next, we used unsupervised non-negative matrix factorization to identify gene expression signatures that represent the diversity of cell states in the murine ENS. We then applied a pattern-based projection analysis using the Human Gut Cell Atlas, highlighting the presence of MENS' transcriptional signatures in human gut cells and their relative increase with age. To test the relevance of these distinct ENS cells to human disease, we used an agnostic approach to test pattern- and cell-specific gene expression signatures for their disease-association in summary statistics from more than 80 GWAS studies of relevant phenotypes. By using an unsupervised approach via single cell transcriptomics to correctly identify the true transcriptional and developmental heterogeneity in an important neural tissue, our study provides a framework that can be implemented across multiple tissue and organ types to study the relevance of specific cells and their molecular processes in states of health and disease.

Mendelian Phenotypes Posters - Thursday

PB1962*. *RFC1* expansions in Italian population: Molecular analysis, structure/size determination and clinical features of patients with CANVAS/spectrum phenotypes.

Authors:

D. Di Bella¹, S. Magri¹, E. Sarto¹, M. Corbetta¹, M. Balzo¹, C. Gelleri¹, C. Pisciotto², E. Salsano², D. Pareyson², M. Fichera¹, L. Nanetti¹, C. Mariotti¹, F. Taroni¹; ¹Fondazione IRCCS Istituto Neurologico Carlo Besta, Unit of Med. Genetics and Neurogenetics, Milan, Italy, ²Fondazione IRCCS Istituto Neurologico Carlo Besta, Unit of Rare Neurodegenerative and Neurometabolic Diseases, Milan, Italy

Abstract Body:

A biallelic intronic AAGGG expansion in the *RFC1* gene has been identified as the cause of CANVAS, an adult-onset, slowly progressive neurodegenerative disorder characterised in its full form by a combination of neuropathy, ataxia and vestibular disease. Recently, expansions with alternative sequence conformation have been identified whose pathogenic role remains to be established. This study was aimed: 1) to screen Italian patients with full or partial CANVAS phenotype (n=279) or late-onset (n=216) spinocerebellar ataxia for *RFC1* repeat expansion; 2) to assess the frequency of heterozygous carriers of *RFC1* expansion in a healthy control population; 3) to sequence the expanded region with a long-read NGS approach (LRseq) in order to characterize the composition and length of different repeat conformations and to assess their pathogenic role. The *RFC1* locus was analysed by a PCR encompassing the repeat, followed by fluorescent repeat-primed PCRs (Cortese et al, 2019) detecting different repeat motifs. LRseq of expanded alleles was set up on Oxford Nanopore Technologies (ONT) device, using Cas9-based enrichment of the expanded region (Nakamura et al, 2020). Biallelic *RFC1* expansion was identified in 29% (81/279) probands with a CANVAS/spectrum phenotype, and in 12.5% (27/216) of patients with ataxia. Heterozygous carrier rate was 12% in patients (24/198) and 10% (36/360) in controls. Approx. 5% of alleles showed complex repeat conformations with uncertain pathogenicity. Biallelic *RFC1* expansion accounts for a large proportion of ataxia phenotypes in Italian patients (12-29%) and should always be considered in the diagnostic algorithm of patients with sporadic ataxia, ranking first in late-onset cerebellar and sensory ataxia. Given the high heterozygous carrier rate, *RFC1* expansion might be the most common cause of ataxia in Italian patients. Nanopore LRseq data confirm the complex architecture of the expanded repeat which may have crucial implications for diagnosis and counseling.

References: Cortese A (2019) Nature Genetics 51:649-658; Nakamura, H. et al. (2020) JHG, 65(5), pp. 475-480.

Grants: FRRB-CP-20/2018

Mendelian Phenotypes Posters - Wednesday

PB1963. RNF220, RING finger E3 ubiquitin ligase, is required for the generation of p2-derived V2 interneurons and its loss-of-function causes an interneuronopathy with a broad spectrum of neurodevelopmental phenotypes

Authors:

H-G. Kim¹, T-I. Choi², S. Tompson³, J. Mokry⁴, B. Leheup⁵, C. Gordon⁶, D. Stevenson⁷, E. Banne⁸, G. Houge⁹, M. Kukolich¹⁰, C-H. Kim¹¹; ¹Qatar BioMed. Res. Inst., Doha, Qatar, ²Dept. of Biology, Chungnam Natl. Univ., Daejeon,, Korea, Republic of, ³Dept. of Ophthalmology and Visual Sci., WI, WI, ⁴Baylor Coll. of Med., Houston, TX, ⁵Génétique Médicale à la Faculté de Médecine de Nancy Université de Lorraine, Lorraine, France, ⁶Embryology and Genetics of Human Malformations, Inst. Imagine,, Paris, France, ⁷Univ of Utah, Salt Lake City, UT, ⁸Edith Wolfson Med. Ctr., Holon, Israel, ⁹Dep Med. genetics, Haukeland Univ Hosp., Bergen, Norway, ¹⁰Cook Children's Hosp, Fort Worth, TX, ¹¹Dept. of Biology, Chungnam Natl. Univ., Daejeon,, Korea, Republic of

Abstract Body:

RNF220, a zinc RING finger gene, was identified as a binding partner of ZC4H2, which is a causative gene for Miles-Carpenter syndrome (MCS). MCS is an X-linked intellectual disability (XLID) syndrome characterized by digit contractures, mild spasticity, exotropia along with intellectual disability. As an E3 ubiquitin ligase, RNF220 also bridges the interaction between USP7 and beta-catenin, both of which are mutated in neurodevelopmental disorders. USP7 is an integral component of ubiquitin ligase complex, whereas beta-catenin, encoded by CTNNB1, is the central downstream player in the canonical WNT signaling pathway. Despite these direct physical interactions of RNF220 with the threesome involved in neurodevelopmental disorder, the pathological role of RNF220 in human disease remained unexplored. Here we demonstrate that the knockout of RNF220 in zebrafish exhibits behavioral and ocular phenotypes resulting in the complete loss of V2 interneurons and P2 progenitor domains in the spinal cord and hindbrain. Most persuasively, we identified a dozen heterozygous loss-of-function mutations of *RNF220* in human patients presenting with cognitive impairment and motor defects in gait, speech, and eye movements. These results support the crucial role for RNF220 in the generation of V2 interneurons and this loss-of-function results in a new interneuronopathy with a broad spectrum of neurodevelopmental features.

Mendelian Phenotypes Posters - Thursday

PB1964*. Rothmund Thomson syndrome with congenital cataracts and severe growth restriction is associated with a not previously described gene in an autosomal recessive inheritance pattern in seven families

Authors:

G. Yamamoto¹, R. Di Lazaro¹, T. da Silva², L. Rocha¹, D. Bartholdi³, A. Schaller³, C. Zweier³, R. S. Honjo⁴, C. Utagawa⁵, C. Steiner⁶, C. A. Kim⁴, M. Passos-Bueno⁷, N. C. Hoch⁸, D. R. Bertola⁹; ¹Inst. da Criança - Univ. of São Paulo, São Paulo, Brazil, ²Inst. de Química - Univ. of São Paulo, São Paulo, Brazil, ³Univ. Hosp. Bern, Bern, Switzerland, ⁴Inst. da Criança - Univ. of Sao Paulo, Sao Paulo, Brazil, ⁵Centro Universitário de Volta Redon, Volta Redonda, Brazil, ⁶UNICAMP, Campinas, Brazil, ⁷Univ. of Sao Paulo, Sao Paulo, Brazil, ⁸Univ. of São Paulo, São Paulo, Brazil, ⁹Inst. da Criança - Univ. of Sao Paulo, Sao Paulo, Brazil

Abstract Body:

Introduction: Rothmund Thomson syndrome (RTS) is a rare autosomal recessive genodermatosis characterized by poikiloderma, abnormal hair/nails; short stature; eye and skeletal abnormalities. Based on clinical and molecular data, this syndrome is classified in type I, including the presence of cataracts and variants in a recently identified gene (*ANAPCI*) and type II, without cataracts and increased predisposition to cancer, associated with biallelic variants in *RECQL4*. **Methods:** The cohort is comprised of 6 Brazilian probands and a pair of siblings (a boy and a male fetus product of TOP, from Portuguese/Swiss origin). None of the probands had history of consanguinity and all presented a severe, homogeneous phenotype (wide-spread poikiloderma, congenital cataracts, facial dysmorphisms; severe growth restriction with growth hormone deficiency, hypogonadism and skeletal abnormalities). Exome sequencing in the Brazilian individuals failed to show pathogenic variants in the coding regions of *RECQL4* and *ANAPCI*. Different genetic techniques were applied to identify a novel *locus* associated with RTS, which we propose to name type III, including SNP-chip array for linkage analysis, whole-genome sequencing (WGS), as well as functional analysis in patient's fibroblast cultured cells. **Results:** Despite admixed ancestries, linkage analysis identified a heterozygous shared chromosomal region of 8 Mb in chromosome 10q in all Brazilian individuals. WGS of the Brazilian probands identified biallelic variants in a gene within this region, all sharing one deep intronic variant in intron 4 of this gene in *trans* with intragenic deletions or a missense variant. On the non-Brazilian case, WGS disclosed the same intronic variant inherited from the father with Portuguese ancestry and a frameshift variant in *trans*. Functional and cDNA analysis depicted alternative splicing due to the intronic variant and consequent loss of function. **Conclusions:** We have identified biallelic pathogenic variants in a gene not previously described associated with RTS in seven families, all with a shared haplotype from probable Portuguese origin. We propose that RTS with congenital cataracts and severe growth restriction should be classified as RTS type III. //FAPESP 2013/08028-1; CNPq: 303375/2019-1. //We declare no conflicts of interest.

Mendelian Phenotypes Posters - Wednesday

PB1965. Screening for Genetic Modifiers of *MED12/kto* Using Naturally Occurring Variation in *Drosophila melanogaster*

Authors:

K. Bussey, N. O. N. Yepiz, V. Shankar, R. R. H. Anholt, T. F. C. Mackay; Clemson Univ., Greenwood, SC

Abstract Body:

MED12 is a component of the Mediator Complex which regulates gene expression. Mutations in different domains of the *Med12* protein are associated with five rare genetic disorders: FG Syndrome, Ohdo Syndrome, Hardikar Syndrome, Lujan syndrome, and Non-Specific X-Linked Intellectual Disability (XLID). They are characterized by a wide range of effects including intellectual disability, behavioral, and congenital defects. Different individuals with the same *MED12* mutation present with differing severity of symptoms, suggesting the presence of genetic modifiers of *MED12*-related disorders. However, the rarity and diversity of these disorders make identifying the genetic basis of pathogenicity and genetic modifiers challenging in human populations. *Drosophila melanogaster* has a strongly conserved functional ortholog of *MED12*, *khotalo* (*kto*). Knocking down *kto* expression using RNA interference shows that *kto* affects startle response and sleep and activity traits. We are utilizing natural variation among the fully sequenced inbred lines of the *D. melanogaster* Genetic Reference Panel (DGRP) to identify genetic modifiers altering effects of *kto* mutations. We crossed a CRISPR/Cas9-generated null mutation of *kto* (*kto*-), and a wild type *kto* allele (*kto*+) in the same genetic background, to the DGRP lines, and quantified sleep and activity and startle-induced locomotion in the F1 offspring. We used analysis of variance to estimate the effects of *kto* genotype (G), DGRP line (L) and the genotype by environment interaction (G×L) on these phenotypes. A significant G×L term indicates the presence of modifier loci, (i.e., the effect of *kto* genotype is different across the DGRP lines). We found significant G×L terms for startle behavior ($P = 0.0087$) and night activity bouts, sleep bout lengths, and daytime sleep ($P < 0.001$) with both amelioration or exacerbation of effects in *kto*- relative to *kto*+ genotypes. We can perform genome wide association analyses using the difference in phenotypes between *kto*- and *kto*+ for each DGRP line (significant G×L variance is the same as the variance in this difference) to identify candidate genetic modifiers of *kto*. Identification of top modifier candidates will provide insight into how genetic modifiers may influence the manifestation and severity of *MED12*-related disorders and identify modifiers as potential therapeutic targets.

Mendelian Phenotypes Posters - Thursday

PB1966. Seizures in a patient with Autosomal recessive Intellectual disability type 5

Authors:

E. Aliu¹, M. Hasan², G. Mainali¹, S. Paudel¹; ¹Penn State Hershey Med. Ctr., Hershey, PA, ²Pennsylvania State Hlth.Coll. of Med., Hershey, PA

Abstract Body:

Autosomal recessive Intellectual disability type 5 (OMIM 61109), is caused by biallelic pathogenic variants in the NSUN2 gene which encodes for a methyltransferase involved in several biological processes, ranging from stress responses to neurodevelopment. NOP2/Sun RNA Methyltransferase 2 (NSUN2) encodes a methyltransferase which is required for splicing at C34 of tRNA Leu (CAA) and C47 and C48 of tRNA-Asp (GTC). A proposed mechanism postulates that the loss of 5-methylcytosine increases the angiogenin-mediated endonucleolytic cleavage of tRNAs, leading to an accumulation of 5' tRNA-derived small RNA fragments causing a decrease in the rate of protein translation which activates stress pathways causing reduced cell size and increased apoptosis of cortical, hippocampal, and striatal neurons. This disorder presents with intellectual disability, facial dysmorphism, juvenile cataracts, chronic nephritis, hearing impairment, seizures, cerebellar atrophy, and microcephaly. To date there have been 25 cases reported caused by NSUN2 deficiency but only one patient had seizures treated with valproic acid. We report a 15-year-old male with a history of autism, and moderate intellectual disability, who presented with seizure-like episodes. These episodes were described as impaired consciousness, drooling, eye deviation, and twitching of the corner of the mouth occurring almost on a daily basis for the past year. Each episode would last from 15-20 minutes which is then followed by sleepiness. He underwent 24 hours ambulatory EEG which showed frequent generalized 1-2.5 Hz spikes and slow-wave discharges and sleep-activated multifocal epileptiform discharges. Work-up was initiated to evaluate the etiology of the seizure. In the past, he had genetic evaluations which were not diagnostic. An Autism/ID Xpanded panel (GeneDx, Bethesda MD), identified a homozygous frameshift (c.560del;p.Pro187Leufs*8) variant, predicted to result in loss of function of NSUN2 gene (NM_017755.5). Parental studies were only done on his mother which confirmed she was a carrier. He was started initially on valproic acid but seizures were not fully controlled. After adding clobazam his seizures are controlled. At baseline, the patient is still non-verbal, ambulates mainly by wheelchair, although can walk some with assistance, has generalized hypertonia, and uses a GT for feeding as he is unable to eat by mouth. In conclusion, we are reporting a new pathogenic variant, a 2nd patient with NSUN2 deficiency who developed seizures in his teens, which suggests that epilepsy is part of the spectrum of this condition although an uncommon one, and a successful regimen to control his seizures.

Mendelian Phenotypes Posters - Wednesday

PB1967*. Sequencing analysis of nonsyndromic cleft lip / palate in Brazilian patients reveals *CTNNA1*, *ESRP1* and *PRICKLE1* as novel candidate genes.

Authors:

L. Brito¹, G. Hsia¹, J. Wang², G. Yamamoto³, S. Ferreira², M. Passos-Bueno²; ¹Inst. of BioSci.s, Univ. of São Paulo, São Paulo, Brazil, ²Inst. of BioSci.s, Univ. of São Paulo, Sao Paulo, Brazil, ³Inst. da Criança - Univ. of São Paulo, São Paulo, Brazil

Abstract Body:

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a prevalent birth defect impacting around 1 in 1,000 liveborn children, who frequently face feeding, hearing, speaking and psychological problems. NSCL/P etiology is complex and most cases are sporadic, while 30% of cases are familial and present variable inheritance patterns. The genetic etiology embraces multiple low-effect risk loci, as identified by genome-wide association studies, but also moderate-to-high effect variants, which have been increasingly detected by sequencing approaches, in genes such as *CDH1*, *CTNND1*, *ESRP2* and *ARHGAP29*. We have characterized the landscape of rare pathogenic variants in Brazilian NSCL/P individuals, and preliminary analyses have already implied *CDH1* and *ARHGAP29* as prevalent genes harboring such variants. Here, we present our final results, which included the exome sequencing of 47 NSCL/P probands from familial cases and 49 affected relatives, performed with Illumina HiSeq platform, and the targeted sequencing of 540 probands from familial cases and 56 relatives, for a panel of candidate genes. In this panel, we included key genes involved in the canonical (beta-catenin) or non-canonical (planar cell polarity) Wnt signaling pathways that are expressed during human palate and frontonasal embryogenesis, or that were previously implicated with orofacial clefts in humans or animal models, such as *CDH1*, *CTNNA1*, *CTNNB1*, *CTNND1*, *DVL3*, *ESRP1/2*, *FZD1/2*, *LRP6*, *PRICKLE1/2*, *RHOA*, *SNAIL*, *TGFB1*, *WNT5A*, *WNT10A*, *WNT11* and *ZEB1/2*. Variant prioritization criteria included frequency <0.5% in the gnomAD populations and in the Brazilian variants database ABraOM and CADD score > 20 (i.e., variants assigned in silico to be at the top 1% of human deleterious variants). Among the variants prioritized after both sequencing strategies, we call attention to three probably pathogenic ones located in genes never implicated with NSCL/P in humans to date: a missense variant in *CTNNA1*, a frameshift insertion in *PRICKLE1* and a stopgain variant in *ESRP1*. All these variants showed complete segregation with the phenotype and are potentially disruptive to the protein functions, although functional validation analysis are still ongoing. Additionally, novel and previously reported pathogenic variants were also found in *CDH1* (9 families), *CTNND1* (5 families) and *ARHGAP29* (5 families), which, in conclusion, represent the most common causes of familial NSCL/P in this Brazilian sample (3.2%); most importantly, our work provides the first evidence for the etiological contribution of *CTNNA1*, *ESRP1* and *PRICKLE1* to human NSCL/P.

Mendelian Phenotypes Posters - Thursday

PB1968. Severe ABCA4-retinal dystrophy caused by homozygosity for a novel double mutated allele.

Authors:

S. Aslaksen¹, I. Aukrust², M. Hamre Bu¹, M. Innselset Flydal², C. Bredrup², E. Bratland², P. Knappskog²; ¹Univ. of Bergen, Bergen, Norway, ²Haukeland Univ. Hosp., Bergen, Norway

Abstract Body:

Pathogenic variants in the *ABCA4* gene are the leading cause of vision loss in inherited retinal diseases. More than 1500 diseases-causing variants in *ABCA4* have been reported, giving rise to a heterogeneous clinical spectrum of ABCA4-retinal dystrophies. The range of clinical phenotypes and disease severity is believed to be determined by the patient's genotype defined as the specific combination of pathogenic variants in the *ABCA4* gene. ABCA4-retinal dystrophies are inherited in an autosomal recessive manner, either by homozygous or compound heterozygous variants. Here we report a rare case of severe retinal degeneration caused by *two* homozygous pathogenic variants in *ABCA4* (NM_000350.3) detected by exome sequencing; the most frequent pathogenic variant c.5882G>A p.(Gly1961Glu), associated with mild phenotypes and later disease onset, and c.634C>T p.(Arg212Cys), associated with more severe phenotypes. The patient had normal visual acuity in his early childhood but experienced relatively rapidly reduced visual acuity from the age of 12. Upon examination at 13 years of age, he had best corrected visual acuity (BCVA) od:0.15, os: 0.2 and a red-green color deficit. Bilateral macular atrophy was found, but no pallor of the optic disc. Electroretinogram (ERG) showed severely reduced cone function. This severe phenotype observed in the patient homozygous for a novel double mutated allele indicates a synergistic effect of the two pathogenic variants p.(Gly1961Glu) and p.(Arg212Cys). The impact of this disease-causing variant combination on the functional properties of ABCA4 will be examined in future experiments.

Mendelian Phenotypes Posters - Wednesday

PB1969. Shared and divergent transcriptional dysregulation is present across brain regions and cell types in adult *Chd8* haploinsufficient mice

Authors:

A. Nord, S. Lozano, C. Canales, K. Cichewicz, A. Wade, E. Smith, N. Seban, R. Ortiz, J. Zhu, M. Corea, D. Rahbarian; UC Davis, Davis, CA

Abstract Body:

De novo mutations in the chromatin-remodeling factor *CHD8* (Chromodomain-Helicase DNA-binding protein 8) are strongly associated with autism spectrum disorder (ASD) and more generally with neurodevelopmental disorders (NDDs). Mice with heterozygous germline loss-of-function mutation in *Chd8* exhibit genomic, neuroanatomical, and behavioral pathology. With the goal of identifying the molecular consequences of *Chd8* haploinsufficiency in the postnatal brain across regions and cell types, we used bulk RNA-sequencing in cerebral cortex, hippocampus, and cerebellum, and single nucleus (sn)RNA-seq in the adult cortex to compare wild type and heterozygous *Chd8* mutant mice. We identified differentially-expressed genes (DEGs) that were altered across all three brain regions, as well as region-specific signatures, with the greatest dysregulation identified in the cerebellum. DEG pathways included neuroinflammatory, metabolic, and synaptic processes. We identified a set of genes that were consistently perturbed across brain regions. Male and female DEG signatures showed some differences in DEG pathways and magnitude of effects. Single nuclei resolution experiments revealed transcriptional dysregulation in cortex. We found changes that were present across all neurons, as well as unique signatures within excitatory and inhibitory subsets. Changes in microglia and oligodendrocyte lineages were also identified. Similar to bulk RNA-seq data, males and female mutants had both overlapping and sex-specific DEG signatures. Validation of regional and cell-type specific changes is in progress. Overall, our findings show that decreased *Chd8* dosage alters gene expression in adult brain both via shared pathways, and in region-specific and sex-specific manners, with some transcriptional aberration attributable to specific cell types in the cortex. Our results present a systems-level characterization of molecular and cellular pathways that are disrupted in adult *Chd8* mutant brain, and potentially captures generalizable dysfunction driven by mutation to chromatin-associated NDD risk genes.

Mendelian Phenotypes Posters - Thursday

PB1970. Siblings with hypophosphatemic rickets and homozygous *CYP27B1* variants

Authors:

L. Wang¹, M. Shanmugasundaram¹, E. Cooney^{1,2}, P. Lee^{1,3}; ¹Univ. of Texas Med. Branch, John Sealy Sch. of Med., Galveston, TX, ²Univ. of Texas Med. Branch, Dept. of Pediatrics, Div. of Med. Genetics & Metabolism, Galveston, TX, ³Univ. of Texas Med. Branch, Dept. of Pediatrics, Div. of Endocrinology, Galveston, TX

Abstract Body:

Vitamin D-dependent rickets type 1A (VDDR1A) is a rare autosomal recessive disorder caused by biallelic pathogenic variants in *CYP27B1*, resulting in deficient 1 α -hydroxylation of 25-hydroxyvitamin D (25OHD) to 1,25-dihydroxyvitamin D (1,25(OH)2D). We report siblings presenting with hypophosphatemic rickets and homozygous *CYP27B1* variants. The proband was a 20-month-old Hispanic female presenting with a 2-month history of bilateral wrist swelling, gross motor delays, and deceleration of height from 11% at 11-months-old to 1% at the time of the first encounter. Evaluation showed radiographs consistent with rickets, calcium (Ca) 9.1 mg/dL (RR 8.6-10.6), phosphorus (Phos) 2.8 mg/dL (RR 3.5-6.7), alkaline phosphatase (ALP) 1748 U/L (RR 150-370), 25OHD 46 ng/mL (RR 25-80), 1,25(OH)2D 42.8 pg/mL (RR 19.9-79.3), and parathyroid hormone (PTH) 902.6 pg/mL (RR 12-88). A clinical diagnosis of hypophosphatemic rickets was made. After empiric treatment with calcitriol and phosphate, her Phos, PTH, and ALP normalized and the rickets, growth failure, and developmental delay resolved. The proband's younger brother was a well-appearing 7-month-old with Ca 10.2 mg/dL (RR 7.8-11.2), Phos 3.9 mg/dL (RR 4.5-6.7), ALP 790 U/L (RR 185-430), and mild bone demineralization on skeletal survey. By 13 months, he developed radiographic evidence of rickets. Empiric treatment for hypophosphatemic rickets was started. Whole exome sequencing of both children showed homozygous *CYP27B1* c. 1040T>A (p. Ile347Asn) mutations classified as variant of uncertain significance (VUS). Variants were not found in the genes currently associated with familial hypophosphatemic rickets (*PHEX*, *FGF23*, *KL*, *DMPI*). Mother was heterozygous for the VUS; although father was unavailable, he was presumed an obligate carrier. The VUS is relatively uncommon in population databases (rs556530774, Exome Aggregation Consortium 0.009%). Predictive algorithms (SIFT, PolyPhen-2, Align-GVGD) suggest the VUS is disruptive. The VUS was clinically interpreted in the context of family segregation and *a priori* risk resulting in the decision to use VDDR1A instead of hypophosphatemic rickets as the working diagnosis. Calcitriol was continued. Phos supplementation was discontinued. To date, the children's growth and development have remained appropriate and their labs are within normal limits. In summary, we report an uncommon biallelic *CYP27B1* VUS presenting with hypophosphatemic rickets. This case emphasizes the variable clinical presentation of monogenic rickets and highlights the clinician's duty to interpret VUS within the context of *a priori* risk and family segregation to inform medical management.

Mendelian Phenotypes Posters - Wednesday

PB1971. Six patients with genetic disorders with atypical symptoms: how to perceive them

Authors:

T. Uehara¹, M. Inaba¹, S. Mizuno²; ¹Aichi Developmental Disability Center, Kasugai-city, Japan, ²Aichi Developmental Disability Center, Kasugai, Aichi, Japan

Abstract Body:

[Background] In recent years, methods of genetic analysis, including next-generation sequence, have made remarkable progress. That has contributed to an improvement in the diagnostic rate of undiagnosed patients with rare diseases. Progress of genetic analysis have also helped to elucidate the causes of previously undiagnosed diseases. As a result, many new diseases have been reported in the past few years. A part of diagnosed patients by genetic analysis with rare or novel diseases have atypical or unknown phenotypes. Here we report patients with rare or novel diseases who show atypical or unreported phenotypes. [Methods] We investigated patients with congenital anomaly showing with atypical or previously unreported phenotypes who had visited our clinic in the past year, based on their electronic medical record descriptions. [Results] We identified the following six individuals as having interesting phenotypes. Patient 1 is a 9-year-old girl definitely diagnosed by genetic analysis as Coffin-Siris syndrome. She has had seizures from the age of 7 and is currently on antiepileptic medication. Patient 2 is a 21-year-old boy diagnosed by genetic analysis as DLG3 mutation. He showed postnatal overgrowth. Patient 3 is a 14-year-old boy diagnosed by genetic analysis as CUL3 mutation. He showed macrocephaly (+2SD) and short stature (-2SD). Patient 4 is a 18-year-old boy diagnosed by genetic analysis as CDK19 mutation. He showed extreme growth failure (-7SD) hand flapping, and dysmorphic features. Patient 5 is a 16-year-old boy diagnosed by genetic analysis as ODC1 mutation. He showed premature tooth eruption and retardation of bone age. Patient 6 is a nine-year-old boy diagnosed by genetic analysis as CSTF2 mutation. He showed craniosynostosis. [Discussion] The phenotypes identified by this investigation have not previously been reported in patients with the same disease or were thought to be very rare in the in patients with the same genetic mutation. There are three possible explanations for such cases: 1) they are coincidental complications, 2) they should be considered as phenotypic extensions, or 3) they should be considered as different diseases due to differences in variants, even if the mutated genes are the same. Complications of seizures in CSS patients have been increasingly reported in recent years and may occur more frequently than was initially thought generally. Each phenotypes observed in the remaining five patients are difficult to compare with previous cases, due to new diseases whose reports are still few. It is important to accumulate cases to clarify which of the three patterns is applicable.

Mendelian Phenotypes Posters - Thursday

PB1972. *Slc9a6* mutation causes Purkinje cell loss and ataxia in the *shaker* rat

Authors:

K. Figueroa¹, C. Anderson¹, S. Paul¹, W. Dansithong¹, M. Gandelman¹, D. Scoles², S. M. Pulst¹; ¹Univ. of Utah, Salt Lake City, UT, ²Univ of Utah, Salt Lake City, UT

Abstract Body:

The *shaker* rat carries a spontaneous mutation leading to Purkinje cell (PC) degeneration resulting in body tremor and progressive ataxia. In previous work, we fine-mapped the *shaker* locus to the distal end of the X chromosome. In this current study, we sought to identify the mutation underlying *shaker*, with confirmation through functional complementation. We fine-mapped the *shaker* locus using an F2 hybrid strategy to the distal end of the X chromosome and performed transcriptomic studies to identify deleterious variants. We identified a XM_217630.9 (*Slc9a6*):c.[191_195delinsA] variant in the *Slc9a6* gene segregating with disease that was predicted to generate a truncated sodium-hydrogen exchanger 6 (NHE6) protein, p.(Ala64Glufs*23). Having identified solute carrier family 9 member A6 (*Slc9a6*) as the likely causative gene, we generated an adeno-associated virus (AAV) targeting *Slc9a6* expression to PCs using a mouse L7-6 (L7) promoter. We administered AAVs through intracerebroventricular injection at 5 weeks of age, prior to PC death. We tracked motor ataxia through 25 weeks of age, and following motor experiments, we harvested cerebella for histology and molecular analyses, quantifying mRNA and protein expression of *Slc9a6*/NHE6, as well as several cerebellar health markers. Administration of AAV9-PHP.eB expressing rat *Slc9a6* prior to symptom onset reduced the *shaker* motor, molecular, and cellular phenotypes. Functional complementation restoring an average of ~10% wildtype NHE6 not only reduced motor ataxia, but also significantly increased both mRNA and protein expression of several key cerebellar markers, CALB1, PCP2, and RGS8. In conclusion, *Slc9a6* is mutated in *shaker*. Human *SLC9A6* loss-of-function mutations cause Christianson syndrome, an epileptic encephalopathy. AAV-based gene therapy may be a viable therapeutic strategy for Christianson syndrome, and the *shaker* rat model may aid in therapeutic development.

Mendelian Phenotypes Posters - Wednesday

PB1973. Splicing noise is variable across human introns, tissues and age and modelling its characteristics can improve our understanding of age-related diseases of the human brain.

Authors:

S. Garcia-Ruiz^{1,2}, D. Zhang^{1,2}, R. Reynolds^{1,2}, E. Gustavsson^{1,2}, S. Guelfi², J. Botia^{3,2}, L. Collado Torres⁴, M. Rytén^{5,2}; ¹Dept. of Genetics and Genomic Med. Res. & Teaching, UCL GOS Inst. of Child Hlth., London, United Kingdom, ²Dept. of Neurodegenerative Disease, Queen Square Inst. of Neurology, UCL, London, United Kingdom, ³Dept. of Information and Communications Engineering, Univ. of Murcia, Murcia, Spain, ⁴Lieber Inst. for Brain Dev., Baltimore, MD, ⁵UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom

Abstract Body:

Alternative splicing is a characteristic of most multi-exonic human genes and is used to generate transcriptomic and proteomic diversity across cells. However, the analysis of RNA-sequencing data from human tissues suggests that this process can be noisy with the consistent identification of reads that cannot be mapped to the known transcriptome. We characterised mis-splicing using RNA-sequencing data from 16,955 samples and 47 human tissues, provided by the Genotype-Tissue Expression Consortium v8, in the process of generating a database of annotated-novel intron pairs which we provide as a web resource (https://rytenlab.com/intron_db/ and https://soniagarciaruiiz.shinyapps.io/intron_db/; docker image: https://hub.docker.com/repository/docker/soniaruiz/intron_db). We focused on reads mapping with a gapped alignment to the genome that could be assigned to a transcript through sharing of known acceptor or donor splice sites but were absent from annotation. Analysing this data, we found that mis-splicing occurs at significantly higher rates at acceptor versus donor splice sites, that novel donor and acceptor sites are located in close proximity to annotated sites in a pattern consistent with the known molecular architecture of splicing, and that mis-splicing is heterogeneous across introns. We estimated that 63% of mis-splicing events detected would be expected to be deleterious. Using linear regression models, we found that both local sequence information and gene-level features significantly impacted splicing noise. By measuring splicing noise across all human tissues, we showed that the effect of 5'/3' consensus sequence and conservation scores on mis-splicing varied in magnitude across human tissues even for a common set of introns, suggesting that splicing accuracy might be affected by RNA-binding protein (RBP) expression. Consistent with this hypothesis, we found that siRNA knockdown of spliceosomal RBPs generated predictable patterns of increased splicing noise and that significant increases in splicing noise with age correlated with a reduction in RBP expression. Interestingly, the effect of age on RBP expression and splicing noise was most apparent in the human brain, with the most mis-spliced introns being significantly enriched in synaptic genes. Together, our findings suggest that understanding patterns of mis-splicing could provide novel insights into age-related human diseases, particularly of the human brain.

Mendelian Phenotypes Posters - Thursday

PB1974*. *SRSF1* haploinsufficiency is responsible for a syndromic developmental disorder with intellectual disability and variable marfanoid habitus

Authors:

A. Garde^{1,2}, E. Bogaert³, T. Gautier⁴, Solve-RD DITF-ITHAC, Y. Duffourd², T. Erin⁵, A. Munnich^{6,7}, C. Besmond⁸, E. Roeder⁹, A. Trimouille^{10,11}, S. Naudion¹⁰, M. Legendre¹⁰, M. Doco-Fenzy¹², M. Seri¹³, A. Yeung¹⁴, J. Lupski¹⁵, T. Mitani¹⁵, C. Cesario¹⁶, C. Brasch Andersen¹⁷, M. Lyons¹⁸, D. Genevieve¹⁹, F. Faletra²⁰, K. Mowrey²¹, H. Northrup²¹, P. Callier²², A-S. DENOMME-PICHON^{2,23}, the Solve-RD Consortia, SRSF1 Research Group, C. Thauvin-Robinet^{2,23,24}, L. Faivre^{1,2}, J. Govin⁴, B. Dermaut³, A. Vitobello^{2,23}; ¹Ctr. de Référence Anomalies du Développement et Syndromes Malformatifs, FHU TRANSLAD, Hôpital d'Enfants, CHU Dijon, Dijon, France, ²UMR-Inserm 1231 GAD team, Génétique des Anomalies du développement, Université de Bourgogne Franche-Comté, FHU-TRANSLAD, Dijon, France, ³Ctr. for Med. Genetics, Ghent Univ. Hosp., Ghent, Belgium, ⁴Inserm U1209, CNRS UMR 5309, Univ. Grenoble Alpes, Inst. for Advanced BioSci.s, Grenoble, France, ⁵GeneDx, LLC, Gaithersburg, MD, ⁶Fédération de Génétique et Inst. Imagine, Université Paris Descartes, Hôpital Necker Enfants Malades, Paris, France, ⁷INSERM UMR1163, Inst. IMAGINE, Paris, France, ⁸INSERM UMR1163, Inst. IMAGINE, France, France, ⁹Dept.s of Pediatrics and Molecular and Human Genetics, Baylor Coll. of Med., San Antonio, TX, ¹⁰Service de Génétique Médicale, CHU de Bordeaux, Bordeaux, France, ¹¹Laboratoire MRGM, INSERM U1211, Université de Bordeaux, Bordeaux, France, ¹²Service de Génétique, CHU de Reims, Reims, France, ¹³Med. Genetics Unit, IRCCS AOU Bologna, Univ. of Bologna, Bologna, Italy, ¹⁴Victorian Clinical Genetics Services, Parkville, Australia, ¹⁵Baylor Coll. of Med., Houston, TX, ¹⁶Lab. of Med. Genetics, Dept. of Lab., Bambino Gesù Children's Hosp., Rome, France, ¹⁷Dept. of Clinical Genetics, Odense Univ. Hosp., Odense, Denmark, ¹⁸Greenwood Gen Ctr, N Charleston, SC, ¹⁹Département de Génétique Médicale, Maladies rares et Médecine Personnalisée, CHRU de Montpellier, Montpellier, France, ²⁰IRCCS Burlo Garofolo, Trieste, Italy, ²¹Dept. of Pediatrics, McGovern Med. Sch. at the Univ. of Texas Hlth.Sci. Ctr. at Houston (UTHlth.) and Children's Mem. Hermann Hosp., Houston, TX, ²²Laboratoire de Génétique Chromosomique et Moléculaire, CHU Dijon, Dijon, France, ²³UF Innovation en diagnostic génomique des maladies rares, CHU Dijon, Dijon, France, ²⁴Ctr. de Référence Déficiences Intellectuelles de causes rares, FHU TRANSLAD, Hôpital d'Enfants, CHU Dijon, Dijon, France

Abstract Body:

SRSF1 (also known as ASF/SF2) is an evolutionary conserved non-snRNP splicing protein that belongs to the serine/arginine rich domain (SR) family. It is known to be a master regulator of constitutive and alternative splicing. Through international data sharing, we gathered 16 patients (9 females and 7 males) carrying 14 different germline *SRSF1* variants including frameshift, nonsense, and missense variant alleles, as well as one microdeletion of the region 17q22 that included *SRSF1*. Variants occurred mostly *de novo*, however in one family a germline mosaicism was suspected to explain the recurrence in two siblings, while in another case the inheritance was unknown. Main clinical features included developmental delay (DD), intellectual disability (ID), hypotonia, behavioral disorders, skeletal and cardiac anomalies. Three out of 16 patients showed marfanoid habitus with learning disabilities or ID/DD. *In silico*, missense variants indicated predicted conformational changes of the protein surface. Functional studies supported pathogenicity, as established by *in vivo* splicing assay in *Drosophila*. Eye-specific overexpression of SF2, the *Drosophila* ortholog of SRSF1, induced a severe developmental phenotype due to missplicing of key genes involved in normal eye development. In the current study, we replicated this finding and found that overexpression of the human *SRSF1* was a phenocopy pointing towards functional conservation of SF2 and SRSF1. In addition, similar to previously characterized

ASHG 2022 Annual Meeting Poster Abstracts

splicing-deficient forms of SRSF1, clinical missense variants lost the capacity to cause an eye phenotype proving their loss-of-function nature. Overall, these results indicate that haploinsufficiency of *SRSF1* is responsible for a syndromic neurodevelopmental disorder trait with recurrent marfanoid habitus and ID/DD.

Mendelian Phenotypes Posters - Wednesday

PB1975. Subtle genomic duplication on Chromosome 6: a hypothesis for a causative mechanism

Authors:

A. John¹, R. Collins¹, V. Tonk², L. Nagy¹; ¹Texas Tech Univ. Hlth.Sci. Ctr., Lubbock, TX, ²Texas Tech Univ Hlth Sci Ctr, Lubbock, TX

Abstract Body:

Background: Duplications of Chromosome 6, especially those of the long arm (q), are a very rare genetic condition that can lead to a spectrum of defects. Anomalies include growth restriction, heart and brain defects, various dysmorphic physical features, genitourinary system abnormalities, and developmental delay. **Case Report:** A 15-month-old male presented with failure to thrive, hypertonia, daily emesis, repetitive banging of head against wall, and inconsolable screaming. Confirmatory imaging noted a Chiari I malformation with a pre-syrinx and a tethered cord. The patient was born at 36 weeks without complications during pregnancy or birth and had normal newborn screenings. Congenital defects included hypospadias and 2 small atrial septal defects. His genetic workup revealed a 177kbp duplication of cytogenetic band 6q27 that involved the T gene (OMIM 601397). The parents do not share this duplication. The patient underwent Chiari decompression and filum sectioning that resolved his presenting symptoms. **Conclusion:** To our knowledge, this is the smallest case of a duplication (177 kbp) of Chromosome 6q that led to the wide spectrum of manifestations documented in trisomy 6q. We propose that due to the hypothesized tight communication amongst band regions in chromosome 6q, duplications regardless of size or location can lead to similar manifestations.

Mendelian Phenotypes Posters - Thursday

PB1976. Sunken symptoms of *SRCAP* mutation: A case report of atypical phenotype of Floating Harbor Syndrome

Authors:

N. Nesheiwat¹, W. K. Chung²; ¹Div. of Clinical Genetics, Dept. of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY, ²Div. of Molecular Genetics, Dept. of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY

Abstract Body:

Floating-Harbor syndrome (FHS) is an autosomal dominant condition characterized by short stature, delayed bone growth, skeletal anomalies, seizures, language impairment, and intellectual disability. Individuals can have distinct facies including deep-set eyes, thin lower lip, triangular-shaped face, and a long broad nose. The *SRCAP* gene encodes for a chromatin-remodeling ATPase, which serves as a co-activator for CREB-binding protein (CREBBP) and CBP-mediated transcription. The CREBBP plays a key role in regulating cell growth and division and is important for normal development. Here we report a 14-year-old female of Ashkenazi Jewish descent with a history of moderate developmental delay, language impairment, amenorrhea, progressive severe sensorineural hearing loss, short stature with poor response to growth hormone therapy over 8 years ($Z = -3.95$), umbilical hernia, and congenital duplicated ureters. Upon examination, she displayed hypernasal speech, broad nose, low-set ears, bilateral clinodactyly, pectus excavatum, and hypotonia. She also had unique findings of the development of severe lower lip angioedema of unknown etiology at 12 years of age and bilateral proptosis. Extensive workup, including multiple biopsies of the lip, have been negative. SNP microarray returned negative. Trio exome sequencing identified a de novo pathogenic c.7330 C>T (p.Arg2444Ter) variant in exon 34 of the *SRCAP* gene. This nonsense variant is predicted to result in protein truncation, as the last 787 amino acids are lost. Truncating variants in exons 33 and 34 of *SRCAP* are associated with FHS. The p.Arg2444Ter variant has been reported in patients with the characteristic FHS phenotype. The prevalence of FHS is not known, yet seventy-three individuals with heterozygous *SRCAP* pathogenic variants have been reported in the medical literature to date. In addition to some of previously mentioned classical FHS features, our patient also had atypical findings of lip angioedema, proptosis and sensorineural hearing loss, which are contrasting to typical features of thin lips, deep-set eyes, and conductive hearing loss typically seen in FHS. These unique features have not been previously reported with this condition. There remains the possibility that these are new traits associated with FHS, however secondary causes cannot be ruled out. This suggests the phenotype of the condition is quite variable and our patient's distinct features of proptosis, lower lip angioedema, sensorineural hearing loss further add to the expansion of the clinical spectrum and variable expressivity of FHS. Additional cases will help expand the phenotype and knowledge of this rare disorder.

Mendelian Phenotypes Posters - Wednesday

PB1977. Systematic bioinformatic analysis promotes a missense variant in *XK* as potential cause of X-linked intellectual disability.

Authors:

T. Litster¹, M. Corbett^{1,2}, T. Dudding-Byth³, J. Gecz^{1,2,4}; ¹Univ. of Adelaide, Adelaide, Australia, ²Robinson Res. Inst., Adelaide, Australia, ³Hunter New England Area Hlth.Service, Newcastle, Australia, ⁴South Australian Hlth.and Med. Res. Inst., Adelaide, Australia

Abstract Body:

Despite only making up 5% of the genome, the X chromosome shows enrichment for genes involved in intellectual disability (ID), housing approximately 15% of known ID genes. Despite this enrichment, many cases of ID with a clear X-linked inheritance pattern remain without a genetic diagnosis. These cases provide opportunities to discover novel ID genes and variants on the X chromosome and enhance our understanding of the genetic landscape of ID. One such case is the MRX11 family (Kerr et al. 1992), who present with multiple males affected by mild to moderate, non-syndromic ID. This family has undergone extensive genetic testing and remains unresolved.

Here we applied a systematic bioinformatic analysis pipeline to this family to identify the disease causing variant. Genome-wide linkage analysis defined the linkage interval between rs7067109 and rs5918300 (hg38 chrX:32775936-42171940) with maximum LOD score of 3.34. Analysis of short-read whole genome sequencing data revealed no coding, splicing or regulatory variants predicted to be damaging in known ID genes within this region. RNA-seq data from patient-derived lymphoblastoid cell lines showed no pathogenic aberrant splicing or expression events. Finally, long-read sequencing with Oxford Nanopore using adaptive sampling to enrich for sequences on the X-chromosome revealed no damaging structural variants.

After extensive analysis, only one variant remained that was predicted to be damaging and segregated with disease; NM_021083.4(*XK*): c.872A>G: p.(Asn291Ser). The *XK* protein forms part of the Kell antigen complex on the surface of erythrocytes, and acts as a phospholipid scramblase involved in apoptotic pathways in other cell types. Pathogenic, loss of function variants in *XK* are typically implicated in McLeod Syndrome which is characterised by neuroacanthocystosis (spine-like protrusions that form on erythrocytes) and a late onset neurodegenerative disorder that is distinct from the phenotypes of affected males in the MRX11 family. Therefore, our extensive and systematic bioinformatic analysis, through a process of elimination, has revealed *XK* as a potential novel ID gene with a phenotypic spectrum separate from McLeod syndrome.

Mendelian Phenotypes Posters - Thursday

PB1978*. Systematic identification of disease-causing near-coding variants in 29,759 individuals with rare disease

Authors:

N. Whiffin^{1,2}, A. Martin-Geary², E. Ng², N. Wieder², M. Ferla²; ¹Univ. of Oxford, Oxford, United Kingdom, ²Wellcome Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom

Abstract Body:

Genomic regions directly adjacent to the protein coding sequence, such as untranslated regions (UTRs) and gene promoters (collectively referred to as ‘near-coding’ regions), are known to play important roles in transcriptional and post-transcriptional gene regulation. Variants that disrupt these elements have been shown to cause severe disease phenotypes. However, these regions are not routinely included in clinical diagnostic pipelines. The extent of ‘missed’ near-coding diagnoses that have the potential to be identified by expanding coding specific pipelines is currently unknown.

Here, we analysed near-coding de novo variants in 12,020 rare disease trios in the Genomics England 100,000 Genomes dataset (GEL). We selected variants within near coding regions of MANE transcripts for genes flagged as potentially involved in each proband’s phenotype (PanelApp - ‘Green genes’), with a dominant mode of inheritance. Rare variants in each region were annotated with UTRannotator, SpliceAI scores, polyA signatures, miRNA binding sites, transcription factor binding sites, and other regulatory features.

We found a total of 480 rare de novo near-coding variants, in dominant genes considered diagnostic for a patient’s phenotype; 58 within 5’ UTR exons; 360 within 3’ UTR exons; 4 within 5’ splice regions (SpliceAI > 0.1); and 58 variants within the gene promoter (based on ENCODE cCREs). Of the 443 individuals with a candidate variant in this group, 361 (81%) do not have a diagnosis linked to a coding variant.

Likely disease-causing near-coding variants included previously identified variants in both *PAX6* and *MEF2C*, along with several novel diagnostic candidates. These include a 5’UTR splice donor variant in *SETD5* in a patient with intellectual disability (SpliceAI=0.97), and a 5’UTR variant that creates an upstream start-codon out-of-frame with the coding sequence of *SLC2A1* in a patient with motor difficulties and abnormal glucose homeostasis. Finally, using matched participants from the GEL cancer cohort as controls we performed burden testing in 29,759 rare disease probands to quantify the enrichment of variants in undiagnosed patients within each of our candidate near-coding elements.

It is common for individuals suffering from rare disease to be faced with, often prolonged diagnostic odysseys. Here, we demonstrate the power of incorporating near-coding regions into existing diagnostic pipelines to uncover additional diagnoses that are missed when using a coding-centric approach.

Mendelian Phenotypes Posters - Wednesday

PB1979. Taking the *CPLANE INTU* the light: further defining the orofacioidigital syndrome phenotype association with biallelic *INTU* alterations.

Authors:

J. Hunter^{1,2}, R. Rushforth^{2,1}, D. Nolan¹, D. Koboldt³, R. Stottmann³, M. Mori¹; ¹Nationwide Children's Hosp., Columbus, OH, ²The Ohio State Univ., Columbus, OH, ³Nationwide Childrens Hosp., Columbus, OH

Abstract Body:

Alterations in the *INTU* gene are provisionally associated with orofacioidigital syndrome XVII (OMIM: 617926, OFD17), short-rib thoracic dysplasia 20 with polydactyly, (OMIM: 617925, SRPS) and possibly nephronophthisis (NPHP). To date, only 4 patients from three families have been reported with an OFD phenotype, two patients with a SRPS phenotype and one patient with a NPHP phenotype. The OFD phenotype appears to consist of growth retardation, facial dysmorphisms, tongue hamartomas, polydactyly, cardiac malformations, and renal malformations. The SRPS phenotype is similar to the OFD phenotype plus short rib thoracic dysplasia and appears to be more severe or even lethal in utero. Finally, a single patient with a rare homozygous missense *INTU* alteration was reported to have only short stature and nephronophthisis. The *INTU* gene codes for the intuned protein, a critical subunit of the *CPLANE* protein complex that is required for proper cilia function and reported phenotypes are consistent with a ciliopathy. We present here a fifth patient with a phenotype consistent with OFD. This male patient presented prenatally with congenital heart defects and was found postnatally to have feeding difficulties, and early growth concerns. He has demonstrated developmental delay including motor and language delay. As with other reported patients, he has tongue hamartomas, bilateral great toe polysyndactyly, and slight facial dysmorphisms. Clinical trio exome sequencing revealed a paternally inherited missense alteration (p.Thr240Ile; c.719C>T) and a maternally inherited truncating alteration (p.Arg572Ter; c.1714C>T) in the *INTU* gene (transcript: NM_015693.4). We are investigating these alterations in an *in vitro* model to further determine their effects on ciliary form and function and confirm the suspected role of *INTU* in ciliopathies. This patient represents only the fifth reported patient with a phenotype consistent with OFD that harbors potentially pathogenic biallelic alterations in *INTU*. This OFD patient, along with *in vitro* and *in vivo* mouse models of disease being studied, will help bring this novel syndrome *INTU* the light and one step closer to clinical validity of the gene-disease relationship.

Mendelian Phenotypes Posters - Thursday

PB1980. Targeted deep sequencing analyses of long QT syndrome in a Japanese population.

Authors:

Y. Nagata¹, R. Watanabe¹, C. Eichhorn², S. Ohno³, T. Aiba⁴, T. ISHIKAWA⁴, Y. Nakano⁵, Y. Aizawa⁶, K. Hayashi⁷, N. Murakoshi⁸, T. Nakajima⁹, N. Yagihara¹⁰, H. Mishima¹¹, T. Sudo¹, C. Higuchi¹², A. Takahashi⁴, A. Sekine¹³, T. Makiyama¹⁴, Y. Tanaka¹⁵, A. Watanabe¹⁶, M. Tachibana¹⁷, H. Morita¹⁸, K-i. Yoshiura¹⁹, T. Tsunoda^{20,21}, H. Watanabe²², M. Kurabayashi²³, A. Nogami²⁴, Y. Kihara⁵, M. Horie²⁵, W. Shimizu²⁶, N. Makita⁴, T. Tanaka²⁷; ¹Tokyo Med. and dental Univ., Tokyo, Japan, ²Private Univ. of the Principality of Liechtenstein, Triesen, Liechtenstein, ³Natl. Cerebral and Cardiovascular Ctr, Suita, Japan, ⁴Natl. Cerebral and Cardiovascular Ctr., Suita, Japan, ⁵Hiroshima Univ., Hiroshima, Japan, ⁶Intl. Univ. of Hlth.and Welfare Narita Hosp., Chiba, Japan, ⁷Kanazawa Univ. Graduate Sch. of Med. Sci., Kanazawa, Japan, ⁸Univ. of Tsukuba, Tsukuba, Japan, ⁹Gunma Univ. Graduate Sch. of Med., Gunma, Japan, ¹⁰Niigata Univ. Graduate Sch. of Med. and Dental Sci., Niigata, Japan, ¹¹Nagasaki Univ., Nagasaki, Japan, ¹²Natl. Inst.s of BioMed. Innovation, Ibaraki, Japan, ¹³Chiba Univ., Chiba, Japan, ¹⁴Kyoto Univ., Kyoto, Japan, ¹⁵Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, ¹⁶Okayama Med. Ctr., Okayama, Japan, ¹⁷Sakakibara heart Inst. of Okayama, Okayama, Japan, ¹⁸Okayama Univ. Graduate Sch. of Med., Okayama, Japan, ¹⁹Nagasaki Univerisity Graduate Sch. of BioMed. Sci., Nagasaki, Japan, ²⁰Tokyo Med & Dental Univ, Tokyo, Japan, ²¹RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, ²²Niigata Univ. Graduate Sch. of Med. and Dental Sci., Yokohama, Japan, ²³Gunma Univ. Graduate Sch. of Med. Maebashi, Gunma, Japan, ²⁴Univ. of Tsukuba, Tshukuba, Japan, ²⁵Shiga Univ. of Med. Sci., Otsu, Japan, ²⁶Nippon Med. Sch., Tokyo, Japan, ²⁷Tokyo Med. and Dental Univ., Tokyo, Japan

Abstract Body:

Long QT syndrome (LQTS) is one of the most common inherited arrhythmias and multiple genes have been reported as causative. Presently, genetic diagnosis for LQTS patients is becoming widespread and contributing to implementation of therapies. However, causative genetic mutations cannot be detected in about 20% of patients. To elucidate additional genetic mutations in LQTS, we performed deep-sequencing of previously reported 15 causative and 85 candidate genes for this disorder in 677 Japanese LQTS patients. We performed in-silico filtering of the sequencing data and found 60 novel variants in 40 genes of 71 cases. These variants were predicted to be damaging to coding proteins or to alter the binding affinity of several transcription factors. They might be useful for screening of LQTS patients who had no known genetic factors. In addition, when the mechanisms of these variants in the development of LQTS are revealed, it will be useful for early diagnosis, risk stratification, and selection of treatment.

Mendelian Phenotypes Posters - Wednesday

PB1981. *TCEAL1* loss-of-function results in an X-linked dominant neurological syndrome and drives the neurological disease trait in Xq22.2 deletion

Authors:

H. Hijazi¹, J. Posey¹, L. Reis², D. Pehlivan^{1,3,4,5}, J. Bernstein⁶, M. Muriello², E. Syverson², D. Bonner⁶, M. A. Estiar^{7,8}, Z. Gan-Or^{7,8,9}, G. A. Rouleau^{7,8,9}, K. Lyulcheva-Bennett¹⁰, L. Greenhalgh¹⁰, M. Tessarech^{11,12}, E. Colin^{11,12}, A. Guichet¹¹, D. Bonneau^{11,12}, R. van Jaarsveld¹³, A. M. A. Lachmeijer¹³, L. Ruaud^{14,15}, J. Levy¹⁵, A-C. Tabet¹⁵, R. Ploski¹⁶, M. Rydzanicz¹⁶, L. Kepczynski¹⁷, K. Polatynska¹⁸, Y. Li¹, J. M. Faith¹, D. Marafi^{1,19}, J. A. Rosenfeld^{1,20}, Z. H. C. Akdemir^{1,21}, W. Bi^{1,20}, R. A. Gibbs^{1,22}, G. M. Hobson²³, J. V. Hunter²⁴, C. Carvalho^{1,25}, E. V. Semina^{2,26}, J. Lupski^{1,22,27}; ¹Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, ²Dept. of Pediatrics and Children's Res. Inst., Med. Coll. of Wisconsin and Children's Wisconsin, Milwaukee, WI, ³Section of Pediatric Neurology and Dev.al NeuroSci., Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, ⁴Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, ⁵Jan and Dan Duncan Neurological Res. Inst. at Texas Children's Hosp., Houston, TX, ⁶Dept. of Pediatrics, Div. of Med. Genetics, Stanford Sch. of Med., Stanford, CA, ⁷Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, ⁸The Neuro (Montreal Neurological Inst.-Hosp.), McGill Univ., Montreal, QC, Canada, ⁹Dept. of Neurology & Neurosurgery, McGill Univ., Montreal, QC, Canada, ¹⁰Liverpool Ctr. for Genomic Med., Liverpool Women's Hosp., Liverpool, United Kingdom, ¹¹Dept. of Med. Genetics, Angers Univ. Hosp., Angers, France, ¹²Mitovasc Unit, UMR CNRS 6015-INSERM 1083, Univ. of Angers, Angers, France, ¹³Dept. of Genetics, Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, ¹⁴INSERM UMR1141, Neurodiderot, Univ. of Paris, F-75019, Paris, France, ¹⁵APHP.Nord, Robert Debré Univ. Hosp., Dept. of Genetics, F-75019, Paris, France, ¹⁶Dept. of Med. Genetics, Med. Univ. of Warsaw, Warsaw, Poland, ¹⁷Dept. of Genetics, Polish Mother's Mem. Hosp. – Res. Inst. (PMMH-RI), Lodz, Poland, ¹⁸Dept. of Dev.al Neurology an Epileptology, Polish Mother's Mem. Hosp. – Res. Inst. (PMMH-RI), Lodz, Poland, ¹⁹Dept. of Pediatrics, Faculty of Med., Kuwait Univ., Safat, Kuwait, ²⁰Baylor Genetics, Houston, TX, ²¹Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ²²Human Genome Sequencing Ctr. and Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, ²³Nemours BioMed. Res., Nemours Children's Hlth., Wilmington, DE, ²⁴Dept. of Pediatric Radiology, Texas Children's Hosp., Houston, TX, ²⁵Pacific Northwest Res. Inst., Seattle, WA, ²⁶Dept.s of Ophthalmology and Visual Sci. and Cell Biology, Neurobiology and Anatomy, Med. Coll. of Wisconsin, Milwaukee, WI, ²⁷Dept. of Pediatrics, Texas Children's Hosp., Houston, TX

Abstract Body:

An Xq22.2 region upstream of *PLP1* has been proposed to underly a neurological disease trait when deleted in 46,XX females. Deletion mapping revealed that heterozygous deletions with a smallest region of overlap (SRO) spanning six Xq22.2 genes (*BEX3*, *RAB40A*, *TCEAL4*, *TCEAL3*, *TCEAL1*, *MORF4L2*) associate with an early-onset neurological disease trait (EONDT) consisting of hypotonia, intellectual disability (ID), neurobehavioral abnormalities, and dysmorphic facial features. To date, none of the genes within the smallest region of overlap (SRO) have been associated with monogenic disease. Through local and international collaborations facilitated by GeneMatcher and MatchMaker Exchange, we have identified and herein report seven novel *de novo* variants involving *TCEAL1* in seven unrelated families: three hemizygous truncating alleles; one hemizygous missense allele; one heterozygous *TCEAL1* full gene deletion; one heterozygous contiguous deletion of *TCEAL1*, *TCEAL3*, *TCEAL4*; and one heterozygous frameshift variant allele. Variants were identified through exome or genome sequencing with trio analysis or through chromosomal microarray. Comparison with previously reported Xq22 deletions encompassing *TCEAL1* identified a more-defined novel syndrome consisting of hypotonia,

ASHG 2022 Annual Meeting Poster Abstracts

abnormal gait, developmental delay/intellectual disability especially affecting expressive language, autistic-like behavior, and mildly dysmorphic facial features. Additional features include strabismus, refractive errors, variable nystagmus, gastroesophageal reflux, constipation, dysmotility, recurrent infections, seizures, and structural brain anomalies. An additional maternally inherited hemizygous missense allele of uncertain significance was identified in a male with hypertonia and spasticity without syndromic features. These data provide evidence that *TCEALI* loss-of-function causes a neurological rare disease trait involving significant neurological impairment with features overlapping the EONDT phenotype in females with the Xq22 deletion.

Mendelian Phenotypes Posters - Thursday

PB1982. Temtamy Preaxial Brachydactyly Syndrome: penetrance in heterozygotes for a *CHSY1* missense variant.

Authors:

P. Sabeh¹, M. Valancy^{2,3}, C. Janelle³, P. M. Campeau^{2,3}; ¹Université de Montréal, Montréal, QC, Canada, ²CHU Sainte-Justine, Montréal, QC, Canada, ³Hôpital Shriners, Montréal, QC, Canada

Abstract Body:

Temtamy Preaxial Brachydactyly Syndrome (TPBS) is an autosomal recessive dysostosis. It is characterized by bilateral preaxial brachydactyly with facial dysmorphism, hyperphalangism, developmental delay or intellectual disability, growth retardation, sensorineural hearing loss and dental abnormalities. Loss of function variants in *CHSY1*, encoding a secreted chondroitin glycosyltransferase, cause TPBS by dysregulating NOTCH and BMP signaling. Here, we report partial penetrance in heterozygotes for a *CHSY1* missense variant. A father and his daughter present the same heterozygous variant (NM_014918.5:c.1386A>C, p.Lys462Asn) in the *CHSY1* gene associated with a milder TPBS phenotype, including hallux valgus, brachymetatarsia, mild brachydactyly, significant clinodactyly and limited metacarpophalangeal flexion. They did not have facial dysmorphisms, cognitive involvement, teeth anomalies nor hearing loss. No other variants were identified on a limb anomaly sequencing panel or on exome sequencing. The variant changes a well conserved lysine amino acid residue to an asparagine, thus changing the charge at this position of the glycosyltransferase domain. The variant is absent from gnomAD, and has a CADD score of 23, thus placing it in the top 1% most deleterious variants. In conclusion, we identified a heterozygous missense variant in *CHSY1* which is likely causally related to the hand and feet anomalies dominantly transmitted in this family, suggesting partial penetrance in carriers, a phenomenon noted in other recessive conditions.

Mendelian Phenotypes Posters - Wednesday

PB1983. TGF β signaling is important for spermatogenesis.

Authors:

U. Kumar, Y-C. Shen, C. Sultan, S. Hammoud; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

The process of spermatogenesis requires intricate interactions between the germline and soma (Leydig, Sertoli, and peritubular myoid cells). While many signaling pathways in Leydig and Sertoli cells have been implicated in various aspects in germ cell maintenance and differentiation, relatively little is known about the contribution of peritubular myoid cells (PMC). To investigate possible signaling pathways between PMCs and germ cells, we performed ligand-receptor analysis using our mouse testis scRNA-seq data and identified TGF β 2 and ALK5 as a high confidence ligand receptor pair. TGF β 2 ligand is expressed in PMCs and TGFBR1 receptor (ALK5) is expressed in meiotic and post-meiotic cells. The Myoid-Meiotic cell communication axis was unexpected since these cells are above the blood-testis barrier. To dissect the role of TGF β R signaling in germ cell development, we initially knocked out the ALK5 receptor in differentiating spermatogonia using a Stra8-Cre. The ALK5 conditional knockout mice have 20% lower testis/body weight ratio as compared to ALK5^{+/+} males, suggestive of spermatid cell loss. As expected, no significant difference was noted in the number of undifferentiated (PLZF⁺), differentiating spermatogonia numbers (Stra8⁺), or the number of proliferating STRA8⁺/PLZF⁺ cells in ALK5^{-/-} males, but we did get a significant decrease in the ratio of round to elongating spermatid in the testis. Together, these results suggest that TGF β signaling through the ALK5 receptor in germ cells is important for reproductive fitness and future studies will examine the mechanisms by which TGF β signaling maintains optimal steady state spermatogenesis.

Mendelian Phenotypes Posters - Thursday

PB1984. The benefits and yields of whole exome sequencing data reanalysis in Iranian undiagnosed patients with intellectual disability

Authors:

Z. Fattahi^{1,2}, M. Babanejad¹, F. Peymani¹, F. Ghodratpour^{1,2}, M. Beheshtian^{1,2}, S. Arjangi¹, M. Mohseni^{1,2}, S. Banihashemi¹, F. Larti¹, K. Kahrizi¹, H. Najmabadi^{1,2}; ¹Genetics Res. Ctr., Univ. of Social Welfare and Rehabilitation Sci., Tehran, Iran, Islamic Republic of, ²Kariminejad-Najmabadi Pathology & Genetics Ctr., Tehran, Iran, Islamic Republic of

Abstract Body:

Nowadays, molecular diagnosis is enhanced by implementing whole-exome sequencing (WES) in studying Mendelian disorders, especially in intellectual disability (ID). Although WES is considered a first-line genetic analysis tool, the diagnostic yield has been still limited to less than 50% in different studies. Over recent years, reanalysis of WES data has allowed genetic diagnosis in patients with negative results from the initial evaluation and has increased the molecular diagnosis in 10-15% of patients. This can be obtained by updates in bioinformatics tools, algorithms, databases, literature, and patient clinical features. Furthermore, simultaneous WES-based detection of CNVs plays an important role as CNVs have a high positive rate with regard to ID. To evaluate the added value of WES reanalysis in molecular diagnosis of autosomal recessive intellectual disability (ARID) in patients with a high rate of consanguineous background, we reanalyzed WES data of 162 negative cases. This cohort was obtained from our previous large-scale study in which whole-exome and whole-genome sequencing had been performed in 404 consanguineous Iranian families with final detection rate of 54%. WES reanalysis was performed by applying the most updated versions of algorithms for variant detection and annotation, as well as WES-based CNV analysis using the GATK GermlineCNVCaller. Thus far, WES reanalysis of all 162 patients has been accomplished, of which the possible causative variant is detected in 17 cases, equal to a 10.5% increase in the detection rate. As the WES-based CNV analysis for all the 162 patients is still in process, we expect to see a higher detection rate in the future. Reanalysis identified pathogenic/likely pathogenic variants in known ID genes (*CDK10*, *ASPM*, *RBMX*, *VARSI*, *MANBA*, and *HSD17B4*) in seven cases. In addition, a duplication on Xq13.2q13.3 in a family with an established X-linked inheritance pattern was detected by linkage analysis and whole-genome sequencing. We also identified the second case with likely pathogenic variants in *FSCN1* and *AGBL2* genes, which were previously reported in similar patients once in the literature. Furthermore, novel candidate genes were detected in six cases: *AKTIP*, *MAZ*, *MIX23*, *SERPINB12*, *CDC25B*, *TTII*, and *GPR126*. Updates in the genetic databases and literature and patients' clinical phenotype were among the main reasons for the ascertained results. In conclusion, this project has shown a 10.5% improvement in detection yield, introduced seven possible novel candidate genes, and provided a second report for the recently identified genes providing a more precise genotype-phenotype correlation.

Mendelian Phenotypes Posters - Wednesday

PB1985. The ClinGen Kidney Disease Clinical Domain Working Group: Defining solutions to the challenge of nephropathy ontology and nomenclature.

Authors:

A. Byrne¹, H. Dziadzio¹, E. Edoh¹, A. Girod¹, R. Webb¹, T. P. Sneddon², M. G. Sampson^{3,4}, A. J. Mallett^{5,6}; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²UNC Hlth., Chapel Hill, NC, ³Boston Children's Hosp., Boston, MA, ⁴Harvard Med. Sch., Boston, MA, ⁵Townsville Hosp. and Hlth.Service, Townsville, Australia, ⁶The Univ. of Queensland, Brisbane, Australia

Abstract Body:

Kidney disease affects one in ten people with 10% of affected individuals having a monogenic form. The ClinGen Kidney Disease Clinical Domain Working Group (CDWG) was established in 2019 with the goal of creating a comprehensive, standardized knowledge base defining the clinical validity of gene-disease relationships and expert classifications of variant pathogenicity across the full spectrum of nephropathies in order to improve patient care. The most significant challenge to this goal however, lies in defining the disease entity. For many nephropathies, current disease nomenclature (e.g., OMIM) and ontology (e.g., MonDO) systems do not accurately reflect the underlying biology of a given disorder and are not aligned with contemporary knowledge of disease pathogenesis. Inaccuracies and inconsistencies in disease naming can therefore hinder clinical recognition and impact patients by restricting eligibility and access to appropriate testing, monitoring, and treatment. The ClinGen Kidney Disease CDWG, in coordination with the ClinGen Disease Naming Working Group, are currently working to define consensus recommendations for kidney disease naming, with input from representatives across the global renal community, including clinicians, scientists, and patients. Naming is suggested to follow the general, stepwise system of 'core disease name - gene name', with disease distinguishing phenotype and/or genotype descriptors added where appropriate to differentiate related disorders with well-established differences in e.g., aetiology, presentation, management, or outcome. Implementing these changes across the multiple related, but not interdependent, ontology and nomenclature systems will likely be a challenging undertaking that requires agreement and support from the community, but is one that is expected to have immense benefits in clinical practice.

Mendelian Phenotypes Posters - Wednesday

PB1986. The genetic and phenotypic landscape of *STUB1/TBP*-associated disorders: Two genes and three patterns of inheritance sharing a continuous neurodegenerative phenotypic spectrum.

Authors:

S. Magri¹, L. Nanetti¹, C. Gellera¹, E. Sarto¹, M. Balzo¹, E. Rizzo¹, A. Mongelli¹, G. Di Fede², C. Mariotti¹, D. Di Bella¹, F. Taroni¹; ¹Fondazione IRCCS Istituto Neurologico Carlo Besta, Unit of Med. Genetics and Neurogenetics, Milan, Italy, ²Fondazione IRCCS Istituto Neurologico Carlo Besta, Div. of Neurology 5 and Neuropathology, Milan, Italy

Abstract Body:

Spinocerebellar ataxias type 17 (SCA17) and type 48 (SCA48) are both characterized by cerebellar-cognitive-behavioural features and incomplete penetrance. While SCA17 is caused by a CAG/CAA (polyQ) repeat expansion in the *TBP* gene, with a full penetrance for alleles with >49 repeats and a reduced penetrance for intermediate 41-49 alleles, SCA48 is attributed to heterozygous pathogenic variants in the *STUB1* gene. Notably, biallelic *STUB1* pathogenic variants cause a recessive ataxia (SCAR16) characterized by a SCA17/SCA48-like phenotype which manifests only in homozygotes or compound heterozygotes. We recently demonstrated a digenic inheritance of the *STUB1/TBP* genotype which explains the incomplete penetrance in SCA17 and SCA48, showing that SCA17 is a monogenic disorder for *TBP* expansions with >47 polyQ and a "true digenic" *TBP/STUB1* disorder for intermediate *TBP* alleles (41-46 polyQ) (Genet Med 2022, PMID 34906452). Here, we present new insights to address questions concerning 1) the pathogenic role of "high-normal" 39-40 repeat *TBP* alleles; 2) the discrepancy between the expected genetic prevalence of the *STUB1/TBP* genotype and the reported disease prevalence of *STUB1/TBP* associated-diseases; 3) the phenotypic overlap in monogenic and digenic SCA17, and SCA48. Analysis of 170 index cases with ataxia carrying "high-normal" 39- and 40-repeat *TBP* alleles identified 10 novel families carrying a *STUB1* pathogenic heterozygous variant (5.9%), supporting a pathogenic role for 39-40 alleles, previously considered to be benign. Interestingly, screening of 94 index cases primarily referred for dementia revealed 18 cases with *TBP*₃₉₋₄₀ alleles, indicating a slight enrichment of these "high-normal" alleles in this group of patients. Notably, a *STUB1* pathogenic heterozygous variant was identified in 3 of these cases, suggesting that the *STUB1/TBP* genotype may underlie a wider spectrum of phenotypes. Comparison of clinical and neuroimaging findings reveal that an earlier and more severe cognitive deterioration and a more widespread pattern of cerebellar atrophy characterizes the SCA17^{Digenic} phenotype in comparison with SCA17 monogenic disease. By contrast, the highly homogeneous MRI morphological findings in SCA48/SCA17^{Digenic} and SCAR16 patients strongly emphasizes the existence of a continuous clinical-neuroimaging and genetic spectrum in *STUB1* gene-related disorders. Due to the relatively high frequency of *TBP*₄₀₋₄₆ (f=0.015) and *TBP*₃₉ (f=0.088) alleles, and that of *STUB1* pathogenic variants (f=0.003) in the general population, SCA17^{Digenic} represents a very demanding task for genetic counseling. (Funding: FRRB-CP-20/2018 to FT)

Mendelian Phenotypes Posters - Thursday

PB1987. The genetic background in a group of 155 Polish pediatric patients with unexplained epilepsy.

Authors:

B. Chalupczynska¹, E. Ciara¹, M. Rydzanicz², P. Stawiński^{1,2}, D. Siestrzykowska¹, P. Halat-Wolska¹, D. Jurkiewicz¹, M. Pelc¹, D. Piekutowska-Abramczuk¹, A. Madej-Pilarczyk¹, D. Wicher¹, R. Płoski², K. Chrzanowska¹; ¹The Children's Mem. Hlth.Inst., Warsaw, Poland, ²Warsaw Med. Univ., Warsaw, Poland

Abstract Body:

Introduction: The studies conducted so far indicate a high clinically and genetically heterogeneous of epilepsy syndromes. A large number of genes, with wide ranging functions, are implicated in their etiology. Next-generation sequencing (NGS) has contributed to the identification of many monogenic epilepsy syndromes and is favoring earlier and more accurate diagnosis in a group of pediatric patients with epilepsy.

Materials and Methods: In the present study, molecular profile analysis of pediatric patients with a variety of epilepsy types was performed using NGS based multi-gene targeted panels, containing the majority of epileptic significant genes or whole-exome sequencing (WES).

Results: Molecular profile analysis was conducted in 155 epileptic patients. A total of 85 pathogenic/likely pathogenic variants in 53 genes were identified, including 32 known and 53 novel alterations. The most commonly mutated was *SCN1A* gene, associated with Dravet syndrome. Sporadically variants in *ALG13*, *AMT*, *ANKRD11*, *ATPIA3*, *CACNA1A*, *CDKL5*, *CHD2*, *CLN6*, *COL4A1*, *CTSD*, *CUX2*, *EFHC1*, *EHMT1*, *FGFR3*, *FLNA*, *GABRA1*, *GABRG2*, *GALC*, *GAMT*, *GRIN2A*, *ITPA*, *KANSL1*, *KCNA2*, *KCND3*, *KCNJ11*, *KCNQ2*, *KMT2D*, *LMBRD2*, *MBD5*, *MECP2*, *PCDH19*, *PIGN*, *POMT2*, *PTCH1*, *RYR2*, *SCN1B*, *SCN2A*, *SCN3A*, *SCN8A*, *SETBP1*, *SLC16A2*, *SMARCA2*, *SMARCB1*, *SMCIA*, *SPTAN1*, *STXBPI*, *SYNI*, *SYNGAP1*, *TPP1*, *TUBA1A*, *TUBB4A*, *WDR45* were identified. The diagnostic yield of the entire study cohort was 45%.

Conclusions: Targeted massively parallel sequencing is an effective diagnostic test for pediatric patients with epilepsy, with over 40% diagnostic efficiency. The *SCN1A* it is the most essential factor in the etiology of epilepsy in the studied group (12%). The remaining genes were observed in single cases, which confirms the high genetic heterogeneity. Patients with earlier seizure onset and drug-resistance epilepsy were the best candidates for testing. Our study contributes to further delineation of the epilepsy molecular profile, and for pediatric patients with epilepsy, genetic diagnosis is important for accurate prognosis and treatment.

Partially supported by CMHI project: M39/2019

Mendelian Phenotypes Posters - Wednesday

PB1988. The genetic landscape of *ATP7B* provides insights into Wilson Disease variant penetrance and prevalence.

Authors:

G. del Angel¹, W. Mowrey¹, J. Bissonette², L. Chunn², M. Kiel², D. Nefcy², S. Wasilewski³, T. Defay¹; ¹Alexion AstraZeneca Rare Disease, Boston, MA, ²Genomenon, Ann Arbor, MI, ³AstraZeneca, Cambridge, United Kingdom

Abstract Body:

Wilson Disease (WD) is a monogenic disorder of copper metabolism caused by biallelic loss-of-function variants in the gene *ATP7B*. A comprehensive understanding of the *ATP7B* genetic landscape is thus required for accurate estimation of WD prevalence and diagnosis rates. The UK Biobank (UKB), a public biomedical database aggregating genetic and clinical data for ~500,000 individuals, enables the discovery of novel *ATP7B* variants and the resolution of ambiguous variant interpretations.

We identified 1,211 protein-altering *ATP7B* variants in UKB Whole-Exome Sequencing (WES), including 37 novel protein-truncating variants. We assessed these variants' pathogenicity using ClinVar and Genomenon's Mastermind. These resources apply ACMG criteria (Richards et al., 2015) to complementary sets of variant data: where ClinVar archives submitted reports of variant clinical significance, Mastermind archives published literature reports of variants. ClinVar and Mastermind identified 140 and 227 variants, respectively, as potentially pathogenic. Subjects biallelic for pathogenic *ATP7B* variants were rare in UKB (0.03%), though subjects diagnosed with disorders of copper metabolism were even more rare (0.003%). We found zero overlap between *ATP7B* biallelic subjects and possible WD cases. Only one subject in UKB had additional diagnoses indicative of WD, including tremor, anemia, and corneal pigmentation. This subject was homozygous for a series of common *ATP7B* polymorphisms, but had no pathogenic variants that were identifiable by WES.

The absence of WD phenotypes in other biallelic subjects were attributable to pathogenic *ATP7B* variants with low-penetrance, or that were in *cis* (and thus not biallelic). The variants c.2972C>T and c.122A>G, identified as low-penetrance by Wallace and Dooley (2020), contributed to 111 of 128 biallelic cases. A variant pair present in another 8 biallelic patients, c.524_525del and c.778dup, was found to be in *cis* as a single pathogenic allele. Excluding these problematic cases left just 9 putative biallelic patients (0.002%) with 13 unique variants. The lack of WD phenotype in these subjects suggests that they may represent additional cases of low-penetrance *ATP7B* variants.

Our findings show that UK Biobank is an important resource for defining the genetic landscape of *ATP7B*, revealing both novel pathogenic variants and low penetrance in variants assumed to be pathogenic. Targeted investigation of these variants will improve our understanding of WD genetics, as well as the quantification of WD prevalence and diagnosis.

This research has been conducted using the UK Biobank Resource under application number 26041.

Mendelian Phenotypes Posters - Thursday

PB1989. The genetic landscape of neurodevelopmental disorders in a large cohort of multiplex consanguineous families from Turkey.

Authors:

A. Reis¹, E. Gümüşlü¹, E. Gümüş², Ö. Öz³, M. Özkan⁴, K. Kararer⁵, A. Ekici⁶, E. Pembegül Yildiz⁷, N. Aydinli⁷, A. Rauch⁸; ¹Friedrich-Alexander-Univ. Erlangen-Nürnberg, Erlangen, Germany, ²Mugla Sitki Kocman Univ., Mugla, Turkey, ³Harran Univ., Sanliurfa, Turkey, ⁴Medeniyet Univ., Istanbul, Turkey, ⁵Pamukkale Univ., Pamukkale, Turkey, ⁶Friedrich-Alexander-Univ. Erlangen-Nürnberg, Erlangen, Germany, ⁷Istanbul Univ. Sch. of Med., Istanbul, Turkey, ⁸Univ Zurich, Schlieren-Zurich, Switzerland

Abstract Body:

Neurodevelopmental disorders (NDDs) are genetically and phenotypically highly heterogeneous. Autosomal recessive (AR) forms are enriched in families from consanguineous marriages, but these families can present all other forms of inheritance, making a diagnosis challenging. In recent years, diagnostic yield increased from 40% to more than 60% in a recent study, with *de novo* and X-linked causes in 8-19% of diagnosed cases.

We now investigated a large cohort of consanguineous families with multiple affected individuals with NDDs in order to investigate the genetic landscape focusing on diagnostic consistency within families, which is commonly assumed in such families. We recruited a total of 183 affected individuals from 82 consanguineous families with two or more children with NDDs through clinical genetic and neuropsychiatric consultations from various academic hospitals in Turkey. Exome sequencing in all affected individuals and their parents allowed identification of pathogenic or likely pathogenic variants for 116 individuals in 85 established or known candidate genes for NDDs. This diagnostic yield of 63% is in line with the most recent report, but was only achieved when also sequencing parents, compared to only 47% when studying affected siblings only. Progress in the field becomes apparent by 8% of diagnoses based on genes published in the last 18 months. In 34 individuals (18.5%) from 23 families, we identified previously described pathogenic variants in 19 genes, indicating a high rate of founder mutations. Similarly to previous studies, we found *de novo* variants and variants in X-linked genes in 19 cases (10%), and in 54 individuals (29.5%) we found multiple molecular NDD diagnoses. Unexpectedly, only in 14 families (17%) our hypothesis of shared homozygous variants among all affected individuals of that family could be confirmed, as most families presented different molecular diagnoses.

Our study shows that multiplex consanguineous families often represent chance familial aggregation of different genetic disorders and modes of inheritance, requiring individual analysis of each affected child and inclusion of parents in the analysis. Only this way, a diagnostic yield of >60% can be achieved. Founder effects with 19% of cases are relatively frequent, a finding that is only now becoming evident given the growing amount of variant data available. Similarly, progress in gene identification now allows detecting multiple molecular diagnoses, which were seen in 30% of cases, a much higher frequency than in individuals from unrelated parents. Further investigation of novel candidate genes is ongoing and will help further increase the diagnostic yield.

Mendelian Phenotypes Posters - Wednesday

PB1990*. The impact of NGS on rare and complex epilepsies: identification of novel causative variants for an efficient diagnostic pathway and personalized treatments

Authors:

C. Gellera¹, B. Castellotti¹, S. Magri¹, G. Messina¹, F. Ragona¹, E. Freri¹, R. Solazzi¹, L. Canafoglia¹, M. Soldovieri², P. Ambrosino³, I. Mosca², M. Tagliatela⁴, F. Taroni¹, J. C. Di Francesco⁵, T. Granata¹; ¹Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, ²Univ. of Molise, Campobasso, Italy, ³Univ. of Sannio, Benevento, Italy, ⁴Univ. of Naples Federico II, Naples, Italy, ⁵Università Milano Bicocca, Monza, Italy

Abstract Body:

Epilepsy affects 1% of children and is the most frequent chronic neurological condition in childhood. The etiology remains unknown in 50% of cases, although it has been estimated that the large majority have a genetic determinant. The systematic use of NGS technology has allowed a) to increase the diagnostic rate in patients with rare phenotypes; b) to identify new pathogenetic variants; and c) to describe new related mechanisms. In recent years, the number of genes causative of severe and complex generalized epilepsies and developmental epileptic encephalopathies has increased significantly. We started using targeted gene panels with ≤ 100 genes to the current one with 295 genes. In total, we have analyzed 697 patients from the clinical, instrumental, radiological, and genetic point of view. Our studies led to a molecular diagnosis in 151 patients (22%) (81 carrying class 5 variants, and 70 class 4 variants); while 152 patients (22%) had class 3 (VoUS) variants and 394 patients (56%) did not carry variants of possible pathogenic significance. The most frequently mutated genes encoded channel proteins: 13.4% in SCNs channel genes and 12.7% in KCNs channel genes. Interestingly, 7% of causative identified variants belong to genes coding for SLC family neuronal transporter proteins. Among the 390 negative patients, we selected 37 families for WES analysis (31 sporadic cases, 4 with an autosomal recessive trait and 2 autosomal dominant). We reached a genetic diagnosis in 6 families (16%), 4 sporadic cases, and 2 autosomal recessive families, and we identified variants of possible pathogenic significance and compatible with the patient's phenotype in 4 additional cases (11%). Through in vitro electrophysiological studies in CHO cells, we demonstrated the pathogenic role of some new variants in potassium channel genes and evaluated in vitro the ability of molecules acting on specific channels, such as retigabine or gabapentine (KCNQ2, KCNQ3), quinidine (KCNT1), and fluoxetine (KCNC1) to revert variant-induced channel dysfunction. Notably, in most cases, the use of these drugs in patients carrying specific variants prompts a significant clinical improvement, thereby highlighting the efficacy of this approach for personalized treatment approaches.

Mendelian Phenotypes Posters - Thursday

PB1991. The mechanism and current concepts and future research direction of rare hereditary disease
White Sponge Nevus

Authors:

W. Cai; Tongji Univ. Sch. of Med., Shanghai, China

Abstract Body:

White sponge nevus (WSN) is an autosomal dominant rare hereditary disease. Keratin 4 (*KRT4*) and Keratin 13 (*KRT13*) gene mutations were involved in the WSN. In this study we will explore the pathogenesis of WSN. We recruited two WSN Chinese families, The mutations of *KRT4* and *KRT13* gene were detected by PCR and direct sequencing. The multiple alignments of *KRT13* from 28 diverse species homology analysis were performed by the ClustalW program. The *KRT13* expression was measured by Real-Time PCR and Western blot analysis. We revealed the pathological pathway of the WSN expression profile by RNA sequencing (RNA-seq). Sequencing analysis revealed two mutations of *KRT13* gene: one mutation was 332T>C (Leu111Pro). Another mutation was 340C>T (Arg114Cys). The sequence of *KRT13* was highly conserved. Real-Time PCR and Western blot analysis results show that *KRT13* expression level is lower in patient but keep almost no change in mRNA level. When treat cells with MG132, *KRT13* protein level was increased and kept almost the same in normal and patient cell. We identified two heritable mutations in the *KRT13* gene, which were associated with the development of WSN. The abnormal degradation of *KRT13* protein of WSN may probably associate with abnormal ubiquitination process. Further RNA-seq analysis suggests that the abnormal degradation of *KRT13* protein of WSN may probably associate with Keratin 7 (*KRT7*) and abnormal ubiquitination process. All the efforts may contribute to the molecular therapy for WSN. Gene-based diagnosis and gene therapy for WSN may become available in the near future while provide reference and instruction for treating other keratin associated diseases.

Mendelian Phenotypes Posters - Thursday

PB1992. The Spectrum of PLEC Sequence Variants and Related Plectinopathies Including Novel Association with Epidermolysis Bullosa Pruriginosa

Authors:

J. Uitto¹, H. Vahidnezhad², L. Youssefian¹, H. Mahmoudi³, A. Saeidian⁴; ¹Thomas Jefferson Univ, Philadelphia, PA, ²Thomas Jefferson Univ., Philadelphia, PA, ³Tehran Univ. of Med. Sci., Tehran, Iran, Tehran, Iran, Islamic Republic of, ⁴Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Plectin, a cytoskeletal linker and intermediate filament protein ubiquitously expressed in many cell types, is encoded by PLEC. Plectin consists of three main domains that determine its functionality: the Nterminal domain, the rod domain, and the C-terminal domain. Molecular defects in PLEC correlating with the functional aspects lead to a group of rare heritable disorders, plectinopathies. These multisystem disorders include an autosomal dominant form of epidermolysis bullosa simplex (EBS-Ogna), limb girdle muscular dystrophy (LGMD), and an autosomal recessive form of epidermolysis bullosa simplex (EBS), which may associate with muscular dystrophy (EBS-MD), pyloric atresia (EBS-PA), and/or congenital myasthenic syndrome (EBS-MyS). In this study, genotyping of over 800 Iranian patients with epidermolysis bullosa by a gene-targeted sequencing panel or whole-exome sequencing identified 15 patients with disease-causing PLEC variants. Analysis of the clinical spectrum of plectinopathies in our cohort and in the literature indicated a relationship between the variant locations in PLEC and phenotypic manifestations. Most EBS-MD patients are found to have at least one variant in the central rod domain of the plectin and/or the C-terminal domain, whereas EBS, EBS-PA, and EBS-MD typically present with variants outside these domains. In addition, a novel association of a homozygous nonsense variant in PLEC was observed in a patient with EB pruriginosa, MD, and MyS. This study integrated our seven novel PLEC variants with previously published data totaling 115 variants to provide the most complete overview of pathogenic PLEC variants and related disorders.

Mendelian Phenotypes Posters - Wednesday

PB1993. Three cases of (likely) pathogenic *FGF10* variants: Examples of phenotypic variability in Lacrimoauriculodentodigital syndrome.

Authors:

M. Drost, S. Demirdas, S. Kant, M. Wessels, L. Hoefsloot; Dept. of Clinical Genetics, ErasmusMC, Rotterdam, Netherlands

Abstract Body:

Background Lacrimoauriculodentodigital syndrome (LADD) is an autosomal dominant condition caused by germline loss-of-function variants in *FGF10* or by tyrosine kinase domain variants in either *FGFR2* or *FGFR3*. The syndrome is characterized by malformed ears and deafness, various dental and digital anomalies and, specifically for *FGF10*, aplasia of the lacrimal duct and/or gland. Here we present three cases of LADD with great phenotypic variability of this syndrome.

Case presentation **Case 1.** A five-year old girl presented with an accessory auricle, conductive right-sided hearing loss and mixed left-sided hearing loss. Whole exome sequencing (WES) followed by targeted analysis of the LADD genes revealed a variant of uncertain significance, NM_004465.1(*FGF10*):c.232C>A, p.(Arg78Ser). Variant segregation was confirmed in four affected family members, with phenotypes including absence of the nasolacrimal duct and/or unilateral hearing loss. This led to reclassification of the variant to a likely pathogenic variant, confirming LADD diagnosis. **Case 2.** A 21-year old man presented with congenital defects including cholestatic liver disease and dysmorphic features including a broad forehead. WES followed by filtering for genes included in a multiple congenital anomaly panel revealed heterozygosity for two variants in the *ABCB4* gene, confirming a diagnosis of progressive familial intrahepatic cholestasis. WES also revealed a pathogenic variant of maternal origin, NM_004465.1(*FGF10*):c.55del, p.(Cys19Aafs*95). LADD diagnosis was considered an incidental finding in this case, reverse phenotyping of both index and mother revealed only minor features of LADD. **Case 3.** A 2-day old baby presented with an accessory auricle, absence of the nasolacrimal duct and severe hypoplasia of the lungs for which extracorporeal membrane oxygenation was required. The baby died after 2 months. WES followed by filtering for genes included in a multiple congenital anomaly panel revealed a pathogenic variant of paternal origin, NM_004465.1(*FGF10*):c.497C>A, p.(Ser166*). *FGF10* phenotypes combined with neonatal lethal pulmonary hypoplasia were described to be additionally associated with deletions or single-nucleotide variants involving *TBX4* (PMID 30639323), which could not be detected in this case. Absence of the nasolacrimal duct was also found in his father, and family members with LADD were only mildly affected.

Discussion We present three LADD cases with associated inter-and intrafamilial variability. Although this syndrome is inherited in an autosomal dominant fashion, its manifestation can differ greatly, stressing the need for caution in genetic counseling.

Mendelian Phenotypes Posters - Thursday

PB1994. Three new founder Bukharian Jewish mutations in *COL4A4* are causative of autosomal dominant and recessive Alport syndrome

Authors:

M. Levy¹, L. Bazak¹, L. Basel-Salmon², L. Basel-Salmon¹, I. Maya³; ¹The Raphael Recanati Genetic Inst., Rabin Med. Ctr., Petah Tikva, Israel, ²Recanati Genetic Inst., petah tikva, Israel

Abstract Body:

Introduction. *COL4A4* pathogenic variants are causative of collagen IV nephropathy or Alport syndrome, in both autosomal dominant and recessive inheritance patterns. To date, no founder pathogenic variant has been reported in this gene in the Bukharian Jewish population. **Methods.** We reviewed medical records of all individuals with suspected Alport syndrome, referred to genetic counseling in our Nephrology clinic between the years 2012-2022. The molecular basis of Alport syndrome using various molecular methods including linkage analysis, Sanger sequencing and next-generation sequencing was determined. Clinical and molecular characteristics of all families of Bukharian Jewish origin were recorded. **Results.** Molecular diagnosis was confirmed in 20/38 (52.6%) patients and 12/22 (54.5%) families of Bukharian Jewish origin. 20 patients were molecularly diagnosed, with at least one of the following three disease-causing *COL4A4* variants: c.871-6T>C, c.338G<A (p.Gly113Asp) or c.3022G>A (p.Gly1008Arg) was detected. two were obligate carriers. 16 patients were found to be heterozygous, and 6 patients were homozygous or compound heterozygous. Each one of the three variants was detected in more than one unrelated family. At the time of referral to our clinic, all patients had hematuria. Proteinuria has been seen at the earliest age in homozygotes for c.3022G>A variant. End-stage renal disease was diagnosed in one patient at the age of 38 years who was found to be compound heterozygous. Hearing deterioration was detected in 3/8 patients heterozygote for c. 338 G<A variant and appeared as early as the age of 40 years. In our local database, the frequency of the three variants was 1:22 to 1:111. All three variants were classified by all the laboratories performing the analysis and by variant classification tools as variants of uncertain significance (VUS) according to the ACMG guidelines. **Conclusion.** Our results show that currently three VUSs c.3022G>A, c.871-6T>C, and c.338G>A in *COL4A4* gene are common among Jews of Bukharian ancestry and cause Alport syndrome in both dominant and recessive inheritance patterns

Mendelian Phenotypes Posters - Wednesday

PB1995. TMEM161B mediates radial glial scaffolding in neocortical development

Authors:

L. Wang¹, J. Gleeson²; ¹UCSD, San Diego, CA, ²Univ California, La Jolla, CA

Abstract Body:

TMEM161B encodes an evolutionarily conserved widely expressed novel 8-pass transmembrane protein of unknown function in human. Here we identify *TMEM161B* homozygous hypomorphic missense variants in our recessive polymicrogyria cohort. Patients carrying *TMEM161B* mutations exhibit striking neocortical polymicrogyria and intellectual disability. *Tmem161b* knockout mice fail to develop midline hemispheric cleavage, whereas knock-in of patient mutations and patient-derived brain organoids show defects in apical cell polarity and radial glial scaffolding. We found that TMEM161B mediates actin filopodia, functioning upstream of the Rho-GTPase CDC42. Our data reveals the function of TMEM161B in apical polarity regulating radial glial scaffold formation during neocortical development.

Mendelian Phenotypes Posters - Thursday

PB1996*. Toward precision medicine in Neurodevelopmental disability: identifying phenotypic differences in individuals with global developmental delay.

Authors:

T. Cuppens¹, J. Shatto², A-H. Ng², A. Kumar³, M. Leclercq⁴, B. Finlay⁵, J. Zwicker⁵, I. Dunham⁶, A. Droit¹, F. Bolduc⁷; ¹Ctr. de recherche du CHU de Québec - Université Laval, Québec, QC, Canada, ²Univ. of Alberta, Dept. of Pediatrics Neurology, Edmonton, AB, Canada, ³EMBL-EBI, Cambridge, United Kingdom, ⁴Ctr. de recherche du CHU de Québec - Université Laval, Québec, QC, Canada, ⁵Univ. of Calgary, Sch. of Publ. Policy, Calgary, AB, Canada, ⁶European Bioinformatics Inst., Open target and European Molecular Biology Lab., Cambridge, United Kingdom, ⁷Univ Alberta, Edmonton, AB, Canada

Abstract Body:

Introduction: Global developmental delay (GDD) is a neurodevelopmental disorder (NDD) subtype that is often diagnosed at an early age and has a prevalence of up to 3%. Many clinical presentations have been described so far, but mostly in case reports and not on a large scale. The influence of age and sex on the clinical presentation of individuals with GDD has therefore never been described. This limits our ability to take advantage of precision medicine.

Objective: We were interested in determining the most common symptoms, clinical signs, and genes in individuals with GDD, both in the overall cohort and sorted by age and sex.

Methods: Using the largest cohort of individuals with GDD, Deciphering Developmental Disorders (DDD), characterized by a systematic approach, we extracted phenotypic information from 6588 individuals with GDD. We analyzed variability in phenotypic presentation using statistical analysis to identify significant variation based on age and sex.

Results: We observed that some GDD-related phenotypes were more prevalent than others, such as autistic behavior, lack of speech, seizures, strabismus, and constipation. Subsequently, we found significant differences in expressed neurodevelopmental, dysmorphic, neurological, and general phenotypes between different age groups and with sex. For example, at preschool age the prevalence of seizures is lower than in adulthood, increasing from 9.11% to 24.91%. We observed regarding sex, males with GDD had a significantly higher prevalence of autistic traits and autism diagnosis. Microcephaly was more present in females with GDD while macrocephaly was more frequent in males.

Implications: Our study highlights the importance of dissecting phenotypes from large cohort databases to better understand the relationships between phenotype, age, and sex of individuals with GDD in order to develop specific interventions. Indeed, although clinically well-defined, GDD represents a broad spectrum of possible phenotypic manifestations. This study allows us to identify the most common phenotypes, and specific phenotypes according to the patient's age and sex. We hope that our results will prove useful in guiding both clinical practice and the design of future clinical trials for individuals with GDD.

Mendelian Phenotypes Posters - Wednesday

PB1997. Transcriptional response of *Fundulus heteroclitus* embryos exposed to carbaryl

Authors:

O. Torano¹, W. Huang¹, D. Bencic¹, R. Flick¹, B. Clark², A. Biales¹; ¹US Environmental Protection Agency, Durham, NC, ²US Environmental Protection Agency, Narragansett, RI

Abstract Body:

Carbaryl is an acetylcholinesterase inhibitor widely used for pest management in agriculture and landscaping. The breadth and volume of carbaryl use results in frequent contamination of aquatic environments and subsequent aquatic organism exposures. In-vitro studies have demonstrated carbaryl to be acutely toxic with developmental effects on exposed fishes at different life stages. To better understand the molecular responses underlying known adverse outcomes, mRNA-seq was employed to assess the transcriptional response of embryonic *Fundulus heteroclitus* exposed to carbaryl at three different concentrations (100, 1000, and 10,000 µg/L). Embryos (n=20) were exposed to a single dose of carbaryl and allowed to grow for seven days after which they were transferred to clean water and grown an additional three days before samples were collected. The highest-dose exposure was prematurely terminated due to embryo death before the conclusion of the experiment and these embryos were not processed for mRNA analysis. RNA-seq analysis using STAR and RSEM showed that low and medium carbaryl exposures resulted in few differentially expressed genes or isoforms compared to untreated control at a 5% FDR cutoff. However, the Ingenuity Pathway Analysis of top-ranked genes showed significant enrichment of cancer, organismal injury and abnormality, and endocrine system disorder-related genes, and results were consistent between the low and medium doses. In summary, the results suggest that carbaryl may have little or no lasting effects on *F. heteroclitus* at the low and medium doses, though the immediate response is not clear and high-dose exposure is lethal. The lack of response in expression to the low- or medium- dose exposure, but acute toxicity at the high-dose may result from narrow timing of gene expression response and a steep dose-response curve to increasing toxicant concentration.

Mendelian Phenotypes Posters - Thursday

PB1998. Translated mutant *DSPP* mRNA expression level impacts the severity of dentin defects.

Authors:

J-W. Kim¹, Y. Kim¹, Y. Lee¹, H. Zhang², F. Seymen³, M. Koruyucu⁴, S. Bayrak⁵, N. Tuloglu⁵, J. Simmer², J. Hu²; ¹Seoul Natl. Univ. Sch. of Dentistry, Seoul, Korea, Republic of, ²Univ. of Michigan, Ann Arbor, MI, ³Altinbas Univ., Istanbul, Turkey, ⁴Istanbul Univ., Istanbul, Turkey, ⁵Private Practice, Eskisehir, Turkey

Abstract Body:

Hereditary dentin defects are conventionally classified into three types of dentinogenesis imperfecta (DGI) and two types of dentin dysplasia (DD). Mutations in the dentin sialophosphoprotein (*DSPP*) gene have been identified to cause DGI type II, III and DD type II; therefore, these are not three different conditions but allelic disorders. In this study, we recruited three families with varying clinical phenotypes from DGI-III to DD-II and performed mutational analysis by candidate gene analysis or whole exome sequencing. Three novel mutations including a silent mutation (NM_014208.3: c.52-2del, c.135+1G>C and c.135G>A; p.(Gln45=)) were identified, all of which affected pre-mRNA splicing. Comparison of the splicing assay results revealed that the expression level of *DSPP* exon 3 deletion transcript correlated with the severity of the dentin defects. This study did not only expand the mutational spectrum of *DSPP* gene but also advanced our understanding of the molecular pathogenesis impacting the severity of hereditary dentin defects.

Mendelian Phenotypes Posters - Wednesday

PB1999. Truncating variants in *RFC1* cause cerebellar ataxia, neuropathy vestibular areflexia syndrome (CANVAS).

Authors:

S. Das¹, R. Ronco², C. Perini³, R. Curro², N. Dominik², A. Gary¹, M. Delfeld¹, P. Kandikatla¹, J. Chen⁴, C. Gomez¹, J. D. Schmahmann⁴, D. Gosal⁵, H. Houlden², A. Cortese²; ¹Univ. of Chicago, Chicago, IL, ²UCL Queen Square Inst. of Neurology, London, United Kingdom, ³Inst. of Molecular Genetics IGM-CNR "Luigi Luca Cavalli-Sforza", Pavia, Italy, ⁴Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, ⁵Manchester Ctr. for Clinical NeuroSci.s, Salford Royal Hosp., Manchester, United Kingdom

Abstract Body:

Cerebellar Ataxia, Neuropathy and Vestibular Areflexia Syndrome (CANVAS) is an autosomal recessive neurodegenerative disease characterized by adult onset and slowly progressive sensory neuropathy, cerebellar dysfunction, and vestibular impairment. The genetic basis of the disorder has recently been identified to be due to a common biallelic (AAGGG)_n repeat expansion in the second intron of the Replication Factor Complex subunit 1 (*RFC1*). A large percentage of patients with a diagnosis of CANVAS have been shown to have biallelic (AAGGG)_n repeat expansions in the *RFC1* gene. However, a small number of cases with typical CANVAS do not carry the common biallelic repeat expansion. We set out to determine the genetic basis of CANVAS patients who do not harbor the biallelic repeat expansion in *RFC1*. Fifteen individuals diagnosed with CANVAS and carrying only one heterozygous (AAGGG)_n expansion in *RFC1* were selected to test for the presence of a second variant in *RFC1* by either whole genome sequencing or exome sequencing. A truncating variant in the *RFC1* gene was identified in seven patients from five unrelated families with clinically defined CANVAS carrying a heterozygous (AAGGG)_n expansion, which included: c.1267C>T (p.Arg423Ter), c.1739_1740del (p.Lys580SerfsTer9), c.2191del (p.Gly731GlufsTer6) and c.2876del (p.Pro959GlnfsTer24). The truncating variant was identified to be present in *trans* with the (AAGGG)_n expansion in all cases. Patient fibroblasts containing the c.1267C>T (p.Arg423Ter) or c.2191del (p.Gly731GlufsTer6) variants demonstrated nonsense-mediated mRNA decay and reduced RFC1 transcript and protein. This is the first description of truncating variants in the *RFC1* gene in patients with CANVAS and our result expands the genotype spectrum of *RFC1* disease. Full *RFC1* sequencing is recommended in cases affected by typical CANVAS and carrying monoallelic (AAGGG)_n expansions. Our results also shed further light on the pathogenesis of *RFC1* CANVAS as it supports the existence of a loss of function mechanism underlying this complex neurodegenerative condition.

Mendelian Phenotypes Posters - Thursday

PB2000. Tumor burden among members of the DiscovEHR cohort harboring deleterious constitutive variants in *NF1*

Authors:

A. Pemov¹, J. Kim², J. S. Haley³, H. Rao⁴, DiscovEHR Collaboration, D. J. Carey⁵, D. R. Stewart⁶; ¹NIH, NCI, Rockville, MD, ²Natl. Cancer Inst., Rockville, MD, ³Geisinger, Danville, PA, ⁴Geisinger, Danville, PA, ⁵Geisinger Clinic, Danville, PA, ⁶Natl. Cancer Inst., ROCKVILLE, MD

Abstract Body:

Background. Large-scale exome sequencing linked to electronic health records (EHR) and broad ascertainment of participants' clinical phenotypes provides a unique opportunity to investigate genotype-phenotype correlations, improve the estimates of prevalence and penetrance of Mendelian disorders, and ultimately, advance disease prognosis, risk stratification and care for patients. **Methods.** Utilizing a genome-first approach, we examined exomes of 175,449 participants from the DiscovEHR cohort and identified deleterious variants in *NF1*, the causative gene in neurofibromatosis type 1 (NF1). We identified carriers of pathogenic/likely pathogenic (P/LP) variants classified by either ClinVar or InterVar. Additionally, we identified and included all remaining loss-of-function (LoF) variants and large intragenic *NF1* deletions (CNVs) in the analysis. We then examined EHR and tumor registry records for the participants carrying these variants. **Results.** Out of 175,449 participants, we identified 124 who carried P/LP ClinVar/InterVar variants, and 17 who carried large deletions in *NF1*. Thus, the prevalence of *NF1* deleterious variant carriers in this cohort was 1:1,243. Of these 141 carriers, 39 belonged to 19 multiplex pedigrees, with ³2 family members per pedigree sharing a deleterious *NF1* variant. There were 80 participants in the entire cohort, diagnosed with neurofibromatosis type 1 (1:2,193); 58 of them (72.5%) carried a deleterious variant in *NF1*, including 44 P/LP ClinVar, 9 P/LP InterVar and 5 intragenic CNVs. The odds ratio (OR) for association of the NF1 diagnosis with the carrier status of a deleterious *NF1* variant was 5,568 (p-value<10⁻⁵ by Fisher's exact test). As anticipated, we identified elevated ORs for the following malignancies commonly observed in NF1 patients: "peripheral nerves and autonomic nervous system" (OR=381, p-value<10⁻⁵); "spinal cord, cranial nerves and other parts of central nervous system" (OR=46, p-value=10⁻³); "adrenal gland" (OR=33, p-value=10⁻⁴); and "neoplasm of brain" (OR=11, p-value=7x10⁻⁴). For the analysis of blood neoplasms, for which association with NF1 is not well-established, we combined *NF1*-carriers (13/141) and compared them to non-carriers (7,077/175,308) diagnosed with this type of tumors (OR=2.41, p-value=0.008). **Conclusions.** By utilizing a genome-first approach in the DiscovEHR cohort, we confirmed existing tumor associations with NF1 and identified potential novel risks for neoplasms in NF1 patients. Our findings will aid in the disease prognosis, risk stratification and care for the patients.

Mendelian Phenotypes Posters - Wednesday

PB2001. Two novel compound heterozygous variants identified in a new case of classic Donohue syndrome and literature review

Authors:

H. Wang, C. Grossheim, S. Ramanathan, R. Clark; Loma Linda Univ. Sch. of Med., Loma Linda, CA

Abstract Body:

INTRODUCTION: Donohue syndrome (DS, Leprechaunism) is an extremely rare, autosomal recessive disorder caused by a functional defect in insulin receptor (INSR). It is characterized by severe insulin resistance and distinctive facial features. Limited cases have been reported. Here we report a new case with two novel compound heterozygous variants in INSR gene. **CASE DESCRIPTION:** A 48-day old, 36w5d gestation male born vaginally with SGA, coarse features, large nipples, umbilical hernia/mass, large scrotal sac, abdominal distention, and hypoglycemia, transferred from outside hospital. Mother, a G3P2012 ethnically Asian woman, with pregnancy was complicated by gestational diabetes and IUGR. The birth weight was 1450 g (<3rd%ile), length 40.5 cm (<3rd %ile), HC 32 cm (21st %ile). Family history was significant for a previous pregnancy termination for severe IUGR with concern for skeletal dysplasia at 25 weeks' gestation. Consanguinity was denied. Physical exam showed emaciated dysmorphic infant, coarse facies, bulging eyes, thick eyelids, proptosis, infraorbital creases, low set ears with thick and overfolded helices superiorly, broad nose with anteverted nares, open mouth with thick patulous lips, gingival hypertrophy, mild micrognathia, prominent nipples with small white cyst on each nipple, distended abdomen, hypertrichosis and decreased subcutaneous tissue. Echo showed mild dilatation and mild hypertrophy of the right ventricle. Abdominal ultrasound showed early medullary nephrocalcinosis. Chromosome microarray was normal. The clinical picture was consistent with classic Donohue syndrome. *INSR* gene analysis detected two novel variants: c.3792_3793dup (p. Val1265Glufs*16) and c.344_348delinsCCTTG (p. Phe115_Phe116delinsSerLeu) that had not been previously reported. Parental testing revealed biparental inheritance of the variants. The laboratory updated the two variants to likely pathogenic. Therapy with IGF was discussed and patient was transferred to another hospital. **DISCUSSION:** DS has an incidence of 1 in 4 million live births but this may be underreported. Currently 87 variants of the insulin receptor gene in DS have been reported in ClinVar, of which only 19 variants have been classified as pathogenic. This infant has classic features of DS, with two novel compound heterozygous likely pathogenic variants. There is no effective treatment, postprandial hyperglycemia does not respond to insulin or metformin in DS. When hyperglycemia persists, high dose insulin may be needed. Cardiomyopathy can be treated with beta blockers. Recombinant human IGF-1 is being studied; its effectiveness is unclear.

Mendelian Phenotypes Posters - Wednesday

PB2002. Understanding the role of *TTC5* in the developing mouse brain: A novel null *TTC5* mouse model.

Authors:

R. Rushforth^{1,2}, R. Stottmann^{1,2}; ¹Nationwide Children's Hosp., Columbus, OH, ²The Ohio State Univ., Columbus, OH

Abstract Body:

Tubulin autoregulation is a long-known, but poorly understood phenomenon. This is the process regulating tubulin mRNA degradation in direct relation to the amount of tubulin protein in the cell. A recent study proposed that the tetratricopeptide 5 (*TTC5*) gene has a role in mediating tubulin autoregulation with the *TTC5* protein binding the amino terminus of tubulin peptides immediately upon exit from the ribosome. Additionally, eight patients with variants in the *TTC5* gene have been previously identified with a spectrum of shared clinical neurodevelopmental phenotypes including moderate to severe intellectual disability, simplified gyral patterns, and mild ventriculomegaly. One additional patient with bi-allelic variants in *TTC5* has been reported to have a brain malformation resembling a tubulinopathy. Literature describing *TTC5* is limited, and no studies have been done to look at the role of *TTC5* in development. There is no known *in vivo* experimental model. Here we report on a novel *TTC5* null mouse generated in our lab. This mouse contains a ~ 800 bp deletion in the *TTC5* gene, resulting in a total null allele. Homozygous null mutants experience 100% perinatal lethality ~ 24 hours after birth. Structural examination of null embryonic day 18.5 (E18.5) brains show multiple phenotypes including periventricular heterotopia and hippocampal malformations. We therefore hypothesize that total loss of *TTC5* leads to defects in neuronal migration in the developing mouse brain. Further studies are being conducted to understand the molecular mechanism(s) underlying the neuronal migration pattern in *TTC5* null mice. As *TTC5* has been reported to have a role in tubulin autoregulation, and at least one patient presents with a bi-allelic *TTC5* mutation representing a tubulinopathy, we plan to examine the relationship between tubulin and our null mouse model. Using a set of novel epitope tagged alleles of tubulin genes in our lab, we can query multiple loci to understand how loss of *TTC5* specifically affects protein levels of TUBA1A, TUBA1B, TUBA1C, TUBB2A, and TUBB2B. We plan to investigate the relationship between these tubulins and *TTC5* via Western blot analysis as this will provide insight into the proposed cellular relationship between tubulin and *TTC5* in an *in vivo* model.

Mendelian Phenotypes Posters - Thursday

PB2003. Using artificial intelligence driven NGS re-analyses followed by full genome profiling (HiFi-GS) for unsolved pediatric white matter disorders

Authors:

I. Thiffault¹, E. Farrow¹, A. S. A. Cohen¹, T. Zion¹, E. C. Boillat¹, J. J. Johnston¹, S. Younger¹, J. Meas¹, W. Cheung², S. Fournier³, S. Perrier³, T. Truong⁴, G. Evrony⁴, G. Bernard³, T. Pastinen¹; ¹Children's Mercy, Kansas City, MO, ²Children's Mercy, Kansas City, KS, ³Res. Inst. of the McGill Univ. Hlth.Ctr., Montreal, QC, Canada, ⁴New York Univ. Grossman Sch. of Med., New York, NY

Abstract Body:

Leukoencephalopathies are a group of heterogeneous genetic diseases leading to progressive disabilities and often associated with poor prognosis. Collectively, their incidence is ~ 1:7500 births. Although some subtypes of white matter disease (WM) have a high (>70%) diagnostic rate by short-read exome (srES), other forms with more complex presentation (neuro+) remain mostly unsolved even by short-read genome sequencing (srGS). Beyond increased diagnostic yield by AI-driven reanalyses, we hypothesize that variations in “srES/srGS blindspots” (repeat expansions, structural variants (SV), intronic/regulatory variants) account for missed molecular diagnoses. Such variations can be detected by Long-read GS using PacBio Sequel IIe (HiFi-GS). Furthermore, HiFi-GS can be utilized for interrogation of non-coding/regulatory variation via methylation detection as well as RNA (HiFi-IsoSeq) and can overcome limitations of reference genomes by *de novo* assembly. **Methods:** In the context of a collaborative genomic medicine program “Genomic Answers for Kids” with >3800 families with suspected genetic disorders; we have developed one of the largest HiFi-GS data resources by integrating an enhanced HiFi-GS workflow for unsolved pediatric cases (> 1000 samples). We recently demonstrated incorporating SVs accounted for 13% of diagnostic yield. To determine the utility of HiFi-GS in WM diseases, we established a cohort of ~270 patients. All patients entering the program have had srES/srGS performed. Prior deploying HiFi-GS, existing srES/srGS data are systematically reanalyzed by AI-driven methods. **Preliminary results:** As previously reported, subtypes of WM (i.e., hypomyelinating, mitochondrial or metabolic) have a high (70%) diagnostic rate versus neuro+ (22%). In neuro+, we have blended phenotypes among patients with dual diagnoses (8%) and a higher rate of gene discovery (18%). For non-trio cases, if VUS(s) detected by srES/srGS are confirmed *de novo* or *in trans* configuration, the genotype would be diagnostic (18%). As examples of HiFi-GS utility in WM, we solved three unrelated patients by identifying pathogenic variants; *CEP85L* (p.M1?), *TUBB2A* (p.R306H) and a deletion in *AARS2*. These variants were undetected by srES/srGS due to poor coverage or misalignment of homologous sequences. We also identified a private hypermethylated “GCC” expansion in a non-OMIM gene. **Discussion:** We are pursuing full genome profiling of unsolved WM cases by leveraging HiFi-GS, *de novo* assembly, methylation and IsoSeq. Moreover, we are validating the findings in patient derived cells (iPSCs) to systematically study impact in RNA or myelination (oligodendrocytes).

Mendelian Phenotypes Posters - Thursday

PB2004. Using genetic relatedness within biobanks to understand genetic architecture of Retinitis Pigmentosa

Authors:

J. Baker¹, H-H. Chen^{2,3}, M. A. Brantley, Jr.^{4,2}, J. E. Below^{2,3}, D. C. Samuels^{5,6}; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Vanderbilt Genetics Inst., Nashville, TN, ⁴Vanderbilt Eye Inst., Nashville, TN, ⁵Vanderbilt Univ. Sch. of Med., Nashville, TN, ⁶Dept. of Molecular Physiology & Biophysics, Nashville, TN

Abstract Body:

Background: Retinitis Pigmentosa (RP) is a group of inherited eye diseases that cause progressive degeneration of the rod and cone cells in the retina ultimately leading to vision loss. The genetic architecture of RP has been well characterized with pathogenic mutations identified in more than 100 genes with approximately 10% of these variants falling within the *RHO* gene. Despite the well described genetics of RP and the availability of comprehensive genetic testing, only about 60% of RP cases can be definitively explained by a causal variant. Large biobanks can provide greater insight into this lack of causality, allowing researchers to leverage genetic relatedness and electronic health records (EHR) to identify likely carriers of causal rare variants without the need for large-scale sequencing efforts through detection of genomic segments shared identically-by-descent (IBD) due to recent common ancestry. **Methods and Results:** Pairwise IBD segments longer than 3cM were identified using hap-IBD for 69,819 individuals of European ancestry in Vanderbilt University's biobank, BioVU. We constructed networks of individuals that shared an IBD segment overlapping the *RHO* locus. 16,997 networks were identified using this approach. We identified a cluster enriched for diagnosed RP ($p=1.3e-8$) where 4 of 10 individuals were diagnosed in the EHR as having RP. The individuals in this cluster had a mean age of 51.8 years. To identify potentially causal variants, these 10 individuals were whole exome sequenced and the sequencing results were annotated using wANNOVAR. All 10 individuals were confirmed as carrying a reported pathogenic, autosomal dominant, Pro23His variant, chr3:129528801-C-A within the *RHO* locus. **Conclusions:** Only 4 of the 10 individuals are diagnosed with RP in the EHR suggesting that the variant may exhibit incomplete or age-related penetrance rather than the reported autosomal dominance. Another interpretation of these results is that these individuals may not yet have been diagnosed as having RP due to mild symptoms or they may have been diagnosed outside of Vanderbilt's health system. These results highlight the power of this analysis pipeline for rare variant carrier discovery and assessment of variant effects, with implications for precision medicine and prevention. Our approach affordably and efficiently identifies carriers of Mendelian disease variants in health systems prior to disease onset and in the absence of clinical family history data.

Mendelian Phenotypes Posters - Wednesday

PB2005. Using remote methods to assess communication in Phelan-McDermid Syndrome.

Authors:

S. Quadri Valverde¹, J. Klusek², L. Ward¹, K. Phelan³, C. Rogers⁴, N. Powers⁵, L. Boccuto⁶, S. M. Sarasua¹; ¹Clemson Univ., Clemson, SC, ²Univ. of South Carolina, Columbia, SC, ³Florida Cancer Specialists, Fort Myers, FL, ⁴Greenwood Genetic Ctr., Greenwood, SC, ⁵Prisma Hlth., Greenville, SC, ⁶Clemson Univ., GREENWOOD, SC

Abstract Body:

Phelan-McDermid Syndrome (PMS) is a rare genetic disorder caused by deletions involving the distal long arm of chromosome 22 or by pathogenic variants of the *SHANK3* gene localized to 22q13.3. Although individuals with PMS have reduced speech and language abilities, little research has been directed at profiling the communication abilities in this population. Typical communication assessments require in-person evaluations by speech language pathologists (SLPs) which often requires burdensome or prohibitive travel by patients and their families to centralized research centers. Further, clinical evaluation teams rarely include SLPs and thus omit this critical area or rely on over simplified or proxy measures. The objectives of the present study are threefold. The first is to identify the language and communication profile of school-aged children with PMS using a combination of standardized assessments and semi-formal measures such as direct observation and language sampling by a SLP. A second objective is to identify the genetic contributions to the language and communication profile in individuals with PMS. The final objective is to determine the feasibility of remote data collection for research purposes. The remote aspect of this study is a unique feature that will support increased accessibility to participate in studies for families who do not have sufficient funds to finance travel to a testing site or for vulnerable patient populations considering the COVID-19 pandemic. The methods included recruitment of participants facilitated by the Phelan-McDermid Syndrome Foundation (PMSF) with enrollment of up to 20 school-aged children with PMS. Children and their parents participated in research activities remotely using a video-conferencing platform and mail-in saliva samples. Caregivers were asked to independently complete standardized measures of adaptive behavior and social communication abilities. Direct evaluations were remotely conducted across one to two sessions based on family needs. A SLP observed a ten-minute subject-caregiver interaction to collect a language sample and identify nonverbal expressive/receptive and pragmatic communicative functions, followed by formal standardized assessments. The communication profile was correlated with existing genomic data. We will present our findings on remote data collection methods along with preliminary language assessment outcomes. The results of this pilot study demonstrating the efficacy of our methods to assess communication in PMS may be helpful for future evaluations of individuals with PMS as well as others with genetic disorders involving speech-language phenotypes.

Mendelian Phenotypes Posters - Wednesday

PB2006. Variable expressivity in a four-generation ACDMPV family with a noncoding hypermorphic SNV in *trans* to the frameshifting *FOXF1* variant

Authors:

P. Stankiewicz¹, J. A. Karolak², P. Szafranski¹, T. Gambin³, A. Matsika⁴, S. Mcmanus⁴, H. S. Scott^{5,6,7,8,9}, P. Arts⁵, T. Ha^{5,7}, C. P. Barnett^{10,11}, J. Rodgers^{12,13}, E. Yildiz Bolukbasi¹; ¹Dept. of Molecular & Human Genetics, Baylor Coll. of Med., Houston, TX, ²Chair and Dept. of Genetics and Pharmaceutical Microbiol., Poznan Univ. of Med. Sci., Poznan, Poland, ³Inst. of Computer Sci., Warsaw Univ. of Technology, Warsaw, Poland, ⁴Mater Pathology, Mater Hlth., South Brisbane, Australia, ⁵Dept. of Genetics and Molecular Pathology, Ctr. for Cancer Biology, An alliance between SA Pathology and the Univ. of South Australia, Adelaide, Australia, ⁶UniSA Clinical and Hlth.Sci., Univ. of South Australia, Adelaide, Australia, ⁷ACRF Genomics Facility, Ctr. for Cancer Biology, An alliance between SA Pathology and the Univ. of South Australia, Adelaide, Australia, ⁸Dept. of Genetics and Molecular Pathology, SA Pathology, Adelaide, Australia, ⁹Australian Genomics, Melbourne, Australia, ¹⁰Adelaide Med. Sch., Univ. of Adelaide, Adelaide, Australia, ¹¹Paediatric and Reproductive Genetics Unit, South Australian Clinical Genetics Service, Women's and Children's Hosp., North Adelaide, Australia, ¹²Genetic Hlth. Queensland, Royal Brisbane and Women's Hosp., Brisbane, Australia, ¹³Sch. of Med., The Univ. of Queensland, Brisbane, Australia

Abstract Body:

Heterozygous single nucleotide variants (SNVs) or copy-number variant (CNV) deletions involving *FOXF1* or its distant lung-specific enhancer on chromosome 16q24.1 have been identified in 80-90% of patients with Alveolar capillary dysplasia with misalignment of pulmonary veins (ACDMPV), a lethal neonatal lung developmental disorder. Recently, we showed that rare non-coding regulatory hypermorphic SNVs within the *FOXF1* enhancer on the non-deleted/mutated allele of *FOXF1* (in *trans*) may prevent ACDMPV lethality (PMID: 31686214). Here, we describe a four-generation family with a deceased ACDMPV neonate, her sibling from the electively terminated pregnancy, healthy mother with a history of pulmonary arterial hypertension (PAH), an unaffected aunt, an aunt deceased due to findings consistent with ACDMPV, and a reportedly unaffected grandmother, all with the frameshifting variant c.881_902dup (p.Gly302Profs*46) in *FOXF1*, and a deceased great-grandmother with a history of PAH. Genome sequencing analyses in the proband's unaffected mother revealed a non-coding putative regulatory SNV rs560517434-A within the lung-specific distant *FOXF1* enhancer in *trans* to the *FOXF1* frameshift variant. Functional testing of the rs560517434-A variant using an *in vitro* luciferase reporter assay showed that it increases *FOXF1* promoter activity 10-fold. Our studies further demonstrate that non-coding SNVs in the *FOXF1* enhancer region can rescue the lethal ACDMPV phenotype and support the compound inheritance gene dosage (CIGD) model.

Mendelian Phenotypes Posters - Thursday

PB2007. Variants of *LRP2*, encoding a multifunctional cell surface endocytic receptor, are associated with hearing loss and retinal dystrophy

Authors:

R. Faridi¹, R. Yousaf¹, S. Gu¹, S. Inagaki¹, A. Turriff², K. Pelstring³, A. Griffith^{4,5}, S. Adadey⁶, E. Aboagye⁶, G. Awandare⁷, R. Morell⁸, E. Tsilou², C. Brewer⁴, A. Noyes⁹, L. Sulmonte⁹, A. Wonkam¹⁰, H. Azaiez¹¹, I. Schrauwen¹², S. Leal¹³, S. Riazuddin¹⁴, M. Hoa¹⁵, W. Zein², K. Dios³, T. Friedman¹⁶; ¹Natl. Inst. on Deafness and other Communication Disorders, Bethesda, MD, ²Ophthalmic Genetics and Visual Function Branch, Natl. Eye Inst., Bethesda, MD, ³Div. of Med. Genetics, Dayton Children's Hosp., Dayton, OH, ⁴Otolaryngology Branch, Natl. Inst. on Deafness and Other Communication Disorders, Bethesda, MD, ⁵Dept. of Otolaryngology, Coll. of Med., Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, ⁶West African Ctr. for Cell Biology of Infectious Pathogens (WACCBIP), Legon, Ghana, ⁷Univ Cape Town, Legon, Ghana, ⁸NIDCD, Rockville, MD, ⁹GeneDx, Inc., Gaithersburg, MD, ¹⁰McKusick-Nathans Inst. and Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ¹¹Univ. of Iowa, Iowa City, IA, ¹²Columbia Univ., NYC, NY, ¹³Columbia Univ., New York, NY, ¹⁴Allama Iqbal Med. Res. Ctr., Jinnah Hosp. Complex, Lahore, Pakistan, ¹⁵Auditory Dev. and Restoration Program, Natl. Inst. on Deafness and Other Communication Disorders, Bethesda, MD, ¹⁶NIDCD/NIH, Bethesda, MD

Abstract Body:

Deafness and retinal dystrophy are clinically variable and genetically heterogeneous phenotypes. Usher syndrome (USH) is a common cause of deaf-blindness and can be associated with biallelic variants of *MYO7A*, *USH1C*, *CDH23*, *PCDH15*, *SANS*, *USH2A*, *ADGRV1*, *WHRN*, *CLRN1* or *ARSG*. Two small unrelated families segregating deaf-blindness were screened using exome sequencing (ES). However, no pathogenic variants of these genes were detected in affected individuals from these families. In one proband we identified compound heterozygous variants of low-density lipoprotein receptor-related protein 2, *LRP2*, p.(Thr50Ser) and p.(Asn1669Asp). In two deaf-blind sisters of the second family, ES revealed compound heterozygous *LRP2* variants, p.(Tyr3933Cys) and a c.7715+3A>T, which we experimentally confirmed as a splice altering variant. In mouse cochlea, *Lrp2* is expressed most abundantly in stria vascularis marginal cells as shown by smFISH, single-cell RNAseq and single-nucleus RNAseq analyses. These data suggest that a *LRP2* deficiency compromises the endocochlear potential, which facilitates hearing. Damaging *LRP2* variants have been associated with Donnai-Barrow syndrome and other multisystem, highly pleiotropic phenotypes different from the two deaf-blindness phenotypes reported here. Biallelic variants of *LRP2* should be considered in cases with a deaf-blindness phenotype, when a panel of common USH genes does not provide a molecular genetic diagnosis.

Mendelian Phenotypes Posters - Wednesday

PB2008. Vici syndrome in Israel: Clinical and molecular insights

Authors:

B. Poded-Shakked^{1,2,3}, O. Chorin^{1,2}, Y. Hirsch⁴, R. Rock^{5,2}, L. Sheelo Salzer^{6,2}, Y. Goldberg^{6,2}, H. Mandel^{7,8}, T. Hershkovitz^{7,9}, N. Fleischer¹⁰, L. Greenbaum^{5,2,11}, U. Katz^{12,2}, O. Barei^{13,14}, N. Hamed^{15,2}, B. Ben-Zeev^{15,2}, S. Greenberger^{16,3,2}, R. Somech^{17,2}, A. Raas-Rothschild^{18,2}; ¹The Inst. for Rare Diseases, Edmond and Lily Safra Children's Hosp., Sheba Med. Ctr., Ramat Gan, Israel, ²Sackler Faculty of Med., Tel-Aviv Univ., Tel Aviv, Israel, ³The Talpiot Med. Leadership Program, Sheba Med. Ctr., Ramat Gan, Israel, ⁴Dor Yeshorim, Brooklyn, NY, ⁵The Danek Gertner Inst. of Human Genetics, Sheba Med. Ctr., Ramat Gan, Israel, ⁶Raphael Recanati Genetic Inst., Rabin Med. Ctr. - Beilinson Hosp., Petach Tikva, Israel, ⁷Ruth & Bruce Rappaport Faculty of Med., Technion-Israel Inst. of Technology, Haifa, Israel, ⁸Metabolic Disease Unit, Ruth Rappaport Children's Hosp., Rambam Hlth.Care Campus, Haifa, Israel, ⁹The Genetics Inst., Rambam Hlth.Care Campus, Haifa, Israel, ¹⁰FDNA Inc., New York, NY, ¹¹The Joseph Sagol NeuroSci. Ctr., Sheba Med. Ctr., Ramat Gan, Israel, ¹²Pediatric Heart Inst., Edmond and Lily Safra Children's Hosp., Sheba Med. Ctr., Ramat Gan, Israel, ¹³The Genomic Unit, Sheba Cancer Res. Ctr., Sheba Med. Ctr., Ramat Gan, Israel, ¹⁴The Wohl Inst. for Translational Med. and Cancer Res. Ctr., Sheba Med. Ctr., Ramat Gan, Israel, ¹⁵Pediatric Neurology Unit, Edmond and Lily Safra Children's Hosp., Sheba Med. Ctr., Ramat Gan, Israel, ¹⁶Dept. of Dermatology, Sheba Med. Ctr., Ramat Gan, Israel, ¹⁷Pediatric Dept. A and the Immunology Service, Jeffrey Modell Fndn. Ctr.; Edmond and Lily Safra Children's Hosp., Sheba Med. Ctr., Ramat Gan, Israel, ¹⁸The Inst. of Rare Diseases, Edmond and Lily Safra Children's Hosp., Sheba Med. Ctr., Ramat Gan, Israel

Abstract Body:

Introduction: Vici Syndrome (OMIM #242840) is a severe, rare neurodevelopmental disorder with multi-systemic manifestations presenting in infancy, and characterized by agenesis of the corpus callosum, hair and skin hypopigmentation and bilateral cataracts. Additional features include global developmental delay, seizures, progressive microcephaly, failure to thrive, cardiomyopathy and varying degrees of immunodeficiency. First described by Dionisi-Vici and colleagues in 1988, Vici syndrome has been described in less than a hundred individuals worldwide. It is caused by biallelic variants in *EPG5*, resulting in impaired autophagy. *Objective and Methods:* We aimed to characterize the clinical and molecular findings in a series of individuals harboring biallelic *EPG5* variants, recruited from four medical centers in Israel. We further sought to assess the prevalence of the c.1007A>G variant, suggested to be a founder mutation among the Ashkenazi-Jewish population, as well as to utilize a machine learning-based tool to assess facial features typical of Vici syndrome. *Results:* Five previously unreported cases of Vici Syndrome, one of which was diagnosed prenatally with subsequent termination of pregnancy, and a sixth previously-published case, were recruited. Among these four unrelated families, a total of four variants were detected in *EPG5*: two novel variants, c.2554-5A>G and c.1461delC, each considered likely pathogenic; and 2 previously reported variants, c.3447G>A and c.1007A>G, the latter suggested to be an Ashkenazi-Jewish founder mutation. In order to evaluate the carrier rate of the c.1007A>G variant, Next Generation Sequencing (NGS)-based testing was carried out amongst 140,491 individuals from diverse Jewish populations who enrolled in the Dor Yesharim pre-marriage screening program between 2016 and 2021. Of these, 508 carriers were identified, yielding an overall carrier frequency of 0.36% (1 in 276), and a carrier rate of 0.45% (1 in 224) in Ashkenazi Jewish individuals. Finally, based on two-dimensional facial photographs of individuals with Vici syndrome (n=19), a composite facial mask was created using the DeepGestalt algorithm, illustrating facial features typical to the disorder. *Conclusions:* We report a case series of five individuals from four unrelated families, and an additional aborted fetus, affected with Vici syndrome. Our findings contribute to the current knowledge

ASHG 2022 Annual Meeting Poster Abstracts

regarding the molecular basis and phenotypic features of Vici syndrome. Additionally, the deep learning-based facial gestalt adds to the clinician's diagnostic toolbox and may aid in facilitating identification of additional affected individuals.

Mendelian Phenotypes Posters - Thursday

PB2009. Whole exome sequencing identifies novel variants in familial myoclonic epilepsy in Mali.

Authors:

S. Bamba¹, M. Dembele¹, L. Sidibé², W. Ji³, S. Diarra^{4,5}, A. Yalcouyé¹, S. Diallo⁶, A. Maiga¹, O. Traoré¹, C. Guinto¹, G. Landouré¹, S. Lakhani³, The H3Africa Consortium; ¹Univ. of Sci., Techniques and Technologies of Bamako, Bamako, Mali, ²Service de Pédiatrie, CHU Gabriel Touré, Bamako, Mali, ³Yale Univ., New Haven, CT, ⁴NIH, ROCKVILLE, MD, ⁵Pediatric Genomics Discovery Program (PGDP), Yale school of Medecine, New Haven, CT, New Haven, CT, ⁶Service de Neurologie, CHU Gabriel Touré, Bamako, Bamako, Mali

Abstract Body:

Introduction Progressive myoclonus epilepsies (PMEs) are a group of rare hereditary neurological disorders characterized by the recurrent myoclonus seizures, and a progressive neurological and cognitive decline. Several clinical entities and genes are reported worldwide, mostly in population with Caucasian ancestry. However, genetically diagnosed cases are scarce in sub-Saharan African population. Only one case of Lafora disease caused by a novel variant in the *NHLRC1* gene was previously identified in Mali. In this study, we used whole exome sequencing (WES) to identify genetic causes of PME in the Malian population. **Objective:** To clinically characterize patients with PME and identify the underlying genetic defects. **Methods:** Institutional ethical approval was obtained. Proband and their families have been enrolled after an informed written consent. Patients undergone a carefully clinical examination by neurologists, neurogeneticist and pediatrician. Blood chemistries and electroencephalography were performed. DNA was extracted from peripheral blood for WES. Several bioinformatics tools were used to assess deleteriousness of the variants and putative variants were confirmed with Sanger sequencing in all available family members. **Results:** Five patients (4 males and 1 female) from three unrelated families were enrolled. Consanguinity was reported in one family and inheritance pattern was consistent with autosomal recessive in two families and autosomal dominant in one family. Disease started in their first decade and the mean age at diagnosis was 6 years (ranging from one month to 30 years). All patients had myoclonus and tonic and clonic seizures. Furthermore, three different patients presented additional symptoms including a spasm with flexion of upper limbs and a psychological and motor delay. Phenotypes suggested a case of severe neonatal myoclonic epilepsy, dominant adult familial myoclonic epilepsy and Lafora disease, in the respective families. WES identified a novel homozygous missense variant located in the *GRIN1* gene (c.T1703C; p.Leu568Pro) and *NHLRC1* gene (c.T602C; p.Phe201Ser) in two different families. A novel heterozygous missense variant was found in *YEATS2* gene (c.A2963G; p.Lys988Arg) in the dominant family. These variants were predicted to be damaging through several *in silico* prediction tools and are absent from all public SNP databases. **Conclusion:** Our study reports novel findings in patients suffering from myoclonic epilepsy and expands their genetic epidemiology. Including patients with different genetic backgrounds may help uncovering novel clinical and/or genetic entities likely using WES approach.

Mendelian Phenotypes Posters - Wednesday

PB2010. Whole exome sequencing in intellectual disability patients identifies de novo mutations in KCNB1, PPP1R3F, SHANK2, and SYNGAP1 genes

Authors:

A. Alkhateeb, M. Almomani; Jordan Univ. of Sci. and Technology, Irbid, Jordan

Abstract Body:

Intellectual disability etiology still poses a challenge to clinicians and families. Here we aimed to dissect the genes causing intellectual disability in local families. We recruited nine trio families with unexplained intellectual disability, and utilized whole exome sequence to identify causative genes/mutations. Out of nine families, we identified the causative genes in four (44% success rate). Novel and known pathogenic mutations were identified in KCNB1, PPP1R3F, SYNGAP1, and SHANK2. All mutations were de novo. With a highly inbred population it was unexpected to find all our mutations to be de novo representing autosomal dominant inheritance as the major pattern for our sample of unexplained intellectual disability. Our data confirm previous data that de novo mutations in autosomal dominantly expressed genes represent the major cause of unexplained intellectual disability, even in a highly inbred population.

Mendelian Phenotypes Posters - Thursday

PB2011*. Whole exome sequencing reveals known and candidate genes in non-syndromic hearing impairment in Mali

Authors:

A. Yalcouyé^{1,2}, O. Traoré¹, I. Shrauwen³, C. Bope⁴, A. Maiga¹, A. Tamega¹, A. Acharya³, T. Bharadwaj³, C. De Kock², M. Jonas², O. Oluwole², C. Guinto^{1,5}, G. Landoure^{1,5}, S. Leal³, A. Wonkam^{2,6}; ¹Univ. of Sci., Techniques, and Technologies, Bamako, Mali, ²Univ. of Cape Town, Cape Town, South Africa, ³Ctr. for Statistical Genetics, Gertrude H. Sergievsky Ctr., and Dept. of Neurology, Columbia Univ. Med. Ctr., New York, NY, ⁴Dept. of Mathematics and Computer Sci., Faculty of Sci., Univ. of Kinshasa, Kinshasa, Congo, Democratic Republic of the, ⁵Service de Neurologie, CHU Point G, Bamako, Mali, ⁶McKusick-Nathans Inst. of Genetic Med. and Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Abstract Body:

Hearing impairment (HI) is a common burden worldwide, affecting up to 7/1000 living new-borns in developing countries. More than 100 causative genes are reported to date to be associated with non-syndromic (NS) HI. While pathogenic GJB2 variants are prevalent in Europeans, their contribution to HI is nonsignificant in most sub-Saharan African populations. We used Whole Exome Sequencing (WES) to identify causative variants in hearing impaired individuals (N=37) from 12 multiplex and 8 simplex families from Mali segregating NSHI. Nineteen families displayed an autosomal recessive mode of inheritance with nine being consanguineous. One family displayed an autosomal dominant mode of inheritance. Carefully clinical assessment was performed and HI was evaluated using pure tone audiometry. The onset of the disease was prelingual in 94,6% of patients. Putative causal variants identified with WES were validated using Sanger sequencing and segregation was confirmed in family members which has not undergone WES. Bioinformatic tools were used to predict conservation and deleteriousness. Additionally protein modelling was performed. WES identified novel variants in 13 of 16 families (81.2%) in known HI genes. Compound heterozygous variants were found in four families located in OTOGL [c.209-9C>G and c.5685C>A; p.(D1895E)], MYO15A [c.8183G>A; p.(R2728H) and c.9401G>A; p.(R3134Q)], MYO15A [c.6783G>A; p.(W2261X) and c.8548C>T; p.(R2850X)], OTOF [c.1697G>A; p.(R566Q) and c.748C>T; p.(R250W)] genes. In eight families homozygous missense variants were found: CDH23 [c.646C>G; p.(L216V) and c.1353G>T; p.(L451F)], PJK [c.461T>G; p.(V154G)], PDZD7 [c.1811C>A; p.(P604Q)], OTOF [(c.2154G>A; p.(W718X)], CIB2 [c.409C>T; p.(R137W)], MYO15A [c.9401G>A; p.(R3134Q)], and MYO6 [c.2839C>T; p.(R947X)]. Additionally a homozygous splice variant was found in TMC1 (c.2003+2T>C) in one family. The known variants in the remaining three families included two in MYO15A [c.6331A>T, p.(N2111Y) and c.8158G>A, p.(D2720N)] and one in MYO7A [c.3945C>A; p.(C1315X)] genes. Moreover, four candidate genes were identified including UBF1 [c.58G>A; p.(E20K)], BLM [c.2T>C; p.(M1?)], ZSWIM6 [c.309G>T; p.(E103D)], and ODF3L2 [(c.763G>A; p.(A255T) and c.556G>A; p.(V186M)] in four different families. These variants segregate with the phenotype and were found to be damaging with in silico tools and protein modelling. Functional studies are ongoing to confirm pathogenicity. This study confirms the genetic heterogeneity of the African population and suggests novel gene discovery. The data will contribute to the understanding of the pathobiology of HI, globally

Mendelian Phenotypes Posters - Wednesday

PB2012. Whole exome sequencing will miss Congenital Myotonic Dystrophy in the child on an asymptomatic mother.

Authors:

F. Appiah¹, B. L. Marr², A. Eldokla³, R. R. Lebel⁴, A. Sakonju³; ¹SUNY Upstate Med. Univ., Norton Coll. of Med., Syracuse, NY, ²Crouse Hlth., Syracuse, NY, ³SUNY Upstate Med. Univ., Syracuse, NY, ⁴SUNY Upstate Med Univ, Syracuse, NY

Abstract Body:

Myotonic dystrophy is an autosomal dominant multi-system disorder caused by CTG repeat expansion within the *DMPK1* gene (Dystrophica Myotonia Protein Kinase 1). This male was born to a 25-year-old G2P0010>1011 Puerto Rican mother and 36-year-old father, in a non-consanguineous union, at 39 weeks by cesarean route for fetal distress, after a good prenatal care. The pregnancy was complicated by abnormal ultrasound findings: bilateral club feet and renal abnormalities. Resuscitation in the delivery room was by tactile stimulation with blow by oxygen. Apgar scores were 1, 3 and 3 at 1, 5 and 10 minutes respectively. He was transferred to NICU for respiratory support due to unstable airway. Birth weight was 3.13 kg (32nd percentile), length 52 cm (87th percentile) and head circumference 32.5 cm (6th percentile). Physical exams was significant for profound hypotonia, spontaneous eye opening without any other spontaneous movements, bilateral equinovarus and undescended right testicle. Irregular heart rate was noted and EKG on day 7 showed QT prolongation. Repeat EKG at 2 months showed sinus tachycardia, right ventricular hypertrophy and slight right ventricular conduction delay. Echocardiography showed a small patent ductus arteriosus with left to right shunting. EMG showed abnormalities in nerve conduction, and diffuse discharges consistent with myotonic dystrophy. He had mechanical ventilation for 47 days with brief unsuccessful CPAP trials, and a teach collar placed at 6 weeks. He received Glycopyrrolate for excess secretions, and also Oxcarbazepine. Feeding is by G-tube. Karyotype and FISH were normal (46, XY). Whole Exome Sequencing (WES) revealed heterozygous *NEB* variants of uncertain significance, in trans, but histochemical and electronic microscopic examination of his muscle biopsy did not show nemaline rods. A molecular diagnosis of myotonic dystrophy type 1 was accomplished with over 200 CTG repeats in the *DMPK1* gene. Exact number of repeats is indeterminate. It is important to note that WES may not find triple repeats but rather give a false negative. The EMG was positive for myotonia with complex repetitive discharges. His mother's EMG was also positive for classic electrophysiologic myotonia despite having inconsistent grip myotonia and no percussion myotonia. Thus asymptomatic mothers can give birth to children with severe congenital myotonic dystrophy, and EMG testing can show abnormalities consistent with myotonic dystrophy in children as young as 6 days old.

Mendelian Phenotypes Posters - Thursday

PB2013. Whole genome sequencing identifies biallelic *SCLT1* splice-region variants in a child with a suspected ciliopathy: Delineating the *SCLT1*-related ciliopathy spectrum.

Authors:

E. Gillesse¹, P. Au², F. Bernier³, J. Parboosingh¹, Care4Rare Canada Consortium, R. Lamont³, A. Innes³; ¹Univ. of Calgary, Calgary, AB, Canada, ²Univ. of Calgary, Cumming Sch. of Med., Calgary, AB, Canada, ³Alberta Children's Hosp, Calgary, AB, Canada

Abstract Body:

Ciliopathies are a complex class of disorders that are caused by impaired primary cilia function, often impacting multiple organ systems, and show significant phenotypic and genetic heterogeneity. Sodium channel and clathrin linker 1 (*SCLT1*) is an emerging ciliopathy gene that encodes a known core component of the distal appendages of cilia, which are essential for primary cilium assembly and function. To date, there are a few individuals with pathogenic *SCLT1* variants reported with a wide range of named phenotypes, including oro-facio-digital syndrome type IX, Bardet-Biedl syndrome (BBS), Senior-Løken syndrome (SLS), and non-syndromic retinitis pigmentosa. Despite these reports, no disease association is currently listed for *SCLT1* in databases such as OMIM and the gene is excluded from most commercial gene panels and clinical exome analyses. Here we report a female proband with a phenotype suggesting a ciliopathy most consistent with BBS or SLS who initially underwent a commercial ciliopathy gene panel and clinical exome that were both non-diagnostic. She presented with early onset obesity, developmental delay, nystagmus and severe early onset vision impairment suggestive of Leber congenital amaurosis, and later developed renal failure secondary to nephronophthisis. Whole-genome sequencing (WGS) identified compound heterozygous variants *SCLT1* NM_144643.4: c.1218+3A>T and c.1218+3A>G. Further cDNA analysis showed two mis-spliced transcripts and no wild-type transcripts. The first transcript showed skipping of exon 14, resulting in a 72 bp in-frame deletion (r.1147_1218del: p.(Val383_Met406del)). The second transcript showed skipping of both exons 13 and 14, resulting in a 171bp in-frame deletion (r.1048_1218del: p.(Ala350_Met406del)). Segregation of cDNA from the mother revealed that both variants cause both alternatively spliced transcripts seen in the proband. This is the first report of a variant at the exon 14 - intron 14 boundary leading to skipping of exons 13 and 14. Overall, these findings contribute to the growing evidence that variant location and subsequent impact on *SCLT1* protein function, may influence the burden of an individual's phenotype. This report highlights the utility of WGS in the diagnosis of unsolved patients with genetically heterogeneous conditions, and will present a detailed review of the phenotypes associated with *SCLT1*-related ciliopathies.

Mendelian Phenotypes Posters - Wednesday

PB2014. Whole-Exome Sequencing of a French Canadian Cohort Reveals New Candidate Founder Mutations

Authors:

D. Ashton-Beaucage^{1,2}, G. Gagnon^{3,4}, J-C. Grenier⁵, G. Boucher⁵, D. Taliun^{3,4}, C. Stevens⁶, M. Daly^{7,8,9}, J. G. Hussin^{5,2}, V. Mooser^{3,4}, C. Bhérier^{3,4}, J. D. Rioux^{5,2}; ¹Dept. of Med. Genetics, McGill Univ. Hlth.Ctr., Montreal, QC, Canada, ²Dept. of Med., Université de Montréal, Montreal, QC, Canada, ³Dept. of Human Genetics, Faculty of Med. and Hlth.Sci., McGill Univ., Montreal, QC, Canada, ⁴McGill Genome Ctr., Canada Excellence in Res. Chair in Genomic Med., McGill Univ., Montreal, QC, Canada, ⁵Montreal Heart Inst., Montreal, QC, Canada, ⁶The Broad Inst., Cambridge, MA, ⁷Med. and Population Genetics, Harvard Med. Sch., Cambridge, MA, ⁸Analytical and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, ⁹Inst. for Molecular Med. Finland (FIMM), Univ. of Helsinki, Helsinki, Finland

Abstract Body:

Mendelian conditions pose a significant burden on public health. In founder populations, the prevalence of some Mendelian disorders is increased, however population-level prevalence data is often fragmentary or lacking. The French Canadian (FC) population of Quebec comprises over 6 million people descending from 8,500 French settlers in the 17th and 18th centuries. To date, extensive genetic research in this population has identified over 30 Mendelian diseases bearing a founder effect signature, yet a comprehensive genome-scale survey of pathogenic variants is still lacking. The availability of whole-exome sequencing (WES) offers an opportunity to fill this gap and provide much needed carrier frequency data which is of fundamental importance to clinical genetics and epidemiology. Here, we estimated the allele frequencies (AF) of ClinVar pathogenic and likely pathogenic (PLP) variants in 1928 unrelated individuals from a case-control cohort of 2,625 participants included in a WES study of Inflammatory Bowel Diseases (IBD) in Quebec. After standard WES variant- and individual-level quality control and exclusion of IBD-associated genes, a total of 1886 PLP variants were identified. We estimated the relative AF enrichment in the FC population by comparing our AFs to non-Finnish Europeans AFs from gnomAD. In doing so, we discovered 157 PLP variants with an AF > 0.001 that were significantly enriched (one-sided Fisher exact test p-value < 0.01). After manual curation, we identified 34 variants with prior evidence of higher frequency and/or specific shared haplotypes, providing further evidence of a FC specific founder effect. Our carrier frequency estimates show high correlation to available neonatal screening data and to previously published estimates. The remaining 122 enriched candidates represent potential novel FC founder effect variants. These notably include variants linked to metabolic disorders (n = 23), neurological conditions (n = 17), eye diseases (n = 8), hearing loss (n = 14), endocrine disorders (n = 7) and cancer predisposition (n = 3). We will present our analysis of these enriched variants focusing on those candidates with the highest potential clinical interest. In summary, we will present the first comprehensive whole-exome evaluation of PLP variants enriched in the FC population. This data will help guide future efforts to extend this approach to a broader spectrum of communities in Quebec and to improve our understanding of these Mendelian conditions in terms of public health, understanding penetrance of enriched variants and disease pathophysiology, and for the discovery and development of innovative therapeutics.

Mendelian Phenotypes Posters - Thursday

PB2015. Whole-exome sequencing of Pakistani consanguineous families identified variants in novel and known genes of severe intellectual disability

Authors:

A. Abdullah¹, M. Asif², M. Anayat¹, Z. Ali³, G. Udin¹, G. Raja¹, M. Hussain²; ¹Univ. Inst. of Biochemistry and Biotechnology (UIBB), PMAS-Arid Agriculture Univ., Rawalpindi, Pakistan, ²Cologne Ctr. for Genomics (CCG), Univ. of Cologne, Faculty of Med. and Univ. Hosp. Cologne, Cologne, Germany, ³Ctr. for Biotechnology and Microbiol., Univ. of Swat, Mingora, Pakistan

Abstract Body:

Background: Intellectual disability (ID) is a condition of significant limitation of cognitive functioning and adaptive behaviour, with 50% of aetiology attributed to genetic predisposition. **Methods:** We recruited four consanguineous Pakistani families manifesting severe ID and developmental delay. The probands were subjected to whole exome sequencing (WES) and variants were further prioritized based on population frequency, predicted pathogenicity and functional relevance. **Results:** The WES data analysis of one of the families identified a homozygous missense variant of *CACTIN* (NM_001080543.1:c.1040A>T;p.(Asp347Val)), co-segregating with the disease phenotype. This variant is absent in gnomAD and 1000 Genomes and predicted to be 'deleterious' by several in silico tools. The gene is highly intolerant to missense and loss of function variants ($Z=3.2$ and $pLOF=0.14$) and has never been implicated in any inherited disorder. Its encoded Cactin which is a critical component of post-catalytic spliceosome that mediates exon ligation by stabilizing the position of the branch helix. *CACTIN* depleted cells show global splicing defects, while its knock down in zebrafish leads to embryonic lethality or extensive dysmorphogenesis. Remaining three families segregated presumably pathogenic novel variants in known ID genes; *MBOAT7* (NM_024298.4:c.757G>A;p.(Glu253Lys)), *CSTF2* (NM_001306206.1:c.445-8C>A) and *TRAPPC9* (NM_001160372.3:c.670delG;p.(Val224Cysfs*13)). **Conclusion:** Our findings indicate that *CACTIN* is novel gene that when disrupted leads to intellectual disability. These also provide additional evidence to the role of *MBOAT7*, *CSTF2* and *TRAPPC9* in causation of ID.

Mendelian Phenotypes Posters - Wednesday

PB2016. Whole-genome sequencing for comprehensive genetic modifier identification in CMT1A

Authors:

I. Xu¹, M. Dohrn¹, M. Danzi¹, M. Reilly², M. Shy³, S. Scherer⁴, Inherited Neuropathy Consortium, S. Zuchner¹; ¹Univ. of Miami, Miami, FL, ²UCL Queen's Square, London, United Kingdom, ³Univ. of Iowa, Iowa City, IA, ⁴Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Charcot-Marie-Tooth disease type 1A is caused by identical heterozygous duplications of *PMP22*. However, significant heterogeneity of phenotype severity has been observed in many studies. We hypothesize that genetic modifiers will explain some of the clinical variability in phenotypic expression. In this study, we will use whole-genome sequencing (WGS) as a comprehensive approach to identify genetic modifiers.

To do so, the study design encompasses four steps: 1. statistical analyses to define mild vs. severe phenotypes and conduct sample size estimations; 2. phenotype-based patient selection; 3. whole-genome sequencing; and 4. comprehensive bioinformatic analysis of genetic variant clusters in defined phenotype subsets. For the first step, we have analyzed previously collected clinical data sets retrieved from the RDCRN-INC data base (Inherited Neuropathy Consortium). Phenotype information were available in 2,280 patients out of 1,737 families, all confirmed with CMT1A. Including follow-up, we assessed 13,283 visits in total. To understand not only disease severity at baseline, but phenotype dynamics as well, we analyzed the yearly follow-up data for foot dorsiflexion strength (Medical Research Council Scale, MRC) and CMTES-2 (Charcot-Marie Tooth Examination Score Version 2), which were the two most complete data sets overall. Our results confirmed that both parameters correlate with disease duration and age and therefore represent the natural disease course, as well as a need to define mild and severe CMT1A patients with respect to their age. Our results confirmed that both parameters correlate with disease duration and age and therefore represent the natural disease course. Additionally, progression in CMTES-2 and foot dorsiflexion strength did not correlate with age giving the potential for us to use rapid or stagnant progression as a measure of severity. We decided to combine those two parameters in a minimal dataset, together with information on age at onset, age at examination, gender at birth, and ancestry.

Using this minimal data set, we will select suitable patients from collaborating clinical expert centers. Based on preliminary power analyses, we aim at collecting enough minimal datasets to categorize and select a cohort of 500 mildly and 500 severely affected patients for WGS. Interested colleagues are invited to contact us.

Genetic Therapies Posters - Wednesday

PB2017. A conditional knock-in preclinical mouse model for the *TRPV4* skeletal disorders.

Authors:

L. Nevarez, J. H. Martin, D. Diaz, D. Krakow, D. Cohn; UCLA, Los Angeles, CA

Abstract Body:

TRPV4 is a tetrameric non-selective ion channel that principally transports calcium. The gene is selectively expressed in chondrocytes but is also expressed at lower but biologically significant levels in other tissues, including the lung, bladder, peripheral neurons and endothelial cells. Dominant activating mutations in *TRPV4* result in a spectrum of skeletal dysplasias and peripheral neuropathies. The *TRPV4* skeletal disorders are unified by varying severity of short stature, progressive scoliosis, and shortened long bones. These disorders include brachyolmia, spondylometaphyseal dysplasia Kozlowski type (SMDK), and both non-lethal and perinatal lethal metatropic dysplasia. Missense mutations in each of these phenotypes has established the genetic basis of disease but mechanistic studies have been limited by lack of a viable animal model. In humans, the R594H *TRPV4* mutation has been shown to result in numerous genetically independent cases of SMDK. To better understand the mechanism of disease, we developed a conditional floxed knock-in mouse with the R594H mutation. Midgestational induction of the mutation in chondrocytes using *Col2a1-Cre* resulted in mice with short stature, short tails and long bones, craniofacial abnormalities and an increase cervical angle consistent with scoliosis. Postnatal induction at weaning using *Acan-CreER* and tamoxifen produced mice with short stature, short tails, short long bones and craniofacial findings but without the scoliosis observed in the *Col2a1-Cre* mice. Radiographs identified small epiphyses and widened metaphyses, especially at the proximal tibiae, concordant with the radiographic phenotype in SMDK and the other human *TRPV4* skeletal disorders. Growth plate histology revealed disorganized columns of proliferating chondrocytes, suggesting that the activated, mutant TRPV4 ion channels act at the level of chondrocyte proliferation to inhibit the extent and polarity of endochondral bone formation. Thus, this viable mouse model of the *TRPV4* skeletal disorders has facilitated understanding of the effects of *TRPV4* mutations on skeletal development not previously possible. The mouse will also serve as a preclinical translational model for testing inhibitor treatments to ameliorate disease progression and severity in the *TRPV4* skeletal disorders.

Genetic Therapies Posters - Thursday

PB2018. A cross-species approach using an *in vivo* evaluation platform in mice demonstrates that sequence variation in the human *RABEP2* gene modulates ischemic stroke outcomes.

Authors:

H. Lee¹, D. Marchuk²; ¹Duke Univ. Sch. of Med., Durham, NC, ²Duke Univ Sch. of Med., Durham, NC

Abstract Body:

Ischemic stroke, caused by vessel blockage, results in cerebral infarction; the death of brain tissue. Previously, quantitative trait locus mapping (QTL) of cerebral infarct volume and collateral vessel number identified a single, strong genetic locus regulating both phenotypes. Additional studies identified the causative gene, encoding RAB GTPase Binding Effector Protein 2 (*Rabep2*). However, there is yet no evidence that variation in the human ortholog of this gene plays any role in ischemic stroke outcomes. We established an *in vivo* evaluation platform in mice using adeno-associated virus (AAV) gene replacement and verified that both mouse and human RABEP2 rescue the mouse *Rabep2* KO ischemic stroke volume and collateral vessel phenotypes. Importantly, this cross-species complementation enabled us to experimentally investigate the functional effects of coding sequence variation in the human *RABEP2* gene. We chose four coding variants from the human population that are predicted by multiple *in silico* algorithms to be damaging to RABEP2 function. *In vitro* and *in vivo* analyses verify that all four led to decreased collateral vessel connections and increased infarct volume. Thus, there are naturally occurring loss-of-function alleles. This cross-species approach will expand the number of targets for therapeutics development for ischemic stroke.

Genetic Therapies Posters - Wednesday

PB2019. A drug repurposing screen to identify therapies for the rare disease DPAGT1-CDG

Authors:

H. Dalton, A. Berman, C. Chow; Univ. of Utah, Salt Lake City, UT

Abstract Body:

Glycosylation encompasses a wide class of biological pathways involving co- and post-translational sugar modifications. Mutations in glycosylation genes underlie Congenital Disorders of Glycosylation (CDGs) - ultra-rare disorders that can cause seizures, developmental delay, and early death. There are few treatment options available for CDGs, and small patient populations make clinical trials difficult. Thus, there is a great need for alternative approaches to finding treatments for these rare diseases. One such alternative is the use of drug repurposing screens that utilize libraries of small molecules with established safety profiles in humans, allowing for potentially faster patient turnaround.

DPAGT1-CDG is caused by mutations in the gene *DPAGT1*, which encodes the critical first enzyme for N-linked glycosylation. I created a model of DPAGT1-CDG in *Drosophila* using RNAi against *DPAGT1* in the fly eye to cause a small, rough eye phenotype. Using this model, I can assay for drugs that rescue this phenotype by quantitatively measuring its eye size. To find such therapies, I am performing a repurposing screen using 1,500+ small molecules that are 98% FDA/EMA-approved (Prestwick Chemical Library). Drugs are mixed into fly food, flies are exposed until early adulthood, and adult eye size is compared to control flies. The top candidate drugs that rescue the eye phenotype, resulting in a larger eye, are later validated by dose-response or genetic analyses.

As a pilot screen, I tested 240 drugs for their ability to rescue the DPAGT1-CDG model eye size and identified a 3.8% hit rate ($Z \geq 1.5$). My top candidate drug is bumetanide (+25% eye size, $Z=2.04$), which inhibits the Na-K-Cl cotransporter, NKCC1 (Ncc69 in flies). I genetically validated bumetanide by using RNAi against *NKCC1* which also increased eye size (+12.3%). This suggests that ion flux is important in *DPAGT1* deficiency and that this screen can identify new drugs capable of rescuing this disorder. Other hits of interest include the NSAID antipyrine (+20.5%, $Z=1.62$) and its metabolite 4-hydroxyantipyrine (+19%, $Z=1.51$). As these are metabolically-linked drugs that independently rescued eye size, it suggests the antipyrine pathway, or NSAIDs in general, may be important for rescuing DPAGT1-CDG. Taken together, this pilot screen has already revealed multiple candidate drugs that can rescue this model, and my current hit rate projects identifying 40+ more. I will present data on candidate drugs, their validation, and overlapping mechanisms of action from the full screen. These drugs may represent novel therapeutic options for DPAGT1-CDG.

Genetic Therapies Posters - Wednesday

PB2020. A Phase 1, Open-Label, Dose-Escalation Study to Evaluate the Safety and Efficacy of HMI-103, a One-Time Phenylalanine Hydroxylase Gene-Editing Vector in Adult Participants with Classical PKU Due to PAH Deficiency

Authors:

A. Sandhu, C. Simiele, L. Benard, D. Lamppu, J. Von Stetina, J. Thompson, T. Wright, N. Avila, M. Bailey, O. Francone, A. Seymour, J. Wright, J. Jordan; Homology Med.s, Inc., Bedford, MA

Abstract Body:

Phenylketonuria (PKU) is a rare autosomal-recessive inborn error of metabolism. If left untreated, phenylalanine hydroxylase (PAH) deficiency may result in progressive, irreversible neurological impairment. Neither a phenylalanine (Phe)-restricted diet nor currently available therapeutic treatments address the core biological defect of the disease—the presence of biallelic pathogenic variants in the *PAH* gene. This leaves a significant unmet medical need for patients with PKU due to PAH deficiency. HMI-103 is an investigational gene-editing vector designed to: 1) deliver normal copies of the *PAH* gene to hepatocytes, 2) integrate into target *PAH* locus in the genome via non-nuclease-based AAV-mediated homologous recombination, 3) produce the PAH enzyme (via the dual mechanism of integration via homologous recombination and episomal expression), and 4) restore Phe metabolism. This approach is supported by studies in the *Pah*^{enu2} mouse model in which blood Phe was reduced and maintained in juvenile and adult mice and in partially hepatectomized mice regrown with human hepatocytes. No clinical pathology, necropsy findings, or evidence of germline transmission were observed. The data demonstrated efficacy, integration, specificity for the target locus, and preclinical safety of HMI-103 and support initiation of a clinical trial. This Phase 1, open-label, sequential, dose-escalation study (pheEDIT) will evaluate the safety and efficacy of a one-time, intravenous administration of HMI-103 in adult participants aged 18–55 years with classical PKU due to PAH deficiency who have uncontrolled disease despite standard of care or marketed treatments. Three dose levels of HMI-103 will be investigated, with up to 3 participants for each dose cohort. A maximum of 9 participants will be enrolled in the study. Enrollment will be staggered between participants in a given cohort. To decrease potential for immune-response, pheEDIT will use a prophylactic immunosuppressive regimen consisting of a corticosteroid administered in combination with the T-cell inhibitor tacrolimus. The primary study endpoints include incidence and severity of treatment-emergent adverse events (TEAEs) and AEs of special interest, and the mean percent change from Baseline in plasma Phe concentrations within each dose cohort post administration of HMI-103. These will be assessed through Week 104 plus long-term follow-up. Once positive safety and efficacy results are demonstrated in the adult population, Homology plans to enroll younger participants into future studies.

Genetic Therapies Posters - Thursday

PB2021. A Structural Screen Approach and Molecular Simulation Identifies Potential Ligands Against the K700E Hot Spot Variant and Functional Pockets of SF3B1 to modulate splicing in Myelodysplastic Syndrome.

Authors:

R. Garcia, M. Atis, A. Cox, P. Koduru; UTSW, Dallas, TX

Abstract Body:

Myelodysplastic syndrome (MDS) is a blood disorder characterized by ineffective hematopoiesis and risk of acute myeloid leukemia. In terms of genetic defects, chromosome aberrations are detected in up to 60% of cases, notably chromosome 5 and 7 alterations and complex karyotypes. Somatic variants in several genes have been reported in up to 90% and genes associated with aberrant mRNA splicing are frequently mutated. Of these, SF3B1 is by far the most frequently mutated and these are associated with early clonal events and disease development. Current therapies of MDS involve lenalidomide, hypomethylating agents and regenerative hematopoietic stem cells to deter the effects of blood cytopenia. Recently, potential new therapeutic agents are under investigation. Despite these efforts, there is a need to develop novel therapies that target splicing factor alterations. Thus, the objective of the study is to identify potential small molecule modulators against the frequently mutated RNA splicing factor SF3B1(K700E) and functional allosteric sites by using a molecular structure-based approach and a molecular dynamic simulation. To identify potential SF3B1 modulators, we collected a series of chemical compounds from the Zinc and Enamine database. All compounds were converted to single file formats for screening and docking using Vina. Binding affinity scores generated by Vina less than -7.0 kcal/mol were further evaluated using docking suite Schrodinger. The binding affinity between SF3B1 and small molecule modulators was described by pKd values (> 5 were considered). The KDEEP package was used to generate pKd values and the H3B-8800 compound in phase-1 clinical trials for SF3B1 modulation was used as a control. To validate structure-based modeling, a molecular simulation was then performed using NAMD and VMD. A total of 190,000 compounds from the Zinc-Enamine database were collected. The virtual screen and molecular docking identified 23 compounds with affinity to E622, R625, H662, and E700 in SF3B1 with docking scores less than -7 kcal/mol and binding affinity scores greater or equal to 5. A 20-nanosecond simulation showed strong binding between selected compounds and key amino acid residues, including K700E. Total binding energies between these compounds and key residues ranged from -59.4 to -88.2 kcal/mol. In brief, small molecule modulators show strong binding to SF3B1 suggesting these compounds may be used against cells harboring the K700E variant or to preferentially kill neoplastic cells by targeting functional allosteric sites.

Key words: MDS, molecule modulators, molecular docking, molecular dynamics

Genetic Therapies Posters - Wednesday

PB2022. AAV delivery of *ELP1* exon-specific U1 snRNA rescues retinal degeneration in a mouse model of familial dysautonomia.

Authors:

E. Kirchner¹, A. Chekuri¹, J. Bolduc¹, E. Logan¹, M. Salani¹, A. Krauson¹, L. Vandenberghe², F. Pagani³, S. A. Slaugenhaupt¹, E. Morini¹; ¹Massachusetts Gen. Hosp., Boston, MA, ²Massachusetts Eye and Ear, Boston, MA, ³ICGEB, Trieste, Italy

Abstract Body:

Familial dysautonomia (FD) is a rare neurodegenerative disease caused by a splicing mutation in the Elongator complex protein 1 gene (*ELP1*). This mutation results in tissue-specific skipping of exon 20 in the mature mRNA transcript leading to reduced levels of ELP1 protein, mainly in the nervous system. FD patients exhibit diminished pain and temperature perception, proprioceptive ataxia, failure of blood pressure regulation, and recurrent hypertensive vomiting. In addition to this complex neurological phenotype, FD patients also have progressive retinal degeneration that severely affects their quality of life. They suffer from optic neuropathy featuring reduction of the retinal nerve fiber layer (RNFL) due to progressive loss of the macular retinal ganglion cells (RGCs). Patients often become visually impaired or legally blind after their third decade of life. Recently, we have shown that our phenotypic mouse model, the *TgFD9; ELP1^{Δ20/flox}* mouse, recapitulates the tissue-specific *ELP1* splicing defect while modeling the selective RGC loss and the optic neuropathy observed in FD patients. To restore correct exon 20 inclusion, we have developed a novel modified *ELP1* exon-specific U1 snRNA (ExSpeU1) that can promote the binding of spliceosomal machinery downstream of the mutated 5' splice site and restore correct splicing. The therapeutic efficacy of our approach in rescuing retinal degeneration was evaluated by intravitreal injection of our FD phenotypic mice with self-complementary adeno-associated vector capsid 2 (scAAV2.ExSpeU1) expressing ExSpeU1. Transduction efficiency in the retina was visualized using immunohistochemistry, and splicing correction was evaluated using RT-PCR. Our results show that intravitreal delivery of scAAV2.ExSpeU1 (1×10^{12} vg/ μ l) efficiently transduces RGCs and significantly improves *ELP1* splicing in the retina of scAAV2.ExSpeU1 injected mice (n=6) (p<0.001). Importantly, using Optical Coherence Tomography (OCT), we have demonstrated that scAAV2.ExSpeU1 injected eyes have increased thickness of RNFL compared to PBS injected eyes. For the first time, our study shows the tremendous therapeutic potential of intravitreal delivery of scAAV2.ExSpeU1 in stopping progressive retinal degeneration and preventing visual decline in FD patients.

Genetic Therapies Posters - Thursday

PB2023. Adjunct treatment with glycogen synthase (GYS1) antisense oligonucleotides and Enzyme replacement therapy (ERT) reduced glycogen in the Pompe disease mouse model

Authors:

V. Kimonis¹, W. Weiss¹, M. Carrer², A. Shmara¹, K. Nguyen¹, C. Cheng¹, L. Hettrick³, A. Watt³, N. Raben⁴, P. Jafar-Nejad³; ¹Univ. of California, Irvine, Irvine, CA, ²Ionis Pharmaceuticals, Inc., Carlsbad, CA, ³Ionis Pharmaceuticals, Carlsbad, CA, ⁴NIH, Bethesda, MD

Abstract Body:

Pompe disease is a progressive myopathy resulting from the deficiency of acid α -glucosidase (GAA). Enzyme replacement therapy (ERT) with recombinant human (rh) GAA works well in alleviating the cardiomyopathy; however, many patients continue to have progressive muscle weakness from muscle glycogen accumulation produced by muscle glycogen synthase (GYS1). Previous studies have provided proof of principle that knockdown of GYS1 mRNA by phosphorodiamidate morpholino oligonucleotide reduced glycogen. To impart specificity for the muscle variant of the enzyme, we tested over 150 antisense oligonucleotides (ASOs) targeting mouse *Gys1* designed and previously screened in vitro and in vivo by Ionis Pharmaceuticals. In this report, we focused on the three lead molecules to identify the most efficacious *Gys1* ASOs in *Gaa*^{-/-} mice. The results from treatment with the three *Gys1* ASOs (16- doses of 25 mg/kg/weekly) identified *Gys1* ASO#3 as the most effective with 100% downregulation of GYS1 protein, and 49% and 33% efficiency in clearing muscle glycogen accumulation in 1-month and 3-month-old *Gaa*^{-/-} Pompe mice respectively. Combination therapy with *Gys1* ASO#3 and ERT further reduced glycogen accumulation and effectively alleviated the massive autophagic buildup in *Gaa*^{-/-} mice quadriceps to wild type mice level. The reversal of lysosomal and autophagic pathologies led to improved muscle function in treated *Gaa*^{-/-} mice. These results provide proof of principle that using GYS1 inhibitors as adjunct therapy with ERT is beneficial and will lead to future clinical trials in patients with Pompe disease.

Genetic Therapies Posters - Thursday

PB2024. An allele- discriminative approach for efficient and specific CRISPR/Cas9 based gene therapy of Late Onset of Alzheimer's disease

Authors:

B. Kantor, J. Rittiner, O. Chiba-Falek; Duke Univ., Durham, NC

Abstract Body:

APOEε4 is well-established genetic risk factor for late onset Alzheimer's disease (LOAD). It has been demonstrated that transcriptional repression stemming from the formation of closed chromatin organization around ApoE loci could play an important role in controlling the levels of APOEε4 gene expression. It also has been shown that reduction of APOEε4 levels may alleviate LOAD pathology. Here, we aimed to develop a novel approach for specific and accurate modulation of APOEε4 gene expression. We took advantage of the engineered VRER-Cas9/gRNA system which specifically and accurately recognized a novel protospacer adjacent motif (PAM) created with SNP-rs429358 T-C which defines APOEε4 haplotype. The system is based on an all-in-one lentiviral vector harboring gRNA/VRER-dCas9-repressor effector. Using this vector, we were able to efficiently and selectively reduce expression of ApoEε4 in vitro, in APOEε4-human induced pluripotent stem cell (hiPSC)-derived excitatory neurons, and in vivo, in mice. The developed epigenome-editing platform highlights the novel approach towards the development of next-generation of gene therapies for LOAD.

Genetic Therapies Posters - Wednesday

PB2025. An allele specific strategy for targeting mutant senataxin in patients with amyotrophic lateral sclerosis type 4

Authors:

C. Grunseich¹, B. Johnson¹, A. Apfel¹, D. Li², K. Fischbeck¹, A. Winkelsas¹, V. Cheung²; ¹NINDS, Bethesda, MD, ²Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Amyotrophic lateral sclerosis type 4 (ALS4) is an autosomal dominant form of motor neuron disease caused by mutations in the senataxin (*SETX*) gene. The onset of ALS4 is often in the teenage years and is characterized by a slowly progressive course. To date, there is no treatment for ALS4; here, we present data from an ongoing natural history study of ALS4 and an RNA-based method to silence the mutant allele of *SETX*. Senataxin is an RNA/DNA helicase that resolves the three-stranded nucleic acid structures called R-loops. We have previously shown that the gain of function senataxin mutations leads to fewer R-loops in ALS4 patients. In a natural history study, we follow ALS4 patients annually to assess their muscle function, strength, and collect skin samples to measure R-loop abundance in fibroblasts. The disease progression is also evaluated by nerve and muscle imaging. The c.1166 T>C (p.Leu389Ser) mutation leads to ALS4 in our patients. We designed a set of siRNAs to target this mutant allele. To assess the efficiency of allele-specific silencing, we measured allele-specific gene and protein expression in cell-based assays. First, we screened for siRNAs in HEK293 cells with a construct containing WT or the Leu389Ser senataxin fused to a Halo epitope tag. We then transfected each of these siRNAs into patients' primary fibroblasts, and two of these siRNAs showed allele-specific knockdown of the mutant *SETX* allele at the transcript and protein levels by approximately 50%. Molecular characterization is now underway. In this presentation, we will describe the disease course of this form of juvenile-onset ALS and data from RNA-based precision treatment.

Genetic Therapies Posters - Thursday

PB2026*. An alternate translational start site allows 5' nonsense variants to escape NMD so that readthrough compound ELX-02 and modulators can restore CFTR function.

Authors:

A. Bowling¹, C. Merlo², G. Lin¹, N. West², S. Patel², G. Cutting¹, N. Sharma¹; ¹Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ²Johns Hopkins Hosp., Baltimore, MD

Abstract Body:

Variants predicted to introduce a premature termination codon (PTC) are targeted for degradation by nonsense mediated mRNA decay (NMD), resulting in unstable mRNA and lack of protein. However, initiation of translation at start sites downstream of 5' nonsense variants can remove exon junction complexes that engage NMD, resulting in stable transcript. Using cystic fibrosis (CF) as a model system, we evaluated whether stable transcript bearing 5' nonsense variants can be targeted by readthrough agents to introduce an alternate amino acid at the PTC, allowing for production of a full-length CFTR protein and augmentation of function with CFTR modulator drugs.

Expression minigenes (EMGs) containing the complete *CFTR* cDNA and select flanking intron sequences were generated bearing variants E60X, L88X, and Y122X. Immunoblotting showed that HEK293 cells transiently expressing E60X and L88X *CFTR*-EMG generated a truncated CFTR-specific core glycosylated product, consistent with downstream translation initiation. Alternative translation initiation methionine at codons 150, 152 and 265 were individually mutated to valine. Substitution of M265 did not affect processing of full-length CFTR, but did cause loss of truncated CFTR protein in cells expressing E60X, L88X, and Y122X *CFTR*-EMGs, indicating that downstream translation initiation is occurring at M265. Substitution of M150 and M152 did not affect production of truncated CFTR protein.

Primary human nasal epithelial (NE) cells from individuals carrying L88X were conditionally reprogrammed and grown at air liquid interface. RT-PCR of WT/L88X NE cells revealed that L88X mRNA is stable and comparable to WT levels. WT/WT NE cells generated CFTR function of $13.1 \pm 1.5 \mu\text{A}/\text{cm}^2$ (n=12). Untreated L88X/F508del NE cells generated 6% of WT/WT function. CFTR modulators restored function to 83% of WT due to recovery of F508del CFTR. CFTR function increased further to 123% of WT when readthrough compound ELX-02 was included, which we attribute to activation of CFTR generated by translational readthrough of L88X-*CFTR* transcripts. Of note, treatment with the classic readthrough compound G418 and modulators did not result in additional recovery of function. Our results demonstrate that nonsense variants that naturally evade NMD via mechanisms such as downstream translation initiation are ideal targets for readthrough therapies.

Genetic Therapies Posters - Wednesday

PB2027*. An *in vivo* screen identifies small molecule modulators of retinitis pigmentosa and the ER stress response

Authors:

K. Hope, A. Berman, C. Chow; Univ. of Utah, Salt Lake City, UT

Abstract Body:

Retinitis pigmentosa (RP) is a common form of retinal degeneration caused by mutations in the *Rhodopsin* gene. Misfolded rhodopsin protein accumulates in the endoplasmic reticulum (ER) and causes ER stress. Cells respond to ER stress by initiating the unfolded protein response (UPR) that upregulates chaperone protein expression, increases the degradation of misfolded proteins, and inhibits protein translation. Failure to effectively manage ER stress and restore homeostasis results in cellular dysfunction and ultimately apoptosis, a process implicated in RP and other human diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), and spinocerebellar ataxia-3 (SCA3), among others. Identifying small molecules that modulate ER stress may be effective therapeutics for human diseases caused by misfolded protein accumulation. Here, we used a *Drosophila* model of RP that expresses misfolded rhodopsin protein, Rh1^{G69D}, in the eye. Rh1^{G69D} expression induces chronic ER stress and apoptosis, resulting in photoreceptor neuron loss and reduced eye size. We took a drug repurposing approach and screened 1520 small molecules from the Prestwick Chemical Library, the majority of which are FDA-approved, to identify compounds that modulate photoreceptor neuron death in Rh1^{G69D} expressing flies. ER stress-enhancing compounds may reveal novel ER stress pathways, and compounds that suppress ER stress are potential therapeutic candidates for RP and other diseases with an ER stress component. Our *in vivo* repurposing screen identified multiple classes of drugs that enhance or suppress the degenerate eye phenotype, including compounds acting through monoamine neurotransmitters, antibiotics, sodium channels, and the renin/angiotensin pathway. Using an RNAi approach, we found that knockdown of components of the *Drosophila* renin/angiotensin system restored eye size and suppressed the ER stress-induced phenotype in Rh1^{G69D} expressing flies. We will also present data on whether compounds that rescue cell death in the RP model also rescue disease-associated phenotypes in other *Drosophila* models of protein misfolding diseases, such as AD, HD, and SCA3. This work identified potential therapeutic compounds for RP and possibly other human diseases that result from misfolded protein accumulation and ER stress.

Genetic Therapies Posters - Thursday

PB2028. Antisense oligonucleotide-based exon skipping therapeutic strategy for Cohen syndrome.

Authors:

M. Ansar^{1,2}, F. Vacca³, R. Da Costa¹, H. M. A. Baig^{1,4}, E. Lize^{1,5}, I. Anwar¹, M. Choung¹, P. Tiwari¹, L. Faivre⁵, H. Riezman³; ¹Dept. of Ophthalmology, Univ. of Lausanne, Jules-Gonin Eye Hosp., Fondation Asile des Aveugles, Lausanne, Switzerland, ²Advanced Molecular Genetics and Genomics Disease Res. and Treatment Ctr., Dow Univ. of Hlth.Sci., Karachi, Pakistan, ³Dept. of Biochemistry, NCCR Chemical Biology, Univ. of Geneva, Geneva, Switzerland, ⁴Dept. of Biotechnology, Inst. of Biochemistry, Biotechnology and Bioinformatics, The Islamia Univ. of Bahawalpur, Bahawalpur, Pakistan, ⁵Genetics of Dev.al Disorders Team, INSERM UMR 1231, Univ. of Bourgogne Franche-Comté, Dijon, France

Abstract Body:

Cohen syndrome (CS) is an autosomal recessive disorder caused by biallelic loss-of-function mutations in the *VPS13B* gene. To date, more than 200 causative mutations in around 1000 affected individuals have been reported worldwide. CS is characterized by developmental delay, intellectual disability, metabolic syndrome manifestations and visual impairment due to severe myopia, progressive retinal degeneration, maculopathy, and cataract. Since CS is largely manifested by biallelic null mutations and missense variants are rarely pathogenic ($p=0.005$), it is envisioned that skipping short in-frame exons, lying out of the key functional domains of *VPS13B* would not affect the protein function, alternatively, a partially functional protein could be enough to improve the disease phenotypes. Antisense oligonucleotide (AON)-mediated exon skipping as a therapeutic approach has been extensively studied in the past decade, including for the treatment of retinal and muscular dystrophies. Existing clinical applications of AON-mediated exon skipping in Duchenne muscular dystrophy and Usher syndrome encouraged us to develop a similar approach for CS. We propose a therapeutic strategy based on skipping of in-frame exons carrying loss-of-function variants and develop an *in vitro* cell model to test the effectiveness of this strategy for each in-frame exon of the *VPS13B* gene. AONs were designed, using well-described parameters, to target the predicted enhancer splice sites of each of the 24 in-frame exons of *VPS13B*. Exon skipping efficiency of AONs was optimized in HeLa cells. Exon skipping was monitored and confirmed by RT-PCR and Sanger sequencing of the cDNA. Potential knock-down effects of AONs were ruled out by qPCR quantification of the transcripts and deleterious effects on cell viability with an XTT assay. Overall, 16 out of 24 AONs were efficient enough to induce over 25% skipping of their target exon. The ability of potential AONs to correct the aberrant Golgi morphology of *VPS13B*-deficient cells is currently being tested. For this purpose, HeLa cell lines were produced by introducing indel mutations in each of the 24 in-frame exons using CRISPR/Cas9 technology. Potential AONs will also be tested in available patient-derived fibroblasts for the corresponding exons. A high-throughput microscopy-based assay with an automated phenotypic analysis pipeline was developed which very efficiently discriminates the Golgi morphology of mutant cells.

Genetic Therapies Posters - Thursday

PB2029. Characterization of novel mouse model with 35 bp deletion in *Champ1* gene and exploration of AAV gene therapy to correct syndromes arising from mutations in the *CHAMP1*

Authors:

T. Carneiro¹, M. Bakay¹, B. Strenkowski¹, L. Sertori¹, R. Pellegrino², H. Hakonarson¹; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Introduction: Chromosome alignment maintaining phosphoprotein 1 (CHAMP1) is crucial for kinetochore-microtubule attachment and for chromosome segregation during mitosis, important process for neurodevelopment. Mutations in *CHAMP1* were recently associated with global developmental delay, Intellectual disability, hypotonia and dysmorphic features. Reported mutations result in a premature stop codon and loss of protein c-terminal region which is essential for CHAMP1 localization and protein function. CHAMP1 is also necessary for the localization and activity of other proteins in mitosis process, such as MAD2L2 and POGZ. Additional studies demonstrated that *CHAMP1* depletion in cell lines resulted in *MCL-1* and *BCL-2* reduce expression, therefore this protein might be crucial for cell survival. We propose a gene therapy treatment using cell line and animal model in order to rescue the phenotype developed due to *CHAMP1* haploinsufficiency. This approach is also a way to acknowledge whether the role of CHAMP1 in cell survival is related to the pathogenesis of intellectual disability, a question that remains unanswered. **Methodology:** To mimic CHAMP1 dysfunction, both cell line and animal model have a deletion in the third exon. The base pair deletion in neuroblastoma cell line (SH-SY5Y) will be reached through CRISPR/SpCas9 (SYNTHEGO[®], Redwood City, CA, USA). Knockout (KO) efficiency will be assessed through DNA sequencing. Cells with *CHAMP1* deletion will be treated with AAV vector (Penn vector Core, University of Pennsylvania, Philadelphia, PA, USA). The Mice model was developed (C57BL/6NCr1-Champ1em1(IMPC)Mbp/Mmucd) using CRISPR/Cas9 gene editing technology in mouse zygotes (Mouse Biology Program, UC Davis, Davis, CA, USA). **Expected Results:** Growth rate, viability, *CHAMP1* gene expression and CHAMP1, MAD2L2, POGZ, BCL-2 and BAX protein expression will be accessed in cells lines knocked out for CHAMP1, untreated and treated with AAV. Furthermore, immunoprecipitation will be held for MAD2L2 and POGZ in both cases. Mice colony is being breed; females presented difficulty in giving birth due to occurrence of dystocia. Wild-type and C57BL/6NCr1-Champ1em1(IMPC)Mbp/Mmucd mice will have behavior and molecular characterization assessed before and after the treatment with AAV through Intra-venous injections. In conclusion, we propose to explore the use of AAV gene therapy to correct phenotype arising from mutations in *CHAMP1* gene in mouse model. We believe that results from this study will be useful and informative in understanding Intellectual disability (ID) - related disorders and in exhibiting the feasibility of gene therapy for ID.

Genetic Therapies Posters - Wednesday

PB2030. Comparison of AAVmyo and MyoAAV as promising vectors to deliver gene therapy in the VCP^{R155H} KI mouse model

Authors:

L. Weiss; Univ. of California, Irvine, Irvine, CA

Abstract Body:

Background and Objective Pathogenic variants in Valosin Containing Protein (*VCP*) gene cause a unique autosomal dominant disease characterized by inclusion body myopathy, Paget disease of bone and frontotemporal dementia (also known as multisystem proteinopathy (MSP)). *VCP* pathogenic variants lead to hyperactive enzymatic activity, suggesting a gain-of-function. Concomitantly, overexpression of *VCP* pathogenic variants in cells dominantly interfere with autophagy and endolysosomal sorting. To ameliorate the gain-of-toxicity of VCP mutant proteins, an ideal approach is to silence the mutant gene in an allele-specific manner and to leave the wildtype allele intact. We therefore propose to silence VCP with strategies such as microRNA, antisense oligonucleotides and CRISPR technology. A limitation, however, is obtaining sufficient expression in muscle to target the VCP mutant allele. Recently AAVmyo and MyoAAV have emerged as promising vectors to deliver gene targeting strategies to muscle tissue. We have performed studies in the VCP KI mouse model to study the efficacy and immunological response to these vectors as a mean to deliver gene modifying strategies in muscle. **Methods** To determine the tropism of the viral vectors through systemic injection, we used C57BL/6 WT and VCP R155H/+ mice at age 6-8-week-old to deliver a single dose of 1×10^{12} and 4×10^{12} vector genomes (vg) of either AAVmyo-CAG-EGFP, MyoAAV4A-pAM-CAG-TdTomato, or AAV9-CAG-EGFP by retro-orbital injection. After 3 weeks of regular monitoring for viability, immunity and signs of toxicity, the mice were sacrificed. All muscle groups in addition to multiple organs were harvested and analyzed to detect EGFP/TdTomato expression to determine muscle tropism. Serum was collected to examine markers of liver, renal and muscle toxicity. **Results** EGFP/TdTomato expression was significantly higher in both doses of AAVmyo and MyoAAV compared to AAV9 in various muscle groups. Our preliminary studies indicated better muscle tropism and less tropism for the liver with MyoAAV deeming it safer and muscle specific. We plan to combine our gene modifying strategies with MyoAAV for future studies in the mutant mice to study the effect on muscle pathology and function. **Discussion** Preliminary results show that using MyoAAV is a promising approach to deliver gene therapy to muscle tissue -a most challenging tissue to target. This leads to an exciting next step of targeting the mutant VCP allele and if needed replacing with WT VCP in the mouse model. If successful, this strategy has great translational potential in correcting disease pathology in VCP and other autosomal dominant disorders.

Genetic Therapies Posters - Thursday

PB2031. Deep phenotyping APOC3 Knockouts in a population with high consanguinity

Authors:

S. Khalid¹, M. Z. Khan², L. Wu³, M. Jahanzaib², S. Saeed², R. Mian², M. Bilal², R. Sultana⁴, S. Abbas⁵, M. Hussain⁶, F-u. Memon⁷, T-u. Salam⁸, M. Ishaq⁴, P. Frossard², A. Rasheed², J. Millar⁹, D. Rader¹⁰, D. Saleheen¹¹; ¹Columbia Univ. Irving Med. Ctr., NY, NY, ²Ctr. for Non-Communicable Diseases, Karachi, Pakistan, ³Univ. of Pennsylvania, PA, PA, ⁴Karachi Inst. of Heart Diseases, Karachi, Pakistan, ⁵Faisalabad Inst. of Cardiology, Faisalabad, Pakistan, ⁶Services Hosp., Lahore, Pakistan, ⁷Red Crescent Inst. of Cardiology, Hyderabad, Pakistan, ⁸Jinnah Hosp., Lahore, Pakistan, ⁹Univ. of Pennsylvania, Philadelphia, PA, ¹⁰Univ of Pennsylvania, Philadelphia, PA, ¹¹Columbia Univ. Irving Med. Ctr., New York, NY

Abstract Body:

Apolipoprotein C-III (APOC3) plays an integral role in the regulation of triglyceride rich lipoproteins, by inhibiting the clearance of triglycerides carried by VLDL and chylomicron remnants in the blood. Prior work has shown that heterozygous loss of function (LOF) carriers of APOC3 have lower triglyceride levels and a lower risk of coronary artery disease (CAD). The quantitative impact of disease risk remains unknown in complete knockouts (KOs). Additionally, the effects and safety implications of complete APOC3 LOF have not been characterized. APOC3 inhibition is an active therapeutic strategy to lower CAD risk; hence these questions have therapeutic relevance.

Among 37,244 unrelated sequenced individuals, including 19,681 cases of myocardial infarction (MI), in the Pakistan Genomic Resource - a biobank with high levels of consanguinity - we identified 207 heterozygous LOF carriers and 14 KOs. As expected, the KOs had undetectable APOC3 levels. We also observed a decrease in plasma triglycerides ($P = 9E-85$), VLDL-C levels ($P = 2E-73$), APOE levels ($P = 5.4E-9$) and an increase in HDL-C levels ($P = 1E-34$) and APOA1 levels ($P = 6E-5$) consistent with a gene-dosage effect. We observed a significant decrease in the risk of MI among heterozygous carriers ($P = 0.01$); however, we did not observe any protection from MI risk in complete KOs. Conversely, we observed a non-significant increase in the risk of MI in complete KOs compared to non-carriers; of the 14 knockouts identified, 9 were found to have MI. The loss of protection against MI could not be explained by the genetic background or by the increase in levels of homozygosity of the KOs. By recalling complete KOs and their family members, we were able to identify and phenotype an additional 33 complete KOs and 152 heterozygotes and assess other safety concerns related to complete APOC3 inhibition (i.e., glucose intolerance, fat content in the liver, etc.). In conclusion, by leveraging a highly consanguineous cohort, we have identified and phenotyped APOC3 KOs that have, hitherto, not been identified elsewhere. We did not observe APOC3 LOF to confer protection in complete KOs and observed other biomarker and phenotypic associations; these findings should inform existing therapeutic programs targeting APOC3.

Genetic Therapies Posters - Wednesday

PB2032. Development of an AAV gene therapy for GNE myopathy: AAV9-CK8e-SV40-*hGNE1-V5* shows robust GNE expression in mouse muscle tissue.

Authors:

K. Koczwara¹, P. Leoyklang², A. Lek³, A. DeSimone¹, S. Mitrani-Rosenbaum⁴, M. Huizing², J. Crudele⁵, M. Lek¹; ¹Yale Univ., New Haven, CT, ²NIH, Bethesda, MD, ³Muscular Dystrophy Association, Chicago, IL, ⁴Hebrew Univ. of Jerusalem, Jerusalem, Israel, ⁵Univ. of Washington, Seattle, WA

Abstract Body:

GNE myopathy is a rare, autosomal recessive disease, leading to progressive skeletal muscle atrophy and weakness in the limbs. Myopathy occurs due to pathogenic mutations in the *GNE* gene, resulting in reduced enzymatic activity of the GNE protein, a bifunctional enzyme involved in the sialic acid biosynthesis pathway. To date, over 250 pathogenic mutations in *GNE* have been reported, making mutation correction a poor therapeutic approach. While oral supplementation of N-acetyl-D-mannosamine (ManNAc), the substance produced by GNE within the sialic acid biosynthesis pathway, may be a viable therapeutic strategy, its efficacy in treating GNE myopathy patients is still being evaluated (ClinicalTrials.gov: NCT04231266). Alternatively, delivery of a fully functional *GNE* gene to muscle tissue should restore proper GNE function, regardless of the particular pathogenic mutation. AAV-based gene delivery has proven successful for the treatment of other genetic neuromuscular diseases, and the relatively small size of the *GNE* coding sequence (<2.2kb) make this approach an attractive treatment strategy for GNE myopathy.

In this study we designed and optimized a robust gene therapy for GNE myopathy. A library of AAV vectors was developed to evaluate changes in GNE expression due to: (1) the promoter used; (2) the inclusion of an intron; (3) the GNE isoform; and (4) the position and type of protein tag. *In vitro* results determined that the CK8e-SV40-*hGNE1-V5* construct resulted in the highest muscle-specific GNE gene and protein expression. This construct was then packaged into AAV9 particles and a biodistribution study was performed in wild type mice to determine the safety and efficiency of GNE expression across a variety of tissues. C57Bl/6 4-week old mice were injected with 1e13 or 5e13 vg/kg AAV9-CK8e-SV40-*hGNE1-V5*, and human *GNE* expression was analyzed at 8-weeks post-injection. Robust, dose-dependent GNE gene and protein expression was observed in a variety of muscle tissues, supporting further investigation of the efficacy of AAV9-CK8e-SV40-*hGNE1-V5* in GNE myopathy cell and mouse models.

Genetic Therapies Posters - Thursday

PB2033. Dkk1 inhibition normalizes limb phenotypes in a mouse model of *Fzd2* associated omodysplasia Robinow syndromes.

Authors:

R. Stottmann^{1,2}, M. Michalski³, R. Liegel⁴, S. Vaidya⁴, E. Bittermann⁴, E. Finnerty⁴, C. Menke⁴, C. Diegel³, Z. Zhong³, B. Willaims³; ¹Nationwide Childrens Hosp., Columbus, OH, ²Ohio State Univ., Columbus, OH, ³Van Andel Res. Inst., Grand Rapids, MI, ⁴Cincinnati Childrens Hosp. Med. Ctr., Cincinnati, OH

Abstract Body:

State of the art genomic approaches have revolutionized our discovery of Mendelian phenotypes but these still are most effective when paired with mechanistic experiments to directly address pathogenicity. Fortunately, revolutionary genome editing approaches have recently made this a significantly more tractable experimental approach. Frizzled-2 (*FZD2*) is a transmembrane receptor mediating the Wnt signaling pathway. We previously identified a pathogenic human variant of *FZD2* (p.W548*) coding for a premature stop and loss of 17 amino acids including a portion of the consensus DISHEVELLED (DVL) binding sequence (KTxxxW) required for Wnt signal transduction. This precise *FZD2* variant and 6 others have since been identified in 14 patients with similar phenotypes and this can now be called *FZD2* associated omodysplasia Robinow Syndrome. Patients with the p.W548* variant exhibited autosomal dominant omodysplasia and incompletely penetrant cleft palate, whereas a previously published deletion of *Fzd2* in mouse exhibited incompletely penetrant (50%) recessive cleft palate. To model the p.W548* variant, we utilized zygote microinjection and i-GONAD-based CRISPR/Cas9-mediated genome editing to generate an allelic series of 3 mouse mutations recapitulating the patient variant or creating a small deletion. Embryos mosaic (50-90%) for humanized p.W553* knock-in exhibited cleft palate, decreased weight, and shortened limbs, consistent with p.W548* patients. Additionally, we generated two germline mouse alleles with small deletions, *Fzd2*^{D3} and *Fzd2*^{D4}. Homozygotes for each allele survive embryonic development at normal ratios but exhibit a >90% penetrant recessive phenotype of cleft palate, a wide/short snout and perinatal lethality. Homozygotes for both alleles also had shorter fore- and hindlimb bones compared to wild-type embryos. Craniofacial tissue isolated from *Fzd2*^{D4} embryos had decreased canonical WNT signaling. In utero treatment with IIC3a (DKK inhibitor) rescued the limb lengths in *Fzd2*^{D4} homozygotes. The in vivo replication of clinical features seen in patients represents an approach which may be used to further investigate the mechanism of the autosomal dominant omodysplasia phenotypes and validates the utility of CRISPR knock-in mice as a tool for demonstrating pathogenicity of human genetic variants. We also present evidence for a potential therapeutic intervention for *FZD2* associated autosomal dominant Robinow Syndrome.

Genetic Therapies Posters - Wednesday

PB2034. Down-regulation of *SCN8A* as treatment for developmental and epileptic encephalopathy

Authors:

W. Yu¹, S. Hill¹, M. Meisler²; ¹Univ. of Michigan, Ann Arbor, MI, ²Univ Michigan, Ann Arbor, MI

Abstract Body:

Developmental and epileptic encephalopathies (DEEs) are rare genetic disorders characterized by refractory seizures, developmental delay, and sudden death. *De novo* mutations of *SCN8A*, encoding the voltage-gated sodium channel Na_v1.6, have been associated with DEE (OMIM #614588). The most common pathogenic mechanism of *SCN8A*-DEE is a gain-of-function mutation causing premature channel opening or impaired channel inactivation. Nav1.6 is localized at the axon initial segment of neurons, where it regulates the initiation of action potentials. GOF mutations in *SCN8A* result in excess neuronal firing of action potentials, leading to seizures. In a mouse model expressing the DEE mutation R1872W, seizures were delayed by intracerebroventricular administration of an *SCN8A* ASO (Lenk et al, Ann. Neurol. 2020). The effect of the gapmer ASO was transient, requiring readministration after 6 weeks. We are evaluating two genetic therapies for longer-lasting effects. An AAV10-shRNA virus down-regulates *Scn8a* expression via RNA interference (Wong et al, Sci. Repts. 2018). Intracerebroventricular injection of the virus at postnatal day 1 prevented the onset of seizures and extended survival of *Scn8a*-DEE mice. The *Scn8a*-shRNA also rescued survival of Dravet Syndrome mice with haploinsufficiency of *Scn1a*. Stereotactic administration of the virus to hippocampus of adult mice resulted in 60% reduction of *Scn8a* transcript, demonstrating the potential for regional application of shRNA therapy. To evaluate the effectiveness of an allele-specific knockout of the mutant allele, we are targeting the mouse N1768D allele, which contains 8 silent substitutions close to the single pathogenic nucleotide substitution. sgRNAs with 3 or 4 nucleotides that are specific for the N1768D allele are being tested in *Scn8a*-N1768D/+ heterozygous mice to determine whether allele-specific inactivation could be an effective treatment for *SCN8A*-DEE.

Supported by NIH R01 NS34509 and the Dravet Syndrome Foundation

Genetic Therapies Posters - Thursday

PB2035. Drug repositioning network in rare and intractable diseases based on drug target gene analyses.

Authors:

R. Sakate, T. Kimura; Natl. Inst.s of BioMed. Innovation, Hlth.and Nutrition, Ibaraki-shi, Osaka, Japan

Abstract Body:

Interest in drug repositioning has been growing because it can help reduce the time, cost, and risk involved in drug developments. We analyzed the drug repositionability among rare and intractable diseases. Drug development for these diseases has been challenging for decades due to the low prevalence and insufficient information. Drug repositioning opportunities lie in identifying promising pairings of already-approved drugs and new therapeutic diseases. We established a systematic approach to identify such pairings. First, we created a list of rare and intractable diseases based on the designated intractable diseases in Japan ("Nanbyo" in Japanese). Second, from clinical trial data worldwide (ClinicalTrials.gov, EU-CTR, ChiCTR, and JPRN), drug development information for the diseases in the list was extracted. Then the extracted drugs were integrated with drug target gene information. As a result, we found that ~1,600 drug repositioning events have been occurring in rare and intractable diseases during the past 20 years. These events also occurred between diseases from different disease groups, such as neuromuscular diseases (multiple sclerosis, etc.) and digestive diseases (Crohn's disease, etc.). We investigated a score to represent a degree of drug repositionability by drug target gene analyses combining other information such as pathways. Drugs and their target genes are keys to explore the repositionable drugs, since sharing drug target genes indicates the common mechanism of drug action between the diseases. Drug repositioning network can be drawn by binding diseases with a positive repositionability score. This network well visualizes promising disease pairs for drug repositioning and facilitates drug development for rare and intractable diseases. This analysis was based on DDrare, the database of drug development for rare diseases (https://ddrare.nibiohn.go.jp/index_e.html).

Genetic Therapies Posters - Wednesday

PB2036. Dynamin-2 interacts with SPEG and its reduction rescues the skeletal myopathy of SPEG-deficient mouse model

Authors:

Q. Li¹, J. Lin¹, J. Widrick¹, S. Luo¹, G. Li², Y. Zhang¹, J. Laporte³, M. Perrella², X. Liu², P. Agrawal⁴; ¹Boston Children's Hosp., Harvard Med. Sch., Boston, MA, ²Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, ³IGBMC, Illkirch, France, ⁴Boston Children's Hosp., Harvard Med. Sch., Boston, MA

Abstract Body:

Striated preferentially expressed protein kinase (*SPEG*), a myosin light chain kinase, is mutated in centronuclear myopathy (CNM) and/or dilated cardiomyopathy. No precise therapies are available against this disorder, and gene replacement therapy is not a feasible option due to the large size of *SPEG*. We evaluated the potential of dynamin-2 (DNM2) reduction as a therapeutic strategy as it has been shown to revert muscle phenotypes in mouse models of CNM caused by *MTM1*, *DNM2*, and *BIN1* mutations. We determined that *SPEG* interacts with DNM2, and *SPEG* deficiency causes an increase in DNM2 levels. The DNM2 reduction strategy in *Speg*-KO mice was associated with an increase in life span, body weight, and motor performance. Additionally, it normalized the distribution of triadic proteins, triad ultrastructure, and triad number, and restored phosphatidylinositol-3-phosphate levels in *SPEG*-deficient skeletal muscles. While DNM2 reduction rescued the myopathy phenotype, it did not improve cardiac dysfunction, indicating a differential tissue-specific function. Combining DNM2 reduction with other strategies may be needed to target both the cardiac and skeletal defects associated with *SPEG* deficiency. DNM2 reduction should be explored as a therapeutic strategy against other genetic myopathies (and dystrophies) associated with a high level of DNM2. We are utilizing transcriptomic and proteomic approaches to decipher the underlying mechanisms of rescue due to DNM2 reduction.

Genetic Therapies Posters - Thursday

PB2037. ECLIPSE, an automated CRISPR platform for the large-scale generation of cell models for the iPSC Neurodegenerative Disease Initiative (iNDI)

Authors:

P. Deng, A. King, B. Saavedra, M. Morell, A. Chair, D. Sailor, A. G. Gianotti, K. Holden; Synthego, Redwood City, CA

Abstract Body:

The National Institutes of Health led iPSC Neurodegenerative Disease Initiative (iNDI) is the largest iPSC genome engineering project attempted with the goal of generating a widely available and standardized set of diseased cell models for over 100 single nucleotide variants (SNV) mutations associated with Alzheimer's disease and related dementias (ADRD) in isogenic iPSC lines. The standardization of cell models is of vital importance for the generation of reproducible and actionable data in therapeutic development. As part of a multi-institution collaboration, Synthego was selected for the generation of 25 SNVs in the candidate KOLF2.1 iPSC line. Toward these goals, we describe the use of our automated, high throughput CRISPR editing platform, ECLIPSE, for the rapid generation of knock-in iPSC models of ADRD. We leveraged our state-of-the-art knock-in methods and automated pipelines for the design, experimental optimization, and clonal isolation of 23 of the candidate target mutations in iPSCs. For each SNV target, at least 3 clonal homozygous and 6 clonal heterozygous mutation lines were generated for a total of 264 clonal cell lines over a 6-month period. The utilization of automated systems such as our ECLIPSE platform are critical catalysts for the rapid development of relevant cell models in large scale disease initiatives such as iNDI.

Genetic Therapies Posters - Wednesday

PB2038. Effect of respiratory resistance exercise done remotely on respiratory function in individuals with familial myopathy.

Authors:

M. Halseth, R. Mahoney, V. Kimonis; Div. of Genetics and Genomic Med., Dept. of Pediatrics, Univ. of California, Irvine, Irvine, CA

Abstract Body:

The goal of this project is to evaluate the effectiveness of remotely administered respiratory exercise training on lung and respiratory muscle function in patients with familial myopathy as quantified by measuring their maximum inspiratory pressures (MIP, primary outcome) throughout the study. We also measured spirometry, hand dynamometry, and functional rating scale surveys (IBMFRS, ALSFRS) biweekly. The six-minute walk test (6MWT) and timed up-and-go (TUG) measurements were taken monthly. Weeks 0-8 serve as baseline measurements. During weeks 8-40, the subjects performed inspiratory muscle strength training twice a day, 6 days a week starting at 50% inspiratory resistance and building up to 80% of their MIP. On review of the results, MIPs showed a significant average increase of 0.425 cmH₂O every week ($p=0.010$). Spirometry measures showed small non-significant decreases in the exercise phase, consistent with the natural decline of the disease. The ALSFRS survey showed a significant beneficial decrease of 0.042 points per week during the exercise phase ($p=0.001$). No other significant changes were noted, in contrast to the natural decline of this disease. A large benefit shown through the course of this study is the effectiveness of administering this experiment remotely, showing research studies and treatments can be largely accessible to individuals who cannot travel due to expenses, mobility, or life responsibilities. An overall decline would be expected as part of the natural history of the disease, but significant increases in the MIP demonstrate that respiratory resistance exercise is beneficial to individuals with myopathy.

Genetic Therapies Posters - Thursday

PB2039. Efficacy and safety of elamipretide in subjects with primary mitochondrial disease resulting from pathogenic nuclear DNA mutations (nPMD) is investigated through this phase 3 study design.

Authors:

M. Mancuso¹, A. Karaa², **A. Abbruscato**³; ¹Dept. of Clinical and Experimental Med., Neurological Inst., Univ. of Pisa, Pisa, Italy, ²Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, ³Stealth Biotherapeutics, Needham, MA

Abstract Body:

Introduction: Primary mitochondrial disease (PMD) is a group of genetic disorders with underlying disrupted energy metabolism due to impaired oxidative phosphorylation capacity. Although mitochondria are under the dual genetic control of nuclear DNA (nDNA) and mitochondrial DNA (mtDNA), 98% of the mitochondrial proteome is encoded by nDNA. Elamipretide is a novel investigational agent that is being developed for the treatment of a variety of mitochondrial diseases, including PMD. Elamipretide readily penetrates and transiently localizes to the inner mitochondrial membrane where it associates with cardiolipin to improve membrane stability, enhance ATP synthesis, and reduce the production of reactive oxygen species.

Objectives: SPIMD-301 (NCT05162768) is designed to evaluate the efficacy and tolerability of elamipretide in subjects with primary mitochondrial disease resulting from pathogenic nuclear DNA mutations (nPMD).

Methods: SPIMD-301 is a phase 3, randomized, double-blind, placebo-controlled clinical trial consisting of screening (≤ 28 days), treatment (48 weeks), and follow-up (4 weeks). Approximately 130 subjects aged ≥ 18 years and ≤ 70 years with nPMD causing their myopathy will be randomized (1:1) to receive a single daily subcutaneous dose of elamipretide or matching placebo for 48 weeks. The population will consist of 90 subjects who have nPMD associated with pathogenic mutations of the mitochondrial replisome ("replisome-related mutations") for primary analysis and an additional subset of up to 40 subjects who have nPMD associated with other non-replisome-related mutations. Subjects are required to have a diagnosis of nPMD with a predominant clinical manifestation of myopathy, which must include progressive external ophthalmoplegia (PEO) and exercise intolerance and/or skeletal muscle weakness. Subjects must also have genetic confirmation of either nDNA mutation of the mitochondrial replisome or other pathogenic mutations specific to nDNA. The primary efficacy endpoint will be distance walked (meters) on the 6MWT. Biomarkers and pharmacokinetic evaluations will also be explored.

Results: The study has been initiated, with the first site enrollment occurring in November 2021. Further opportunities for enrollment continue in the US and Europe. The estimated completion date is June 2024.

Conclusion: Results will help determine the efficacy and safety of mitochondrial-targeting therapy with elamipretide for the treatment of patients with nPMD.

Genetic Therapies Posters - Wednesday

PB2040. Evaluation of Ataluren efficacy in fibroblasts from Neurofibromatosis Type 1 patients with nonsense mutation

Authors:

S. Kim¹, H. Do², S. Heo¹, J. Kim¹, I. Choi³, H. Bae⁴, G-H. Kim⁵, B. Lee⁶; ¹Asan Med. Ctr., Seoul, Korea, Republic of, ²Asan life Sci. Inst., Seoul, Korea, Republic of, ³Asan Med. center, Seoul, Korea, Republic of, ⁴Med. Genetics Ctr., Asan Med. Ctr. Children's Hosp., Seoul, Korea, Republic of, ⁵Asan Med Ctr, Seoul, Seoul, Korea, Republic of, ⁶Asan Med Ctr, Seoul, Korea, Republic of

Abstract Body:

Neurofibromatosis Type 1 (NF1) is an autosomal dominant human genetic disorder caused by heterogenous mutations on the tumor suppressor gene *NF1*, affecting 1 in 3000 people. Since NF1 regulates cell growth and survival by inhibiting the accumulation of hyperactive RAS (GTP-bound RAS), disruption of NF1 induces cell growth-promoting signals such as MEK-ERK effector pathway. According to our previous data, about 30% patients have nonsense mutations on *NF1* introducing premature termination codon (PTC). Ataluren is a well-characterized small molecule which serves as a nonsense suppressor and has been approved by the European Medicines Agency for the treatment of Duchenne muscular dystrophy (DMD) that is basically produced by PTC in *DMD* gene. Here, we have isolated fibroblasts from 22 Korean patients with nonsense mutation NF1 and evaluated the efficacy of Ataluren treatment. We found that GTP-bound RAS activities were significantly reduced in 5 patients' fibroblasts (23%) after Ataluren treatment. Additionally, the levels of phosphorylated ERK, a downstream effector of the RAS-MEK pathway, decreased in the presence of Ataluren. In conclusion, our finding suggests that Ataluren can be an alternative drug for some patients with nonsense mutation NF1 instead of a standard treatment, MEK inhibitor Selumetinib.

Genetic Therapies Posters - Thursday

PB2041. Generation of eight human induced pluripotent stem cell lines from immortalized lymphoblastoid cells

Authors:

L. Sertori Finoti, C. Hou, L. Matsuoka, M. March, D. Li, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

In order to accurately study timed events that affect gene activation and repression and investigate the timing and dosage of disease genes related to neuron development, it is preferred to use neurons derived from human induced pluripotent stem cells (iPSCs). We have identified a neurodevelopmental syndrome with *de novo* novel missense variants in *U2AF2*, encoding a core component of U2 snRNP in the spliceosome. However, how these variants lead to neuronal differentiation defects is unknown. To investigate this, we utilized a developmental reprogramming strategy to capture a potential critical early neuronal phenotype. Human iPSCs from Epstein-Barr virus (EBV) immortalized lymphoblastoid cell lines (LCL) were generated from 3 healthy matched controls and 4 patients with c.448C>T mutation and 1 patient with a nonframeshift deletion (K329del). LCLs were reprogrammed using non-integrating Sendai viruses expressing Oct3/4, Sox2, cmyc, and Klf4. iPSCs were free of vectors carrying reprogramming genes at the 13th passage. They were positive for the SSEA4 and OCT4 pluripotency markers verified by immunostaining and expressed the LIN28A, NANOG, and OCT4 markers by qPCR analysis. Each line was assessed as negative for large genomic rearrangement using SNP genotyping array. The iPSC cells haven't shown any discrepancy in their morphology; however, we are expecting to see differences during their neuronal differentiation. The generation of these cell lines will allow us to investigate disease mechanisms and test possible therapeutic approaches.

Genetic Therapies Posters - Wednesday

PB2042. Genetic subgroup learnings is presented from the MMPOWER-3 trial: elamipretide improves six-minute walk test in individuals with mtDNA replisome disorders.

Authors:

A. Karaa¹, M. Mancuso², B. H. Cohen³, M. J. Falk⁴, A. Goldstein⁴, M. K. Koenig⁵, M. Hirano⁶, J. Vissing⁷, **A. Sullivan**⁸; ¹Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, ²Univ. of Pisa, Pisa, Italy, ³Akron Children's Hosp., Akron, OH, ⁴Mitochondrial Med. Frontier Program, Div. of Human Genetics, Dept. of Pediatrics, Children's Hosp. of Philadelphia and Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, ⁵Dept. of Pediatrics, Ctr. for the Treatment of Pediatric Neurodegenerative Disease, Univ. of Texas McGovern Med. Sch., Houston, TX, ⁶Neuromuscular Div., Columbia Univ Med Ctr, New York, NY, ⁷Copenhagen Neuromuscular Ctr., Rigshosp. et Univ. of Copenhagen, Copenhagen, Denmark, ⁸Stealth Biotherapeutics, Needham, MA

Abstract Body:

Introduction: Primary mitochondrial myopathy (PMM) is a group of genetic disorders of the mitochondria, adversely affecting predominantly skeletal muscle, leading to decline in quality of life. The mitochondria-targeting peptide elamipretide was recently assessed in a Phase 3, randomized, double-blind, placebo-controlled clinical trial for the treatment of subjects (N=218) with PMM (MMPOWER-3). Subjects had a variety of pathogenic variants in either nuclear (nDNA) or mitochondrial DNA (mtDNA) genes that caused their myopathy. The trial did not meet the primary endpoints in the highly heterogeneous study cohort. Here, we report results of genetic subgroup, post-hoc analyses subsequently performed on the per protocol population.

Methods: Using the MMPOWER-3 per protocol population, elamipretide effects on change from baseline in the six-minute walk test (6MWT) were examined as a function of gene variants.

Results: Individuals with primary mtDNA pathogenic mutations represented 74% of the trial population. Elamipretide showed no significant effects on 6MWT in these subjects when compared to placebo. Placebo-randomized subjects with the MT-TL1 pathogenic variant (~33% of mtDNA cohort) increased 42.4±13.4 meters (m) in 6MWT at week 24. After 24 weeks of elamipretide, subjects with nDNA gene mutations walked significantly farther than placebo counterparts (25.5±8.0 versus 0.3±7.7 m; p<0.05). The majority of the nDNA cohort was comprised of subjects with pathogenic variants in genes required for mtDNA genome maintenance (the mtDNA replisome, n=51 subjects). 6MWT at week 24 in subjects with replisome mutations changed 25.2±8.7 m with elamipretide versus 2.0±8.6 m in placebo (P<0.05), an effect even more prominent in individuals with progressive external ophthalmoplegia (PEO; n=32), with elamipretide-treated subjects walking 37.3±10.7 m versus -8.0±9.5 m in placebo at 24 weeks (P<0.05). Pharmacokinetic analyses in the nDNA cohort showed a positive correlation between plasma elamipretide concentration and 6MWT improvement.

Conclusions: The subgroup analysis demonstrated that 6MWT in the mtDNA cohort was confounded by a 'placebo effect', disproportionately driven by subjects with MT-TL1 pathogenic variants. The MMPOWER-3 basket trial design introduced significant heterogeneity and highlights the importance of considering genetic subgroups in developing future trials and treatments for individuals with PMM. The observation of improvement in 6MWT in the nDNA cohort with replisome disorders is encouraging, and future studies are planned to explore the efficacy of elamipretide in this population.

Genetic Therapies Posters - Thursday

PB2043. Genome-wide CRISPR/Cas9 Screening Reveals ATR as a Therapeutic Target That Overcomes Osteosarcoma Chemoresistance.

Authors:

S. Tang¹, P. H. Pandya², X. Wu¹, R. D. Roberts¹, K. E. Pollok², L. Li¹; ¹The Ohio State Univ., COLUMBUS, OH, ²Indiana Univ., Bloomington, IN

Abstract Body:

Osteosarcoma is the most common primary bone malignancy exhibiting remarkable histologic diversity and genetic heterogeneity. Despite extensive research efforts, the treatment and survival outcomes for osteosarcoma patients have stagnated over the past 30 years. The complex nature of osteosarcoma has confounded study of chemo-resistance mechanisms and identification of novel target. As an ideal new strategy to address the genetic complexity and drug-resistance in osteosarcoma, genome-wide CRISPR/Cas9 screen was conducted in one osteosarcoma cell line, SaOS2, parallel with RNA-sequencing in different conditions (control, Cisplatin, Doxorubicin), followed by gene essentiality score calculation. Cisplatin and Doxorubicin combination is among those commonly used as first-line therapies for osteosarcoma treatment. The objective of our studies is to bring forth new perspectives in target identification and chemo-resistance study for osteosarcoma. The gene targets and pathways that become more essential for osteosarcoma cell proliferation in the presence of chemo drugs were revealed. Drugs specific for some of these targets could then be added to the established regimens to overcome existing drug resistance. On the other hand, the genes that become less essential after treatment could provide additional chemo-resistant mechanisms. Multiple gene targets were identified in the screen. Among these, we've successfully identified and validated Berzosertib, an ataxia telangiectasia and Rad3-related protein kinase inhibitor (ATRi), as a synergistic lethality partner to Cisplatin in vitro for osteosarcoma treatment. Efficacy and safety measurement of Berzosertib and Cisplatin combination therapy in a cell-line xenograft mouse model and in several patient-derived xenograft models is in progress and expected to be finished before September.

Genetic Therapies Posters - Wednesday

PB2044. Genotyping analysis in patients with retinitis pigmentosa due to mutations in the *RHO*, *PDE6A*, or *PDE6B* gene: the PHENOROD2 study.

Authors:

D. Chung¹, L. Blouin², A. Celle², A. Le Meur², S. Mohand-Saïd³, J-A. Sahel⁴, C. Zeitz⁵, I. Audo⁶; ¹SparingVision, Philadelphia, PA, ²SparingVision, Paris, France, ³Inst. de la Vision, Paris, France, ⁴Dept. of Ophthalmology, Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, ⁵Sorbonne Université, INSERM, CNRS, Inst. de la Vision, Paris, France, ⁶UMR S 968 INSERM, Paris, France

Abstract Body:

Retinitis pigmentosa (RP) is a rare rod-cone dystrophy (RCD) in which rod photoreceptors progressively degenerate, subsequently triggering the degeneration of cone photoreceptors. Clinical symptoms start with night blindness, followed by slow constriction of the visual field causing tunnel vision and eventually legal blindness. With over 70 different genes implicated, RCDs form a heterogeneous group of poorly characterized inherited retinal disorders.

PHENOROD2 is an ongoing prospective, multicentric, natural history study documenting functional and structural changes in patients with RP due to mutations in *RHO*, *PDE6A*, or *PDE6B*. An interim analysis was performed in the first 44 patients enrolled at the national reference center for rare diseases (REFERET) of the *Centre Hospitalier National d'Ophthalmologie des Quinze-Vingts* (Paris, France).

Genotyping data were collected from the medical records of patients enrolled in PHENOROD2.

Out of 44 patients, 33 (75%) carried a monoallelic variant in the *RHO* gene, and 11 (25%) carried a biallelic variant in either the *PDE6A* gene (n=7, 16%) or the *PDE6B* gene (n=4, 9%). Segregation analysis was performed in all but one patient with autosomal recessive RP

(i.e., *PDE6A* and *PDE6B* patients). A total of 14 unique variants were identified in the 33 *RHO* patients, 9 unique variants in the 7 *PDE6A* patients, and 7 unique variants in the 4 *PDE6B* patients. The majority of disease variants were missense mutations: 94% in *RHO* patients, 50% in *PDE6A* patients, and 38% in *PDE6B* patients. The most common missense variants were p.(Pro347Leu) and p.(Arg135Trp) each found in 5 (15%) *RHO* patients, and p.(Gln569Lys) found in 3 (21%) *PDE6A* carriers. The only patient with homozygous variants carried biallelic missense variants p.(Arg100His) in the *PDE6B* gene.

Further assessment is ongoing on the full cohort of PHENOROD2 that included a total of 82 patients. The prospective analysis of disease progression in these RP patients could allow for genotype-phenotype correlations.

Genetic Therapies Posters - Thursday

PB2045. Harmonization of phenotype ontologies and identification of novel gene/drug-disease associations with Natural Language Processing

Authors:

I. Melas, L. Middleton, M. Tadeja, A. O'Neill, A. Harper, D. Vitsios; AstraZeneca, Cambridge, United Kingdom

Abstract Body:

Integration of data from multiple resources and ontologies around genotypes and phenotypes is a key challenge in life sciences aiming to bridge the gap between basic and clinical research. This integration is obstructed by the use of different nomenclatures in different consortia, initiatives and healthcare systems. Projects such as the Monarch Initiative have achieved the streamlined curation for a number of such ontologies, employing both computational methods and manual curation. Nevertheless, mapping diseases and phenotypes across ontologies remains a daunting task.

Herein, we leverage several ad-hoc frameworks, such as word embeddings and the Monarch dataset, that address the mapping of terms across ontologies and augment on these to provide a universal mapping framework. We use the BioWordVec word2vec embeddings, pre-trained on the entire PubMed corpus, to construct vector representations of free text terms. BioWordVec captures the meaning behind these terms and embeds them in a latent (250-dimensional) space, so that terms with similar semantic context are mapped closer together in the latent space. To improve these semantic mappings, we use Elasticsearch to look up the PubMed corpus for each query term and retrieve extracts around them from all relevant papers. This allows us to construct augmented embedding representations for each query term (e.g. a phenotype or disease) that better capture their biological context.

Moreover, we leverage the curated mappings from Monarch to train a supervised learning model capable of identifying novel associations between phenotypes. This framework combines the power of BioWordVec embeddings with expertly curated mappings from Monarch to infer the biological semantic similarity in any pair of disease/ontology terms. A Deep Neural Net trained to distinguish between true and false (random) associations achieved an AUC of 0.79 (5-fold Cross-Validation). Remarkably, the trained classifier is powerful enough to identify associations spanning across different biological layers, i.e. disease to gene (AUC = 0.75) and disease to drug (ROC AUC = 0.79) associations (OpenTargets derived associations were used as a training dataset in this case). Additionally, we trained a universal classifier on all association types together (i.e. disease to disease/gene/drug associations) achieving a comparable performance with the specialized models, albeit slightly lower (AUCs = 0.76, 0.74, 0.76 respectively). Finally, we accompany our work with a web resource providing both unsupervised and supervised learnt mappings between various disease ontologies, genes and/or drugs.

Genetic Therapies Posters - Wednesday

PB2046. Harnessing the power of natural variations of the human genome to support the discovery of new therapeutics: A systematic review

Authors:

K. Trajanoska^{1,2}, C. Bhérier^{1,2}, D. Taliun^{1,2}, B. Richards^{1,3}, V. Mooser^{1,2}; ¹McGill Univ., Montreal, QC, Canada, ²Canada Excellence Res. Chair in Genomic Med., McGill Univ., Montréal, QC, Canada, ³Lady Davis Inst., Jewish Gen. Hosp., McGill Univ., Montréal, QC, Canada

Abstract Body:

Rationale: The massive investments made in human genetics and genomics over the past decades were anticipated to result in many innovative and impactful therapies or new indications for existing drugs. Has this expectation been met for non-cancer drugs? Twenty years after publication of the first human genome, we considered this question legitimate and the answer critical to guide future investments in the field. **Methods:** We defined as “genetically-driven” those targets for which human germline genetic associations were judged to be sufficiently informative to support a new drug discovery campaign, resulting in regulatory approval of the therapeutic (or new indication). We restricted our analysis to FDA- or EMA-approved therapeutics and excluded cancer drugs, hormonal therapies, vaccines, and antimicrobials. We then applied this definition to a search in PubMed which we complemented with information from the Orphan Drug Designations and Approvals database. All articles were screened first at the title/abstract level and at a full-text stage. **Results:** We identified 31 drug targets that met our definition of being genetically-driven. This corresponds to ~6% of the ~500 human targets of FDA-approved non-cancer drugs. The 31 underlying germline genetic associations led to the development and approval of new therapies for rare (n=27), common (n=3), and both rare and common (n=1, i.e., PCSK9 inhibitors) disorders. Most drugs indicated for rare diseases compensate for disease-causing loss-of-function or misfolding variants in the target gene whereas, in contrast, the 4 genetically-driven drugs for common conditions mimic the disease-protective effects of rare loss-of-function variants found in diverse populations. Most (18/31) genetically-driven targets are in the metabolic area. Only five genetically-driven drugs so far are small molecules, whereas the remaining drugs have onerous cost of goods. In addition, we identified 7 marketed drugs approved (n = 4) or under development (n = 3) for genetically-driven repurposed indication. **Conclusion:** Our systematic review confirms that drug discovery is starting to harvest the fruits of genetics. A wealth of new therapeutics are anticipated to be tested in humans within the next few years. For genomics to be successfully embedded in drug development will require access to a new generation of large, diverse, disease-based cohorts of properly consented, well-phenotyped participants.

Genetic Therapies Posters - Thursday

PB2047. Implementing a Genetics-First Strategy in Clinical Development for Rare Brain Diseases

Authors:

A. Mitchell, C. Paisan-Ruiz, C. Gubbels, C. Whelan, V. Nguyen, T. Cochrane, L. Gedney, S. Fradette, H. Runz; Biogen, Cambridge, MA

Abstract Body:

Biogen's clinical pipeline programs span seven broad therapeutic areas, working across the human lifespan, from neonates to the elderly. Biogen's programs are increasingly focused on rare diseases and genetically-defined subsets of common diseases. These types of programs require extra support from Clinical Genetics, particularly in the development of patient identification strategies.

The strategies for trial site selection and patient recruitment that have previously worked well are insufficient for ultra-rare diseases, for genetically defined subsets of common disease, and for pre-symptomatic mutation carriers of these diseases.

We approach patient recruitment for each specialized trial from many different angles. Depending on availability, we may leverage sponsored genetic testing panels, networks of genetic counselors and other medical providers, patient databases and outreach programs through clinical genetic testing providers, large scale population and disease-specific genotyping initiatives, academic collaborations, and cohorts that may recall participants by genotype. We may do outreach through patient advocacy groups or sponsor a new patient advocacy group, if none exists. We are increasingly turning to natural history studies for trial-ready cohorts, particularly for indications for which there is no current treatment. Here, we highlight three examples of patient identification strategy for three, very different Biogen programs: an ultra-rare indication, a genetically-defined subset of a common, complex disease, and a study of pre-symptomatic mutation carriers.

As the number of clinical programs with genetic inclusion criteria continues to increase, patient identification challenges will continue to grow. Continued success will require multiple, simultaneous efforts, extensive preparation, and - most importantly - support from every level and every department in the company.

Genetic Therapies Posters - Wednesday

PB2048. Interrogation of the role of androgens and their antagonists in the pathogenesis and treatment of Vascular Ehlers-Danlos Syndrome.

Authors:

E. Juzwiak^{1,2}, C. Bowen^{1,2}, A. Zeng^{1,2}, H. Dietz^{1,2}; ¹Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ²Howard Hughes Med. Inst., Chevy Chase, MD

Abstract Body:

There are currently no therapeutic strategies to prevent spontaneous aortic rupture and premature death in patients with Vascular Ehlers-Danlos Syndrome. Male patients with VEDS are at an increased risk for catastrophic arterial rupture during puberty compared to females. Our VEDS mouse model (Col3a1^{G938D/+}) recapitulates this pubertal sexual dimorphism, with 44% vs. 69% survival of males vs. females at 60 days of age, respectively. We hypothesized that pathogenesis of vascular rupture in VEDS involved modulation of the PKC/ERK signaling axis by androgens. Androgens are hormones that increase during puberty and are higher in males during puberty. They can bind to the androgen receptor (AR) and influence both gene expression and molecular signaling. We therefore had incentive to understand the potential mechanisms by which androgen receptor antagonists might contribute to protection from arterial events. We treated Col3a1^{G938D/+} mice with the selective AR antagonist (ARa) bicalutamide, the dual ARa and mineralocorticoid receptor antagonist (MRa) spironolactone, or the selective MRa eplerenone from the time of weaning until 60 days of age. To ensure these small molecules are acting through the AR to modulate VEDS pathogenesis, we also crossed an AR conditional (floxed) allele to Col3a1^{G938D/+} mice that globally express Cre recombinase. We observed that male Col3a1^{G938D/+} AR null mice have improved survival compared to untreated Col3a1^{G938D/+} mice at 60 days (68% vs. 44%), a protective performance mimicked by treatment with bicalutamide (65% survival). Immunoblot analysis of descending aortic lysates shows that AR blockade, either chemical or genetic, normalizes activation of the PKC/ERK axis. Taken together, these data suggest that AR signaling contributes to vascular disease in Col3a1^{G938D/+} mice. Notably, male Col3a1^{G938D/+} mice treated with the dual ARa/MRa spironolactone showed the best performance (87% survival at 60 days). Use of the selective MRa eplerenone afforded intermediate protection, approaching that of spironolactone (81% male survival at 60 days) in selected contexts. Identical trends were observed in female Col3a1^{G938D/+} mice. These data document that both AR and MR signaling are determinants of outcome in Col3a1^{G938D/+} mice and highlight the therapeutic potential of isolated MR antagonism that should maintain normal sexual development.

Genetic Therapies Posters - Thursday

PB2049. Learnings from a randomized clinical trial with ARID1B patients

Authors:

P. J. van der Sluijs¹, K. Safai Pour², C. Berends², R. G. J. Zuiker², G. W. E. Santen¹; ¹Leiden Univ. Med. Ctr. (LUMC), Leiden, Netherlands, ²Ctr. for Human Drug Res. (CHDR), Leiden, Netherlands

Abstract Body:

Background: Few clinical trials target patients with intellectual disability (ID). However, this is likely to change due to accelerating gene discovery and the increasing knowledge about underlying mechanisms. We initiated a randomized, double-blind, placebo-controlled, two-way crossover study evaluating the effect of clonazepam in patients with an ARID1B-related disability. Here we report our experiences and describe potential challenges while conducting a clinical trial with participants with an intellectual disability or developmental delay. **Results:** We approached 47 patients: 23 of these were eligible and willing to participate (figure 1). Eventually, 16 patients were randomized over two study treatment periods. Initially, we planned to conduct the whole study in a single phase in 2021. Since the inclusion rate initially lagged, we allowed other study groups for a second phase in early 2022. Finally, 10 patients were included in the first phase and 6 in the second. Two patients initially declined but eventually participated in the second phase because we could report limited side effects. One participant dropped out on day 4 of the first treatment period because of increased anxiety possible related to clonazepam. **Conclusion:** Even though we had many potential participants, we struggled to achieve sufficient enrolment. The leading cause was that parents were afraid of side effects. Interestingly, our drop-out rate during the study was low, indicating that our patients were highly motivated. We recommend involving parents in the study design and adding multiple moments of inclusion so that experiences of the first testing period, for example, side effects, can be used to inform parents in subsequent testing periods.

Genetic Therapies Posters - Wednesday

PB2050. Leveraging Machine Learning to Improve Guide RNA Design for Efficient and Selective ADAR-Mediated RNA Editing

Authors:

Y. Jiang, B. Banjanin, L. Guo, L. Bagepalli, K. Rupp, Y. Savva, B. Booth, A. Briggs, R. Hause; Shape Therapeutics, Seattle, WA

Abstract Body:

Therapeutic RNA editing by redirecting natural ADAR enzymes offers promise as a safe method of gene therapy without risk of DNA damage or dependency on the delivery of non-human proteins. However, ADAR enzymes are inherently promiscuous, and deterministic rules for how different guide RNA (gRNA) sequences recruit ADAR and influence editing efficiency and specificity remain poorly understood. We demonstrate a successful application of machine learning coupled with a novel high throughput screening (HTS) and validation platform to identify deterministic gRNA features that improve ADAR-mediated RNA editing efficiency and specificity. Our approach explores the enormous gRNA design space to propose highly efficient and specific gRNAs that validate experimentally. We further propose novel machine learning approaches to expand modeling gRNA performances for additional targets.

The RNAfix® platform generates gRNAs that direct natural ADAR enzymes to therapeutically-relevant sites in the transcriptome to correct G->A mutations, control splicing, or modulate protein expression and function. To empirically determine the rules affecting on-target ADAR-mediated RNA editing, we constructed a HTS platform that assesses up to 500,000 structurally unique gRNAs against any clinically relevant target sequence. We used convolutional neural networks (CNN) to model gRNA performances using primary gRNA sequences as inputs and achieved high predictive accuracy for ADAR1 and ADAR2 editing. Input optimization successfully generated novel gRNA designs that outperformed any gRNA from the original HTS and exhibited primary and secondary sequence diversity that expanded vastly from the original HTS screen.

We have successfully established a robust pipeline for integrating supervised learning into our HTS design for each target. However, the ultimate goal of de novo, in silico design requires identifying generalizable rules that predict gRNA editing outcomes across multiple targets. To this end, we derived secondary structural features across gRNAs to model gRNA editing performance using gradient boosted decision trees that identify important structural features to prioritize for future HTS. We also extended CNN models towards better generalizability by fine tuning several novel transformer-based architectures that incorporate global dependencies of RNA sequence and secondary structure space across multiple candidate therapeutic targets. These developments rapidly shorten gRNA discovery timelines through in silico gRNA design for any number of common or rare genetic diseases.

Genetic Therapies Posters - Thursday

PB2051. Localization of KK13 peptide and their potential against ALS.

Authors:

A. Bhuiyan¹, S. Asakawa¹, N. Bailey Kobayashi², T. Yoshida²; ¹The Univ. of Tokyo, Tokyo, Japan, ²Inst. for Advanced Sci., Toagosei Co., Ltd., Tsukuba 300-2611, Japan

Abstract Body:

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are a kind of neurodegenerative disease that is caused by expansion of the hexanucleotide repeat (GGGGCC) in the non-coding region of the C9orf72 gene. Three possible disease mechanisms have been hypothesized - DNA level: reduced levels of C9orf72 transcription (haploinsufficiency), RNA level: formation of toxic RNA foci and sequestration of RNA-binding proteins (RNA toxicity), and protein level: accumulation of the dipeptide repeat protein (DRP) generated by RAN (repeat-associated non-ATG) translation. Among the six DPR proteins, highly basic poly-PR (proline-arginine) was reported to show cytotoxicity and localization at the nucleolus. A novel polypeptide, KK13 was also shown to be located at the nucleolus, showing the possibility of antagonizing poly-PR was demonstrated. To examine the potential of the KK13 peptide as a therapeutic agent, a vector encoding 100 poly-PR repeats under dsRed was constructed. The poly-PR gene expression and their nuclear localization were observed. The KK13 peptide as a potential material against poly-PR dipeptide will be examined.

Genetic Therapies Posters - Wednesday

PB2052*. Long-term efficacy and safety of elamipretide in patients with Barth syndrome is presented through the 192-week open-label extension results of TAZPOWER.

Authors:

H. Vernon¹, W. R. Thompson², R. Manuel¹, A. Abbruscato³, J. Carr³, B. Hornby⁴, **J. Campbell⁵**; ¹Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ²Dept. of Pediatric Cardiology, Taussig Heart Ctr., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ³Stealth BioTherapeutics, Inc., Needham, MA, ⁴Dept. of Physical Therapy, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ⁵Stealth Biotherapeutics, Needham, MA

Abstract Body:

Introduction: Barth Syndrome (BTHS) is a rare mitochondrial disorder characterized by cardiomyopathy, skeletal myopathy, neutropenia, growth abnormalities, and shortened lifespan. BTHS is caused by defects in TAZ, the gene encoding for tafazzin, which is an acyltransferase involved in the final remodeling step to mature cardiolipin, an essential component of mitochondrial function. Elamipretide readily penetrates and transiently localizes to the inner mitochondrial membrane where it associates with cardiolipin to improve membrane stability, enhance ATP synthesis, and reduce the production of reactive oxygen species. This study evaluated the long-term efficacy and safety of elamipretide during the open-label extension (OLE) of the TAZPOWER trial conducted in patients with BTHS.

Methods: TAZPOWER was a 28-week randomized, double-blind, placebo-controlled crossover trial with a 168-week OLE. Patients aged ≥ 12 years with genetically confirmed BTHS received elamipretide 40mg subcutaneously daily or matching placebo for 12 weeks with a 4-week washout between treatments. Patients who entered the OLE continued on long-term elamipretide. The primary outcome measures were the 6-minute walk test (6MWT) and Total Fatigue score on the BarTH Syndrome Symptom Assessment (BTHS-SA). Secondary outcomes included muscle strength by hand-held dynamometry (HHD), Five Times Sit to Stand (5XSST), and SWAY Application Balance assessment.

Results: Ten patients entered the OLE with 8 reaching the week 168 visit and 3 of the 8 patients completed the week 192 visit. Significant improvements on the 6MWT were seen at all timepoints throughout the OLE with cumulative 96.1 meters (w168, $p=0.003$) and 122.7 meters (w192, $p=0.009$) of improvement. Mean BTHS-SA Total Fatigue scores were decreased from baseline at the start of the OLE and maintained below baseline at all OLE timepoints through week 192. An overall increase in the mean muscle strength by HHD from baseline was noted for all subjects over time. The mean changes from baseline were statistically significant at all OLE visits ($p < 0.05$). There was overall decrease in the mean 5XSST time observed for all subjects over time. An overall increase in the mean SWAY balance score was observed for all subjects over time, with statistical significance ($p < 0.05$) at visit weeks 24, 48, 72 and 168.

Conclusion: Long-term therapy with elamipretide was associated with improvements in functional assessments. These results suggest that elamipretide may improve the natural history of BTHS which is usually associated with a progressively deteriorating physical function parameters.

Genetic Therapies Posters - Thursday

PB2053. New class of anaplerotic compounds ameliorates substantial loss of lysine succinylation in propionyl-CoA carboxylase deficient cells

Authors:

A-W. Mohsen¹, A. Karunanidh¹, B. Seminotti¹, C. Van't Land¹, J. Vockley²; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²UPMC Children's Hosp. of Pittsburgh and Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA

Abstract Body:

Background: Propionic acidemia (PA) is an autosomal recessive metabolic disorder caused by deficiency of propionyl-CoA carboxylase (PCC), a mitochondrial enzyme that catalyzes conversion of propionyl-CoA to *S*-methylmalonyl-CoA. The current treatment approach is to control metabolic episodes of acidosis, hypoglycemia, and hyperammonemia. Propionyl-CoA is a common catabolic product of Val, Ile, Met, Thr, and odd-chain and branched-chain fatty acids. While excess propionyl-CoA inhibits *N*-acetylglutamate synthetase causing hyperammonemia, its thiolysis introduces toxic levels of propionic acid and its derivatives and depletion of its downstream pathway product, succinyl-CoA. Succinyl-CoA's depletion disrupts the structure/function of 1200 mitochondrial proteins modulated by Lys succinylation/desuccinylation including various Krebs cycle enzymes. We are developing a new class of compounds that address the underlying cause of PA namely the depletion of succinyl-CoA and accumulation of toxic propionic acid and its derivatives. **Methods:** Histological sections from liver of a PA patient were stained with hematoxylin and eosin to assess cell morphology. Immunofluorescence (IF) staining with antisuccinyllysine and antipropionyllysine antibodies was used to examine Lys succinylation and propionylation. Cultured HEK293 *PCCA*^{-/-} cells were treated with compound PMA010 for 72 hours and Lys succinylation and propionylation was evaluated by IF staining. Propionic acid derivatives were measured in cell culture media using tandem mass spectroscopy. **Results:** Concurrent with changes in hepatocytes' morphology, PA liver tissue IF staining showed substantial loss of Lys succinylation and gain in Lys propionylation compared to control tissue. In HEK293 *PCCA*^{-/-} cells, the loss of antisuccinyllysine staining intensity was corrected with the PMA010 treatment to levels similar to those observed in the parent HEK293 cells. Propionylcarnitine, propionylglycine, and hydroxypropionylglycine measured in the HEK293 *PCCA*^{-/-} cell culture media containing the PMA010 were decreased by 59, 40, and 80%, respectively, compared to their levels in media of cells without the compound. **Conclusions:** Results confirm the substantial decrease in Lys succinylation in patient liver and in HEK293 *PCCA*^{-/-} cells similar to what we previously reported in patient fibroblasts. Restoring Lys succinylation to normal in HEK293 *PCCA*^{-/-} cells and the decrease in propionic acid derivatives in the media identify PMA010 as a promising compound to further optimize its effectiveness in treating PA.

Genetic Therapies Posters - Wednesday

PB2054*. Novel ASO based treatment shows promise in treating CMT2S, preclinical data

Authors:

B. Przychodzen, S. Smieszek; Vanda Pharmaceuticals, Washington, DC

Abstract Body:

Charcot Marie Tooth type 2S (OMIM 616155), is an autosomal recessive type of CMT, an inherited peripheral neuropathy. Rare variants in immunoglobulin mu-binding protein 2 (IGHMBP2) have been shown to cause CMT Type II. These have been shown to result in abnormal RNA processing which likely leads to alpha-motor neuron degeneration. CMT2S is primarily characterized by distal muscle atrophy, weakness with areflexia, and relatively minor sensory involvement. Cassini et al., 2019 reported a patient whose suspected diagnosis was CMT. Whole-genome sequencing has revealed paternally inherited cryptic splice site variant (inherited non-coding variant (c.1235 + 894 C>A) deep in intron 8). RT-PCR analysis was consistent with activation of the cryptic splice site. Bidirectional sequencing of cDNA derived from the patient's cells identified both ends of the splicing alteration. The abnormal transcript was shown to undergo nonsense-mediated decay (NMD), resulting in haploinsufficiency. The objective of this study was to target IGHMBP2 - specifically cryptic splice site variant with a novel ASO designed to avoid NMD. We have obtained patient's fibroblast cell line and confirmed the variant with WGS, furthermore, we also confirmed NMD induced by cryptic splicing. We designed an ASO targeting specifically deep in intron 8 (c.1235 + 894 C>A). The ASO was 19mer targeting the sequence around CACTTCCAC(A)GGGGGAAGA. Analysis of the area of interest led to the design of several ASOs. These were further screened and prioritized based on optimal in silico binding affinity. CMT Type II patient-derived fibroblast cells underwent ASO treatment (10uM) and incubation. Successful cellular entry of the ASO was confirmed with flow cytometry (fluorescein labeled ASO (GFP+ 99.8%)). Upon treatment with ASO we have observed a significant increase (~30% increase) of the IGHMBP2 protein levels in the oligo-treated samples as compared to control (WB antibody Sigma SAB2106426). Additionally, qPCR results confirmed an increased ratio of restored WT transcript to cryptic exon-containing transcript (~1.3-fold). Furthermore, the ASO was shown to have limited off-target effects in silico. While the improved clinical formulation of the tested ASO may be necessary, the current preclinical data are supportive of a potentially efficacious treatment increasing the IGHMBP2 protein levels. The N-of-1 precision medicine approach may prove instrumental to the design of treatments for this highly diverse genetic disorder. This case exemplifies the shifting boundaries with rapid WGS-based clinical diagnoses coupled with the rapid design of personalized ASO-based therapeutics.

Genetic Therapies Posters - Thursday

PB2055. Novel siRNA-based therapeutic approach for Megacystis Microcolon Intestinal Hypoperistalsis Syndrome (MMIHS).

Authors:

I. Baldini, R. Ottalevi, P. Ebadi, N. Sabatini, A. Infascelli, G. Valoti, P. Graziani, M. Masciovecchio, C. Ciace, D. Clementoni, A. Lo Iacono, G. Iodice, M. Luciano, J. Caceres, R. Paone; Dante Labs, L'Aquila, Italy

Abstract Body:

Rationale MMIHS is a rare and severe form of functional intestinal obstruction in the newborn characterized by megacystis, microcolon, and intestinal dysmotility. A classic feature of MMIHS is represented by *ACTG2* mutations that can be both inherited in an autosomal dominant manner or occur as *de novo* events during oogenesis, spermatogenesis, or early embryonic development. *ACTG2* encodes γ -2 enteric actin and its mutations disrupt actin polymerization leading to impaired contraction of intestinal smooth muscle cells. Even though survival has improved in recent years, MMIHS is a condition that requires invasive palliative treatments. To meet the need for therapeutic approaches targeting the underlying cause of the disease, we designed a group of therapeutic siRNA to specifically target the *ACTG2* mutant allele, with no effect on the wild type (WT). **Methods** Whole genome sequencing (WGS) was performed to identify *ACTG2* mutations. DNA was isolated from chorionic tissue of a 13th week pregnant patient presenting an echography with alteration of gut and bladder of the newborn and sequenced on Illumina Novaseq6000. Variant analysis was performed using Variant Call Files (VCFs) computed with DRAGEN (Illumina) on the human reference genome GRCh37 and the pathogenic variant was identified with EXTENSA™ software. A Machine Learning (ML) software, fed with siRNA silencing information retrieved from experimental data, has been used as prediction tool for the identification of siRNAs with the most discriminatory power between *ACTG2*^{mutant} and *ACTG2*^{WT} mRNA. *In vitro* preclinical studies have been performed on HEK293 cells to evaluate the specific silencing of the *ACTG2*^{mutant} mRNA and to test the reduction of its expression in an efficient and highly specific manner. **Results** Upon identification of the pathogenic *ACTG2* variant, NM_001615.3:c.532C>T (NP_001606.1:p.Arg178Cys), we have predicted 8 siRNAs specific and effective for R178C-*ACTG2* mutation. Preliminary data suggest that our selected siRNA specifically target the *ACTG2*^{mutant} mRNA and reduce the expression of mutant γ -2 enteric actin while showing no effect on *ACTG2*^{WT}, thus enabling a condition of haplosufficiency. **Conclusions** Our study proposes a therapeutic approach based on siRNA as a novel treatment for MMIHS, identified siRNA sequences as good candidates for the development of a new drug and underscore a translational impact for future strategy to cure this disease. Further studies are necessary to verify that the observed reduced expression of mutant γ -2 actin prevents disease progression and to urgently translate these results into benefits for patients.

Genetic Therapies Posters - Wednesday

PB2056*. Personalized RNA interference approach as a nanotherapy for Crouzon syndrome: design of allele-specific siRNAs targeting FGFR2 mutant allele delivered by recombinant human nanoferritin.

Authors:

F. Tiberio¹, M. Salvati¹, L. Di Pietro¹, G. Tamburrini², P. Ceci³, E. Falvo³, L. Massimi², G. Tisci⁴, O. Parolini¹, A. Arcovito¹, W. Lattanzi¹; ¹Università Cattolica del Sacro Cuore, Rome, Italy, ²Fondazione Policlinico Univ.rio Agostino Gemelli, Rome, Italy, ³Consiglio Nazionale delle Ricerche, Rome, Italy, ⁴Università La Sapienza, Rome, Italy

Abstract Body:

Objective: Crouzon syndrome (CS) is a rare genetic disorder presenting with complex craniofacial malformations mainly due to the premature fusion of skull sutures. CS is caused by missense heterozygous gain-of-function mutations in the *Fibroblast Growth Factor Receptor 2 (FGFR2)* gene, which lead to constitutive activation of the receptor and of its downstream cascade. Currently the treatment is based on multiple surgeries to release the skull constraint that impairs brain growth. Post-surgical relapses occur in up to 40% cases requiring further operations associated with severe morbidities. The aim of this project is to develop a noninvasive therapy for CS using allele-specific siRNAs targeting the mutant *FGFR2* allele, delivered by recombinant human ferritin (HFt) as a highly biocompatible nanocarrier. This strategy is intended to restore FGFR2 signaling, thus hampering the aberrant FGFR2-induced osteogenesis and the subsequent premature suture fusion. **Methods:** Calvarial mesenchymal stromal cells (CMSCs) were isolated from calvarial tissue collected as surgical waste from *FGFR2* mutation-positive CS patients. A set of siRNA specifically targeting the mutant *FGFR2*-alleles of each patient enrolled was designed and tested in patient cells. The efficiency of each siRNA was evaluated through real time PCR. The expression of the receptor of ferritin *CD71* was investigated using real time PCR. CMSCs were incubated with fluorescein isothiocyanate-labeled HFt and the cellular uptake of the construct was analysed using fluorescence microscopy. Lastly, siRNAs have been encapsulated within ferritin cavity performing pH-dependent procedure and HFt-siRNAs complexes were tested in CMSCs using Incucyte Live-cells analysis system. **Results:** Gene expression analysis allowed identifying specific siRNAs with differential knockdown efficiency between wild-type and mutant *FGFR2* alleles in CMSCs derived from CS patients. In addition, expression analysis revealed that *CD71* was up-regulated during osteogenic differentiation of CMSCs. Fluorescence microscopy showed an efficient internalization of HFt within CMSCs cytoplasm. Preliminary results with HFt-based nanoparticles showed so far that, though siRNAs can be encapsulated in HFt, their intracellular trafficking is poorly detectable. **Conclusions:** Our data suggested that allele-specific *FGFR2* knockdown by siRNA represents a desirable strategy to silence *FGFR2* mutant allele in CS patients. We also demonstrated that HFt is suitable for delivery in CMSC. Future studies aim to optimize the intracellular trafficking and siRNAs release upon HFt internalization.

Genetic Therapies Posters - Thursday

PB2057. Pre-clinical antisense oligonucleotide treatment of CMT2E in a human induced pluripotent stem cell (iPSC)-derived motor neuron model.

Authors:

J. Medina¹, A. Rebelo², C. Yanick³, E. Jacobs⁴, S. Zuchner², M. Saporta⁵; ¹Univ. of Miami Miller Sch. of Med., MIAMI, FL, ²Univ Miami, Miami, FL, ³Univ. of Miami Miller Sch. of Med., Miami, FL, ⁴Univ. of Miami, Maimi, FL, ⁵Univ. of Miami, Miami, FL

Abstract Body:

Charcot-Marie-Tooth disease type 2E is caused by mutations in *NEFL* encoding neurofilament light chain (NFL), an integral component of the axonal cytoskeleton. Previous studies have established intracellular toxic NFL-positive aggregates in motor neurons. Increased levels of plasma and CSF NFL have been described as biomarkers for CMT2E. Our study aims to develop a genetic treatment strategy to prevent NFL aggregate formation and decrease NFL supernatant, a hallmark of axonal degeneration. Genetic treatments in dominant diseases have thus far proven challenging. Seven antisense oligonucleotides (ASOs) were developed spanning the allele-specific targetable region surrounding a missense mutation (p.N98S) in the *NEFL* gene. ASOs were designed with 2'-O-methyl modifications flanking 4 base pairs on each end with phosphorothioate modifications between all 18 base pairs to enable RNase H binding and clearance of the p.N98S mRNA transcripts. Stable transduction of Lenti-X 293T cell lines were produced with lentiviral vectors encoding wild type and mutant *NEFL* to assess the ASO allele-specific knockdown ability of all seven ASOs. ASO treatment of transduced Lenti-X 293T cells led to an allele specific knockdown of NFL in a human cell model confirmed via western blot at the protein level. We are currently investigating ASO efficacy in iPSC-derived 2D motor neurons and 3D neurospheres from CMT2E individuals and controls. To further assess allele specificity and potential off-target effects, we will be conducting allele-specific RT-PCR using iPSC-derived 2D motor neurons. To explore reversal of pathomechanistic changes observed, we will also be conducting post-treatment co-immunoprecipitation of NFL and known binding partners required for endogenous NFL-dynamics. Our results suggest a viable genetic targeting strategy for a therapeutic application in this autosomal dominant disease. Optimizing this approach in patient derived pre-clinical iPSC studies is a key strategy under FDA rules to design early human trials.

Genetic Therapies Posters - Wednesday

PB2058*. RNA editing of founder nonsense mutations causing inherited retinal diseases using site-directed endogenous adenosine deaminase acting on RNA.

Authors:

D. Sharon¹, N. Schneider¹, R. Steinberg², J. Valensi¹, A. Eylon², E. Banin¹, E. Levanon², S. Ben-Aroya²; ¹Hadassah Hebrew Univ Med. Ctr., Jerusalem, Israel, ²Faculty of Life Sci., Bar-Ilan Univ., Ramat Gan, Israel

Abstract Body:

Purpose: Targeted RNA editing utilizing the ubiquitous human adenosine deaminase acting on RNA (ADAR) enzyme is a possible new genetic therapeutic approach for the treatment of inherited retinal diseases (IRDs). Utilizing guideRNA (gRNA) to recruit the endogenously expressed ADAR enzyme to a mutated RNA and facilitating the deamination of a specific adenosine to inosine (read as a guanine by the ribosome), allows for the correction of mRNA transcripts in a transient and tunable manner. According to our recent analyses, 40% of single nucleotide variants (SNVs)-causing IRDs are candidates for ADAR-directed editing. Our aim is to design and test gRNAs that induce targeted ADAR editing for 3 common Israeli mutations causing IRDs: *TRPM1*- p.K294*, *FAM161A*- p.R523*, and *KIZ*- p.R76*. **Methods:** After determining Israeli IRD candidate mutations, we used a yeast model to identify candidate gRNAs for these mutations by measuring yeast survival and percent editing in next generation sequencing (NGS). Effective gRNA sequences were then assessed for appropriate chemical modifications and produced as single-stranded RNAs. We developed a fluorescence-expressing plasmid reporter system for ADAR editing by inserting a gene cassette harboring a nonsense mutation in between mCherry and EGFP, and subsequently transfected these plasmids into HeLa cells to test the candidate gRNAs. Successful editing of target gene RNA fragments produced by the reporter plasmid is measured through fluorescent microscopy, Sanger sequencing, and NGS. **Results:** Our yeast model identified three possible gRNAs, one for each candidate mutation previously mentioned. In this yeast model, the gRNAs targeting mutations in *TRPM1*, *FAM161A*, and *KIZ* showed 9%, 1%, and 0.03% editing respectively of a relevant nucleotide flanking each mutation. Using our reporter system in HeLa cells, we found that the gRNA complementary to the target *TRPM1* mutation induced RNA editing levels in our system of up to 55% in Sanger sequencing, 40% in NGS, and the treated cells expressed both mCherry and EGFP. Due to sequence constraints in *FAM161A* and *KIZ*, lower editing levels were obtained- up to 5% and 10% respectively. Off-target edited bases were observed in the vicinity of the *KIZ* edited base, while practically no off-targets were detected in the vicinity of the two other studied mutations. **Conclusions:** Targeted RNA editing utilizing the ADAR enzyme could be a useful novel form of genetic therapy due to its ability to edit SNVs in a tunable manner. Next steps include applying this technology to the appropriate knock-in mouse models and retinal organoids and studying editing efficiency on missense mutations.

Genetic Therapies Posters - Thursday

PB2059. Sustained efficacy and safety up to 3.5 years in adults with glycogen storage disease type Ia (GSDIa): results from a phase 1/2 clinical trial of DTX401, an AAV8-mediated, liver-directed gene therapy.

Authors:

J. Mitchell¹, R. Riba-Wolman², D. Rodriguez-Buritica³, A. Ahmad⁴, M. Couce Pico⁵, T. Derks⁶, D. Weinstein², D. Mitragotri⁷, V. Valayannopoulos⁷, E. Crombez⁷; ¹Montreal Children's Hosp., Montreal, QC, Canada, ²Univ. of Connecticut, Farmington, CT, ³Univ. of Texas McGovern Med. Sch., Houston, TX, ⁴Univ. of Michigan, Ann Arbor, MI, ⁵Hosp. Clínico Univ.rio de Santiago de Compostela, Santiago de Compostela, Spain, ⁶Univ. of Groningen, Groningen, Netherlands, ⁷Ultragenyx Pharmaceutical Inc., Cambridge, MA

Abstract Body:

Background: Glycogen storage disease type Ia (GSDIa) results from a deficiency of glucose 6-phosphatase (G6Pase) which is essential for glycogenolysis and gluconeogenesis. DTX401 is an investigational adeno-associated virus serotype 8 (AAV8) vector expressing the human *G6PC* gene under transcriptional control of the native promoter. **Methods:** This global, open-label, phase 1/2, dose escalation gene therapy trial (NCT03517085) is evaluating the safety and efficacy of a single DTX401 intravenous infusion in adults with GSDIa. Three patients in Cohort 1 received DTX401 2.0×10^{12} genome copies (GC)/kg, and three patients each in Cohorts 2, 3, and 4 received 6.0×10^{12} GC/kg. Cohort 4 includes a prophylactic steroid regimen to prevent transaminase elevation. **Results:** In the 12 patients enrolled in Cohorts 1 through 4, mean (SD [range]) total daily cornstarch intake reduction from baseline to Week 52 was 70.0% (23.1 [28-100%]), $p < 0.0001$. From baseline to last available timepoint (~3.5 years for three patients), mean (SD [range]) total daily cornstarch intake reduction was 73.8% (23.5 [29-100%]), $p < 0.0001$. From Cohort 3 onward, continuous glucose monitoring was implemented. In Cohort 3, the average percentage of time in euglycemia (blood glucose range of 60 mg/dL to 120 mg/dL) from Baseline to Weeks 77 to 80 increased 14%, despite a reduction in average cornstarch intake of 65%. In Cohort 4, the average percentage of time in euglycemia from Baseline to Weeks 49 to 52 remained stable (+0.5%), despite a reduction in average cornstarch intake of 51%. At Week 52 exit interviews, patients reported more energy and stamina, better mental clarity, improved glycemic control independent of cornstarch, improved sleep quality, and improvements in health-related quality of life related to cornstarch intake reduction. Patient Global Impression of Change scores at the Week 52 visit indicated that 67% of patients (n=9) felt their disease was moderately or much improved since the start of the study. All serious adverse events (Grade 1 or Grade 2: n=17; Grade 3: n=2) were classified as serious due to hospitalizations and were determined to be unrelated to study drug by both the investigator and study sponsor; all resolved. No dose-limiting toxicities were reported. **Conclusions:** DTX401 showed a positive efficacy and tolerability profile in all treated patients at Week 52 that was sustained for up to 3.5 years in patients enrolled in Cohort 1. Patients in all cohorts showed a significant reduction in cornstarch needs from baseline to Week 52 that was sustained to the last available timepoint. Additional glycemic control, fasting time, and health-related quality of life data will be reported.

Genetic Therapies Posters - Wednesday

PB2060*. Testing gene therapy for achromatopsia-like visual deficits in nonhuman primates affected by naturally occurring missense mutations in *PDE6C*

Authors:

J. Rogers¹, A. Moshiri², J. Stout¹, S. Thomasy², R. Chen¹; ¹Baylor Coll. of Med., Houston, TX, ²Univ. of California, Davis, Davis, CA

Abstract Body:

Various forms of inherited retinal disease disrupt the function of cone photoreceptors and thus affect both high acuity vision and perception of color. Many photoreceptor disorders are due to mutations in a single gene and are therefore candidates for gene therapy. Phosphodiesterase 6C (*PDE6C*) is a gene known to cause one form of cone-specific visual system dysfunction: achromatopsia. Patients with damaging mutations in *PDE6C* exhibit loss of cone photoreceptor mediated signaling and thus lose color vision and macula-dependent high-resolution vision as well as becoming photophobic due to dependence on rod photoreceptors. We have previously reported that at the California National Primate Research Center, we identified a series of rhesus macaques (*Macaca mulatta*) that displayed evidence of photophobia and apparently diminished visual system function. Whole genome sequencing showed that the two initial probands were both homozygous for an amino acid change affecting the catalytic domain of the PDE6C protein. *In vitro* assays of enzyme function demonstrated that the amino acid change (R565Q) significantly reduces enzymatic function and thus is expected to have profound effects on cone photoreceptor activity. Electroretinograms confirmed loss of cone photoreceptor signaling and retention of normal rod function. We have now, in unpublished work, tested an adenovirus-associated viral vector (AAV5) that contains the macaque *PDE6C* gene driven by the cone specific 1.7-kb L-opsin promoter. Subretinal injection of this construct into two normal control animals showed no serious adverse reaction. We next injected one eye each of two affected macaques with the *PDE6C*-carrying AAV5, one at low dose (1.5 E10 vg/eye) and one with a higher dose (1.5 E11 vg/eye). One month after injection the animal receiving the high dose of gene therapy showed marked improvement of cone photoreceptor function in the treated eye, as measured by photopic single flash and flicker electroretinogram, with no change in contralateral control eye. Six weeks after injection the injected eye showed significant rescue of signal using multifocal ERG, markedly different from the control eye. These results provide initial evidence for feasibility of gene therapy as treatment for achromatopsia due to *PDE6C* deficit. A breeding program is actively producing additional macaque carriers of this mutation in order to facilitate further studies. In parallel we are developing stem cell derived retinal organoids to compare the efficacy of gene therapy versus cell-based replacement therapy.

Genetic Therapies Posters - Thursday

PB2061. The Jackson Laboratory Center for Precision Genetics: Preclinical research for rare diseases

Authors:

C. Lutz¹, S. Murray², K. Snow¹, S. Kneeland¹, A. Zuberi¹; ¹The Jackson Lab., Bar Harbor, ME, ²Jackson Lab., Bar Harbor, ME

Abstract Body:

There are more than 7,000 known rare disease that affect more than 350 million people worldwide. The majority of rare diseases affect children, many of which do not survive to reach their fifth birthday. There are currently no FDA approved drugs for more than 95% of all rare disease. However, the landscape for rare disease is changing. Advances in high-throughput sequencing technology has accelerated the pace of novel disease allele discovery and improved the time to diagnosis for patients. Further, advances in genomic medicine, such as gene replacement, antisense oligonucleotides and genome editing have opened new doors in treatments for rare diseases. This progress has revealed a new bottleneck and challenge: the growing need for experimental systems to validate discoveries, interrogate biological mechanisms, and provide platforms for preclinical development. The mouse is an ideal model system as it recapitulates key features of human disease, and advances in genome engineering technology has enabled the rapid and efficient generation of sophisticated alleles. To fill this gap in experimental models and to facilitate research that will lead to therapeutic advancements in rare disease, the overarching goal of The Jackson Laboratory Center for Precision Genetics (JCPG) is to deliver a comprehensive animal model development pipeline to the rare disease community. Our team works with patient based organizations and clinical investigators to identify unmet need for model systems and engineer precise human disease-relevant alleles in mice. We extensively characterize these mouse models to demonstrate the key disease-relevant phenotypes and establish the line as a suitable model of disease. Finally, the JCPG can perform testing of novel therapeutic efficacy testing in the mouse models that can inform clinical trials in terms of target tissues, dose, timing of treatment and biomarker analysis. For more information on nominations, visit our website at <https://www.jax.org/research-and-faculty/research-centers/precision-genetics-center>

Genetic Therapies Posters - Wednesday

PB2062. Two years of venglustat combined with imiglucerase shows continued positive effects on neurological features and brain connectivity in adults with Gaucher disease type 3

Authors:

I. Batsu¹, **R. Zheng**², P. Minini³, D. Cherkasov³, D. Scott⁴, O. Goker-Alpan⁵, J. Hennermann⁶, K. Sakurai⁷, T. Cox⁸; ¹Sanofi, Bridgewater, NJ, ²Sanofi, Cambridge, MA, ³Sanofi, Chilly Mazarin, France, ⁴Clario, San Mateo, CA, ⁵Lysosomal & Rare Disorders Res. and Treatment Ctr., Fairfax, VA, ⁶Univ. Med. Ctr. Mainz, Ctr. for Pediatric and Adolescent Med., Mainz, Germany, ⁷The Jikei Univ. Sch. of Med., Dept. of Pediatrics, Tokyo, Japan, ⁸Univ. of Cambridge, Dept. of Med., Cambridge, United Kingdom

Abstract Body:

Background: Venglustat is a brain-penetrant glucosylceramide synthase inhibitor under investigation for treatment of Gaucher disease type 3 (GD3). After 1 year of venglustat + imiglucerase treatment in the Phase 2 LEAP trial (NCT02843035/Sanofi), target venglustat concentrations were maintained in CSF, and median glucosylceramide and glucosylsphingosine in plasma and CSF decreased markedly with favorable safety and tolerability. Ataxia scores improved, brain volume increased in patients with adequate venglustat exposure, and fMRI showed stronger connectivity between brain regions at Weeks 26/52 in relation to baseline.

Aims: Evaluate 2-year treatment outcomes in the LEAP trial.

Methods: The Phase 2 LEAP trial evaluated oral venglustat (15-mg daily) + intravenous imiglucerase (usual dose) in 11 adults with GD3 who had received ERT for ≥ 3 years and achieved non-neurological therapeutic goals before enrollment. Primary endpoints were safety, tolerability, and changes in glucosylceramide (GL-1) and glucosylsphingosine concentrations in CSF during 1 year of treatment. Secondary endpoints included plasma biomarkers, neurological assessments, volumetric and functional brain MRI, and systemic disease measures. All patients continued treatment for an additional year; 9 patients are currently in study.

Results: There were no deaths or adverse events leading to study discontinuation. Systemic manifestations remained stable. One patient had low-to-undetectable venglustat exposure after Week 26; although the cause was undetermined, non-adherence could not be ruled out. In all patients, a decrease in plasma GL-1 at 1 year continued at 2 years [median (IQR) % change from baseline: -78% (72-84) to -80% (79-85)] and sustained decreases in plasma glucosylsphingosine were observed from 1 year to 2 years [median (IQR) % change from baseline: -56% (41-60) to -56% (45-70)]. On the Scale for Assessment and Rating of Ataxia score, 91% of patients improved or had no worsening at 1 and 2 years. Brain atrophy on MRI in 10 patients had partially reversed from baseline to year 1 but markedly declined in the patient without sustained exposure to venglustat. In the remaining patients who were restudied at 104 weeks, mean brain volume remained stable, and connectivity across key brain regions (resting state networks) on functional MRI showed additive connectivity enhancement in year 2 between RSNs 1, 2 and 3 with RSN 8 among patients with venglustat exposure.

Conclusions: During two years of treatment, venglustat + imiglucerase treatment showed favorable safety and tolerability outcomes and salutary neurological improvements were maintained.

Genetic Therapies Posters - Thursday

PB2063. Venglustat, a novel, investigational, brain-penetrant glucosylceramide synthase inhibitor, for Gaucher disease type 3: phase 3 LEAP2MONO trial design

Authors:

W. Heine¹, I. Batsu², P. Minini³, D. Cherkasov³; ¹Sanofi, Cambridge, MA, ²Sanofi, Bridgewater, NJ, ³Sanofi, Chilly Mazarin, France

Abstract Body:

Background: Venglustat is a novel, small-molecule, brain-penetrant glucosylceramide synthase inhibitor under investigation as a disease-modifying therapy for Gaucher disease type 3 (GD3). Imiglucerase enzyme replacement therapy (ERT) ameliorates hematologic, visceral, and bone manifestations of GD3; however, there is currently no treatment for the neurologic manifestations. After one year of venglustat+imiglucerase in the Phase 2 LEAP trial (NCT02843035/Sanofi), venglustat concentration was maintained in CSF, median glucosylceramide and glucosylsphingosine in CSF decreased markedly, and potential to ameliorate neurological manifestations was demonstrated. **Aims:** Describe the design of the LEAP2MONO study of venglustat monotherapy for GD3. **Methods:** LEAP2MONO is a Phase 3, randomized, double-blind, trial evaluating efficacy and safety of venglustat vs imiglucerase in adults and children ≥ 12 years with GD3 who have reached therapeutic goals with ERT. Patients will be randomized to receive either once-daily oral venglustat with biweekly placebo infusions or biweekly imiglucerase infusions with once-daily oral placebo for 52 weeks (primary analysis period). Venglustat patients who meet prespecified criteria for decline in GD status can receive ERT rescue therapy. After Week 52, all patients will receive venglustat for 52 weeks (extended treatment period). Safety follow-up will continue through 30-37 days after last treatment. Enrollment is planned for 40 participants, including ≈ 10 pediatric participants. Primary endpoints are change from baseline to Week 52 on the Scale for Assessment and Rating of Ataxia modified total score and the Repeatable Battery for the Assessment of Neuropsychological Status total scale index score. Secondary endpoints include safety, tolerability, and changes from baseline in spleen volume, liver volume, hemoglobin, and platelets. Exploratory endpoints include changes from baseline in glucosylceramide and glucosylsphingosine in plasma and CSF, venglustat pharmacokinetics, and patient-reported outcomes. This study will evaluate whether the effects of venglustat monotherapy translate to clinical benefit on neurological manifestations with maintenance of systemic disease stability in GD3 patients.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2064. A novel human mitochondrial donating cell type promotes maturation of GV oocytes

Authors:

T. Lee¹, Y. Chan¹, T. Li¹, H. Chan¹, Q. Lian²; ¹The Chinese Univ. of Hong Kong, Hong Kong, China, ²Hong Kong Univ., Hong Kong, China

Abstract Body:

Delayed marriage and parenthood have been increasingly common and have become a global issue, causing a significant increase of advanced maternal age (AMA) mothers. Aging leads to the qualitative and quantitative decline of oocytes, which translates to poor embryo development and available mature oocytes for IVF. Currently, there is no proven solution to rescue or increase the number of usable mature oocytes in ART procedure. We recently developed a method to increase the number of mature (MII) oocytes through *in vitro* maturation of immature (GV) oocytes with mesenchymal stem cells (MSCs). This novel non-invasive method allows producing more mature (MII) oocytes for IVF and improves the quality. We demonstrated Green fluorescence Protein (GFP)-labelled mitochondria were transferred to mouse oocytes via tunneling nanotubes (TNT). Here we show an alternative non-stem cell human cell source that can serve similar mitochondria donating function as MSCs. Our results revealed that MSC-mediated mitochondrial transfer can improve mitochondrial function and the average maturation rate of GV oocytes by 38% and 243% respectively. The study thus provides a foundation for creating more usable oocytes for ART procedure.

Funding: The work is supported by Hong Kong Research Grant Council (GRF project no. 14111320).

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2065. Assessing the accuracy of variant detection and coverage bias in mitochondrial DNA (mtDNA) after whole genome amplification of single fibroblast cells with known mtDNA heteroplasmic mutations to effectively optimize preimplantation genetic testing of mtDNA related disorders.

Authors:

A. Novoselska^{1,2}, A. Aggarwal¹, R. Antes¹, Y. S. S. Kwok^{1,2}, M. Ho¹, M. Ho¹, C. Librach^{1,2,3,4,5}, S. Madjunkova^{1,6}; ¹CReATe Fertility Ctr., Toronto, ON, Canada, ²Dept. of Physiology, Univ. of Toronto, Toronto, ON, Canada, ³Dept. of Obstetrics and Gynecology, Univ. of Toronto, Toronto, ON, Canada, ⁴Inst. of Med. Sci., Univ. of Toronto, Toronto, ON, Canada, ⁵Sunnybrook Res. Inst., Toronto, ON, Canada, ⁶Dept. of Lab. Med. and Pathobiology, Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Objectives: Knowledge regarding maternal transmission of mtDNA heteroplasmy (het) to preimplantation embryos is lacking due to the technical bias of whole genome amplification (WGA), which is necessary to obtain enough DNA from 5-10 trophectoderm (TE) biopsied cells during preimplantation genetic testing (PGT). WGA can introduce uneven coverage bias and false positive single-nucleotide changes (SNVs) in genomic DNA, which can lead to inaccurate detection of variants. Our aim was to assess and compare technical bias between two widely used WGA platforms, REPLI-g (RPG) (QIAGEN) and SurePlex (SPX) (Illumina), using two different fibroblast cell lines with known mtDNA levels at specific loci. We also aimed to compare the accuracy of het levels detection with these platforms to the known het levels.

Materials and Methods: We used two different fibroblast cell lines with known het levels: F49 - 95% het at m.8993T>C and F101 - 40% het at m.10191T>C in the mtDNA. DNA was extracted from 1.7 x 10⁶ cells from F49 and from F101 using DNeasy Kit (QIAGEN) to represent the unamplified (UA) source of DNA (n=2). Three technical replicates of 10 single cells from each line were collected for both platforms to represent the mock TE DNA source (n=12). For complete mtDNA enrichment, DNA Prep with Enrichment (Illumina) was performed during library prep using custom probes (Twist Bioscience). Sequencing was performed with NextSeq550 (Illumina) and data analysis with the DRAGEN Enrichment (Illumina) analysis pipeline.

Results: All samples were sequenced at an average (av.) depth of 2971x. The av. het level detected at m.8993T>C for RPG (60%) and for SPX (54%) was similar, but significantly lower than the 95% known het (p<0.05). At m.10191T>C, the av. het levels for RPG (34%) and SPX (30%) were comparable, but significantly lower than the 40% known het (p<0.05). A total of 28 variants were detected in UA-F101, 38 in WGA F101, 10 in UA-F49 and 30 in WGA F49. The concordance of variants between RPG and SPX was 92% for F101 and 77% for F49. Between UA and RPG, variant concordance was 93% for F101 and 91% for F49. Between SPX and UA reference, the concordance was 83% for F101 and 78% for F49. Both WGA kits had allele dropout occurring at 2 positions (m.4769 and m.8860) for both cell lines, and amplification bias occurred at 8 positions only for F49.

Conclusions: Whole genome amplification by RPG or SPX, as used in PGT, has relatively high reproducibility for amplification of mtDNA. Both WGA kits have skewed het detection which should be assessed individually for specific variants to limit misinterpretation of PGT for mtDNA mutations.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2066. Assessing the magnitude of clinical utility of preconception expanded carrier screening for prevention of neurodevelopmental disorders

Authors:

P. Boonsawat¹, A. Horn^{1,2}, K. Steindl¹, A. Baumer¹, P. Joset³, D. Kraemer¹, A. Bahr¹, I. Ivanovski¹, E. Cabello¹, M. Papik¹, M. Zweier¹, B. Oneda¹, P. Sirleto¹, T. Burkhardt⁴, H. Sticht², A. Rauch^{1,5}; ¹Inst. of Med. Genetics, Schlieren-Zürich, Switzerland, ²Inst. of Biochemistry, Erlangen, Germany, ³Med. Genetics, Basel, Switzerland, ⁴Univ. Hosp. Zurich, Zurich, Switzerland, ⁵Univ. Children's Hosp. Zurich, Zurich, Switzerland

Abstract Body:

The magnitude of clinical utility of preconception expanded carrier screening (ECS) concerning its potential to reduce the risk of affected offspring, especially those with disabling neurodevelopmental disorders (NDDs), is unknown. Since determination of such risk-reduction potential in general populations would require huge cohort numbers and long-term follow-up data, we retrospectively studied the parental samples of a cohort of children with disabling NDDs, who previously had undergone extensive diagnostic work-up. To this end, we used exome sequencing data from 700 parents of children with NDD and blindly screened for carrier-alleles in up to 3,046 recessive/X-linked genes. Depending on variant pathogenicity thresholds and gene content, risk-reduction potential for NDD was up to 43.5% in consanguineous, and 5.1% in nonconsanguineous couples. This risk-reduction potential was compromised by underestimation of pathogenicity of missense variants in 16 couples due to non-automatable criteria not available for the screening set-up, which corresponds to a false-negative ECS rate of 4.6%. Furthermore, un-investigated inherited copy-number variants and compound heterozygosity of one inherited and one de novo variant that was not detectable by carrier screening (0.9% each) also mitigated the risk-reduction potential. Adherence to the ACMG recommendations of restricting ECS to genes with high carrier frequency for nonconsanguineous couples would not detect diagnoses in 9 couples, which accounted for more than 50% of the detectable inherited NDD-risk in this group. While such guidelines, as well as most previous studies focused on restricted gene lists of supposed moderate to severe disease genes, screening of, for instance, 1,300 genes, listed as severe by the Mackenzie's Mission Project in 2020, would have missed ~30% of our recessive diagnoses. Thus, for optimized clinical utility of ECS, screening in recessive/X-linked genes regardless of their frequency (ACMG Tier-4) and sensible pathogenicity thresholds should be considered for all couples seeking ECS. We therefore suggest a feasible approach for improved variant classification in carrier screening resulting in high sensitivity without compromising specificity. Taken together, our findings revealed the magnitude of utility of preconception ECS for NDD-risk in relation to parental consanguinity, screened gene content and pathogenicity assessment, and will inform genetic counselling in reproductive medicine.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2067. Assisted reproductive technology and DNA methylation at imprinted genes

Authors:

R. Lyle^{1,2}, Y. Lee³, C. Page^{4,5}, H. E. Nustad^{6,7}, P. Magnus⁷, S. E. Håberg⁷; ¹Oslo Univ. Hosp., Oslo, Norway, ²Norwegian Institution of Publ. Hlth., Oslo, Norway, ³Norwegian Inst. of Publ. Hlth., Oslo, Norway, ⁴Univ. of Oslo, Oslo, Norway, ⁵Norwegian Inst. of Publ. Hlth., Oslo, Norway, ⁶Deepinsight, Oslo, Norway, ⁷Norwegian Inst. of Publ. Hlth., Oslo, Norway

Abstract Body:

Assisted reproductive technology (ART) use is increasing, and around 8 million children have been conceived using ART. As more women postpone childbearing, and egg freezing is becoming more socially acceptable, this increase is expected to continue. ART is associated with several adverse pregnancy outcomes, with implications for childhood and adult health. The impact of ART on long-term health is not clear. Some studies have suggested an influence on neurodevelopment, cardiovascular function, metabolism, growth and malignancies. However, it is unclear whether the differences observed in ART conceived children are caused by the ART procedure itself or by underlying factors associated with parental subfertility.

Epigenetic mechanisms regulate gene activity, cell function, and development. DNA methylation, the attachment of a methyl group to a cytosine preceding a guanine base (CpG), is an epigenetic mechanism essential for normal embryonic development. ART involves the manipulation and culturing of embryos during a period that coincides with extensive genome-wide epigenetic remodeling. It is therefore plausible that ART procedures alter the fetal epigenome, potentially perturbing development and thereby influencing later health. So far, there is limited and inconsistent evidence for ART-induced DNA methylation changes, with the exception of findings in imprinting disorders.

Genomic imprinting is an epigenetic phenomenon that results in parent-of-origin gene expression. Imprinting is regulated by DNA methylation at both primary imprinting control regions (ICRs), which are inherited from the parental gametes and are not changed during epigenetic remodelling, and secondary differentially methylated regions (DMRs), which are reestablished during early development. It is thus important to distinguish between ICRs and DMRs when considering the effects of ART.

To investigate if ART perturbs DNA methylation at imprinted loci, we measured DNA methylation (Illumina EPIC array) in trios of 962 ART and 983 naturally conceived newborns (5835 samples total) from the Norwegian Mother, Father and Child Cohort study (MoBa). Using trios allows us to address the potential influence of underlying parental subfertility on newborn DNA methylation. Here we show that ART conceived newborns display differences in DNA methylation at certain imprinted genes.

Associations persist after controlling for parents' DNA methylation, and are not explained by parental subfertility.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2068. Capacity Building: Exploring the Use of Newborn Screening Data to Understand Neurodevelopmental Outcomes

Authors:

Z. Talebizadeh¹, J. Taylor², G. Tona³, Y. Unnikumar⁴, K. Chan⁵, A. Brower⁶; ¹ACMG, Overland Park, KS, ²American Coll. of Med. Genetics and Genomics, Baltimore, MD, ³NBSTRN, Bethesda, MD, ⁴American Coll. of Genetics and Genomics, Silver Spring, MD, ⁵American Coll. of Med. Genetics and Genomics, Bethesda, MD, ⁶American Coll. of Med. Genetics and Genomics, Dakota Dunes, SD

Abstract Body:

Introduction: Newborn screening (NBS) is the largest public health genetic program with 3.8 million newborns screened for up to 81 genetic conditions each year. Many of these conditions have comorbidities, including autism, intellectual and/or developmental disabilities (IDD), and these comorbidities impact health outcomes across the lifespan. The majority of screened conditions are inborn errors of metabolism (IEM), frequently associated with IDD and/or neurological problems. The Longitudinal Pediatric Data Resource (LPDR) is one of the data tools developed by the Newborn Screening Translational Research Network. The LPDR captures, stores, analyzes, visualizes, and shares genomic and phenotypic data over the lifespan of NBS identified newborns to facilitate understanding of genetic disease and to assess the impact of early identification and treatment. Data deposited in the LPDR is available for secondary use by the research community. We undertook an effort to 1) evaluate the analytical utility of the existing data to comprehend neurodevelopmental trajectories, and to 2) prioritize question-answer sets for developing further common data elements (CDEs) for metabolic disorders. **Methods:** We examined health information from a completed, 10-year natural history study of IEMs, called Inborn Errors of Metabolism Collaborative. We extracted data elements related to either neurodevelopmental status or the associated clinical symptoms. A mixed methods approach (data mining as well as community-based participatory approach, focus group, and systematic reviews) was used to conduct statistical power analysis and a feasibility assessment, including identification of needs and priorities. **Results:** We will describe a strategy roadmap for utilizing the NBS system to understand neurodevelopmental outcomes, by outlining 1) aggregated data points from completed studies within the LPDR, 2) developed CDEs including branching logic to facilitate future data collection and analytical utility, and 3) identified barriers, facilitators, and motivators along with suggested recommendations. **Conclusion:** The present study provides a new capacity assessment of this unique public health system from the standpoint of understanding neurodevelopmental outcomes. Documenting behavioral and neurodevelopmental status of participants in long-term follow up NBS studies, using consensus-based codes and standards, would expand the utility of the LPDR datasets for use by the research community. Findings from this capacity building project can inform future efforts toward advancing research infrastructure for this established public health program.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2069. Cell free fetal (cff) DNA assessment using Y- chromosome specific sequences and amplicon NGS of targeted SNPs for possible implementation in non-invasive prenatal testing of common chromosomal aneuploidies

Authors:

R. Vodicka, J. Bohmova, R. Vrtel, M. Prochazka; Univesity Hosp. Olomouc, Olomouc, Czech Republic

Abstract Body:

Non-invasive prenatal testing belongs to the main in the field of prenatal diagnostics of common aneuploidies. Assessment of the fetal fraction is an important part of the entire test procedure. This study aims to assess the fetal fraction using Y-specific sequences in the created NGS panel of single-nucleotide polymorphisms (SNP) and to estimate the possibilities of using this panel for the NIPT. Particularly, we focused on the coverage analysis of 8 SNPs (rs9786043; rs13305774; rs9786562; rs72617693; rs61797733; rs3853054; rs3900; rs57351463) from the Y chromosome to assess and estimate the total amount of reads originated from cffDNA. The study included 17 samples of DNA isolated from maternal plasma of pregnant women with male fetuses from 13 to 17 weeks of pregnancy. Using a database of human genetic variants and polymorphisms from the UCSC genome browser there was designed *in-silico* set of probes targeting in population common SNPs polymorphisms from chromosomes 13, 18, 21, X, Y and from the reference chromosomes. A total of 408 SNPs have been proposed, with respect that the length of the amplicon should be around 100bp. The design was done and evaluated using Ion AmpliSeq Designer. Of the 408 SNPs, a total of 317 polymorphisms from which 8 is from the Y chromosome, 42 from the X chromosome, 72 from chromosome 21, 78 from chromosome 18, 89 from chromosome 13 and 28 from the others (chr1; chr2; chr4; chr5 ; chr6; chr8; chr10; chr12; chr14; chr16; chr17; chr19; chr20; chr22) could be used for in vitro amplification and NGS sequencing. Particular NGS libraries was carried out using the Ion Chef™ device and the Ion 510™/520™/530™ Kit-Chef kit, which provides automatic library preparation, templating and the chip loading. The Ion 540 chip was used for sequencing on Ion S5™ System. The quality of the sequencing data, mapping and analysis of the coverage, was carried out using Torrent Suite software. The total number of reads ranged from 1291449 to 3166718 (average 2377876) per sample. The average number of reads (excluding Y-specific) per SNP ranged from 4179 to 10245 (average 7264). The average number of Y chromosome specific reads per SNP was used as fetal DNA predictor. The total number of cffDNA reads so were estimated from 4 000 to 263 000 (median/average is 73 000/93 000). A number of mapping reads, after appropriate analysis parameters setting that include rich GC amplicon filtration and for particular SNP, the coefficients of amplification robustness and reliability, could be sufficient to reliably distinguish the most common trisomic pregnancies. The study was supported by Ministry of Health, Czech Republic - conceptual development of research organization (FNOI, 00098892)

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2070. Cell-informed atlas of gene, isoform, and splicing regulation in the developing human brain identifies candidate causal variants for neuropsychiatric GWAS

Authors:

C. Wen¹, PsychENCODE Consortium, R. Dai², P. F. Przytycki³, D. Vo¹, E. Tsai¹, C. Hoh¹, A. Bhattacharya¹, M. Kim¹, D. Clarke⁴, K. S. Pollard³, N. Daskalakis⁵, D. Pinto⁶, C. Liu², M. Gandal¹; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²SUNY Upstate Med. Univ., Syracuse, NY, ³Gladstone Inst.s, UCSF, San Francisco, CA, ⁴Yale Univ., New Haven, CT, ⁵Harvard Med. Sch., McLean, Belmont, NY, ⁶Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Genome-wide association studies (GWAS) have identified hundreds of loci associated with neurodevelopmental psychiatric disorders, yet our understanding of the risk genes and mechanisms underlying these associations remains very limited. Strong enrichment for psychiatric genetic risk loci among fetal brain regulatory elements has prompted several recent efforts to characterize gene expression regulation in the developing human brain through expression quantitative trait loci (eQTL) for integration with GWAS. Yet, previous integrative efforts have been hindered by the small scale of individual studies, and lack of investigation of alternative splicing and cell-type-specific mechanisms. Here, we uniformly process and systematically mega-analyze gene-, isoform-, and splicing-QTLs, as well as cell-type-specific QTLs from 9 deconvoluted cell-types, in the developing human brain across 654 donors spanning 4-39 post-conception weeks (PCW). We identified 17,847 genes harboring a significant QTL, including 2,095 (11.7%) with QTLs detectable only in a cell-type-specific context. Network-based integration of fetal single-cell chromatin accessibility further mapped 18.4% of eQTL-containing genes to specific cell types. Leveraging this resource to gain mechanistic insight into neuropsychiatric disorders, we performed joint statistical fine-mapping of fetal brain QTLs with summary statistics from five neuropsychiatric GWAS, including schizophrenia (SCZ), autism spectrum disorder (ASD), bipolar disorder (BIP), attention deficit hyperactivity disorder (ADHD), and major depressive disorder (MDD). From a total of 485 genome-wide significant loci across these 5 traits, we were able to successfully account for 214 unique loci (44.1%) with a significant GWAS-QTL colocalization (CLPP > 0.01) in our fetal brain functional genomics reference panel. Of these, 95, 137, and 140 colocalizations were captured by eQTLs, isoQTLs, and sQTLs, respectively, highlighting the important contributions of alternative splicing to psychiatric genetic risk mechanisms. Cell-type-specific eQTLs further identified 121 colocalized GWAS loci, including 39 not captured from bulk-tissue eQTLs. Newly prioritized candidate risk genes overlap with orthogonal results from rare-variant sequencing studies of developmental disorders, such as *SP4* and *ASH1L* in SCZ. Altogether, this work provides a comprehensive view of genetic regulation in the developing human brain and demonstrates the importance of developmental context, cell-type specificity, and splicing/isoform regulation when prioritizing candidate mechanisms underlying GWAS variants.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2071. Characterization of the prenatal presentation of novel truncating variants in *ANKRD11*.

Authors:

J. Van Ziffle¹, B. R. Lianoglou¹, U. Hodoglugil², W. Devine¹, P-M. Martin¹, J. Shieh³, A. Slavotinek⁴, T. Sparks¹, M. Norton³; ¹Univ. of California, San Francisco, San Francisco, CA, ²UCSF, San Francisco, CA, ³Univ. of California San Francisco, San Francisco, CA, ⁴Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH

Abstract Body:

KBG syndrome has been described primarily in the pediatric setting, associated with inactivating variants in *ANKRD11*. Characteristic postnatal features of KBG syndrome are developmental delay, intellectual disability, behavioral abnormalities, macrodontia, and characteristic facial features. Skeletal anomalies, short stature, early feeding difficulties, hearing loss, brain malformations and seizures have also been reported. The prenatal phenotype is not well understood aside from growth deficiency. We describe six cases diagnosed in a single institution, three diagnosed prenatally and the remaining at 10, 12 and 21 years of age. The prenatal cases were identified by exome sequencing (ES) referred for abnormalities on routine prenatal imaging. All three cases were found to have *de novo* truncating variants in *ANKRD11*. Case 1 was a male fetus with complete atrioventricular septal defect and coarctation of the aorta. Case 2 was a female fetus with pericardial effusion and elevated MCA Doppler suggesting fetal anemia. Periumbilical blood sampling was performed and confirmed anemia (hematocrit: 24%), followed by red blood cell transfusion at 22 weeks. Case 3 presented with congenital diaphragmatic hernia. The pediatric cases in our cohort were reported to have unremarkable prenatal structural imaging and birth history. Case 4, diagnosed at age 21, presented at age 3 with speech delay, intellectual disability, facial features, short stature and a positive family history for intellectual disability. Cases 5 and 6 are siblings; case 5 was diagnosed at age 10, with initial presentation at age 2 including speech and developmental delays, seizures, aplasia cutis, mild facial dysmorphism; case 6, diagnosed at age 12, presented with speech and developmental delay, non-union of the right tibia after recurrent fractures, right lower limb hemihypertrophy, demineralization of femurs, and bilateral dacryostenosis. These findings expand our understanding of the heterogeneous presentation of KBG. While the postnatal cases presented uniformly with developmental delays, the prenatal cases all had abnormal prenatal imaging including structural or physiologic differences not well described as a part of *ANKRD11*-associated KBG, although the prenatal congenital heart condition is likely related. It remains uncertain if the other prenatal cases represent phenotype expansion or incidental findings. For the postnatal cases, the diagnosis was made 8 and 18 years after the initial parental concern. This report demonstrates the value of ES to identify rare genetic diagnosis in cases of non-specific findings in the prenatal and postnatal period.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2073*. Comprehensive preimplantation genetic testing for monogenetic disorders (PGT-M) and aneuploidy (PGT-A) using whole genome sequencing and haplotyping

Authors:

S. Chen¹, R. Abramov¹, R. Antes¹, M. Madjunkov^{1,2}, C. Librach^{1,2,3,4,5}, S. Madjunkova^{1,6}; ¹CReATe Fertility Ctr., Toronto, ON, Canada, ²Dept. of Obstetrics and Gynecology, Univ. of Toronto, Toronto, ON, Canada, ³Dept. of Physiology, Univ. of Toronto, Toronto, ON, Canada, ⁴Inst. of Med. Sci., Univ. of Toronto, Toronto, ON, Canada, ⁵Sunnybrook Res. Inst., Toronto, ON, Canada, ⁶Dept. of Lab. Med. and Pathobiology, Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Introduction: Wider use of expanded carrier screening, genetic diagnostic tests, improved awareness and higher utilization of in vitro fertilization (IVF) have increased demand for preimplantation genetic testing for monogenetic disorders (PGT-M), as a therapeutic alternative to invasive prenatal tests; not only for common severe childhood and rare disorders, but for serious and mild late onset disorders, including for variants of unknown significance. This requires adaptation of generic genome wide approaches to complement or eventually replace targeted disease-specific PGT-M assays and add the PGT-A testing as a first-tier test. We aimed to evaluate concurrent use of genome wide haplotyping and disease specific assays for PGT-M and high resolution PGT-A in an unselected patient population at risk for monogenetic disorders. **Methods:** 104 couples at risk for monogenetic disorders were analyzed at the CReATe Genetic Laboratory, Toronto, Canada. For each couple DNA from oocyte and sperm provider, related family member/s and whole genome amplified (WGA) DNA (REPLI-g, WGA Kit, Qiagen) from a single trophectoderm (TE) biopsy of embryos created through IVF were used. For PGT-M whole genome SNParray typing (Karyomapping, Illumina, CA) and STR analysis were used to obtain parental haplotypes and Sanger sequencing was used for specific direct variant analysis. PGT-A was performed using Illumina library prep and sequencing on NextSeq 550. The Nxclinical (BioDiscovery) and BlueFuse Multi (Illumina) softwares were used for data analysis. **Results:** 948 embryos, from 144 IVF cycles were analyzed for 53 unique genes (14 cancer predisposing genes, 20 autosomal recessive and 19 dominant) and total of 240 variants. PGT-A analysis showed 32.7% aneuploidy, 9.7% mosaicism and 57.6% euploidy. Euploid and mosaic embryos were tested with Karyomapping that had an average case call rate of 92%, and an average Allele Drop Out rate of 2.7%. Recombination events in 3% of cases prevented assignment of haplotype. The distribution for both recessive and dominant disorders was in Mendelian order. Direct mutation testing was obtained for all samples and was 100% concordant with the predicted haplotypes. In cases with no predicted haplotype, direct testing established the PGT-M diagnosis. To date 45 live births confirmed the PGT-M diagnosis. **Conclusion:** Karyomapping is effective as a generic haplotyping platform, with low risk of undetected recombination events in the genomic region of interest. Simultaneous PGT-A and PGT-M with optional direct variant testing provides optimal patient management that can significantly reduce the time to pregnancy in couples at risk for inherited disorders.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2074. Contribution of BBS9 isoforms to the etiopathogenesis of non-syndromic craniosynostosis

Authors:

D. Sibilia¹, E. Lo Cascio¹, L. Di Pietro¹, A. Vita¹, G. Tamburrini², L. Massimi², O. Parolini¹, A. Arcovito¹, W. Lattanzi¹; ¹Università Cattolica del Sacro Cuore, Rome, Italy, ²Fondazione Policlinico Univ.rio Agostino Gemelli, Rome, Italy

Abstract Body:

Background. Skull development relies on the presence of calvarial mesenchymal stromal cells (CMSCs) residing in the cranial sutures. Congenital alterations that hamper CMSC homeostasis lead to premature suture ossification, causing craniofacial malformations in infants, known as craniosynostosis (CS; affecting 1/2000-2500 newborns). In particular, non-syndromic CS (NCS) are the most prevalent forms (75-80% cases) and have a poorly clarified etiopathogenesis. Nevertheless, a number of osteogenic and developmental molecular pathways have been reported to be impaired in NCS, few of them localizing at the primary cilium, a membrane organelle involved in mechanotransduction and cell-to-cell signaling. The BBSome is a multimeric protein complex that localizes at the primary cilium and translocates cargo proteins throughout this organelle; among its subunits, BBS9 is crucial for the BBSome assembly and integrity. Recent evidence suggested an implication of BBS9 in NCS: i) *BBS9* gene was associated as a susceptible locus for NCS in a GWAS; ii) our group found an overexpression of some BBS9 splice isoform-specific exons in prematurely fused sutures of NCS patients. Our hypothesis is that the BBS9 implied isoforms may lack of the C-term region required for the BBSome assembly with the overexpressed exons falling in the N-term and central regions. **Aims.** Based on this background, our aim is to identify the overexpressed BBS9 isoforms in NCS during the osteogenic process. **Methods.** We used an *in vitro* model of N-CMSCs (from unfused sutures) and P-CMSCs (from fused sutures) isolated from surgical wastes from cranioplastic surgery on NCS pediatric patients. We evaluated BBS9 differential expression during osteogenic differentiation of CMSCs using qPCR and WB techniques. **Results.** Even though the *in silico* model of the BBSome containing a BBS9 isoform lacking of the C-term region showed an impairment in the whole BBSome stability, the transcript level analysis showed a similar expression trend of all the isoforms between N- and P- cells, both in proliferative and osteogenic medium. On the contrary, protein analysis corroborated with the previous results: BBS9 protein is overexpressed in P-CMSC compared to N-CMSC upon osteogenic induction. **Conclusions and future studies.** BBS9 protein isoform expression seems to vary quite reproducibly during osteogenic differentiation of CMSCs and shows a differential trend in N-versus-P cells. Further ongoing studies are aimed to identify the primary structure of selected isoforms and to confirm their role in the BBSome assembly, hence to define their contribution in NCS onset.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2075. *De novo* mutations disturb early brain development more frequently than common variants in schizophrenia

Authors:

T. Itai¹, P. Jia¹, Y. Dai¹, J. Chen², X. Chen¹, Z. Zhao¹; ¹The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ²Univ. of Nevada Las Vegas, Las Vegas, NV

Abstract Body:

Schizophrenia (SCZ) is a neuropsychiatric disease with high heritability. The genetic etiology of SCZ is heterogeneous, consisting of common variants and rare variants, some of which occur *de novo* (DNM). Although a few thousand DNMs have been identified, there is limited knowledge of how they function in SCZ etiology. We compared temporal and cell-type expression data in SCZ associated genes with common variants and DNMs. We collected 2636 DNMs in 2263 genes across 3477 SCZ patients (SCZ-DNMs). Since statistical enrichment tests failed to extract a sufficient number of DNM genes for downstream analyses, we curated two gene lists based on priori knowledge. The first list had 106 genes (SCZ-neuroGenes) from SCZ-DNMs that are intolerant to loss-of-function and missense DNMs and are neurologically essential. The second list had 52 genes (SCZ-moduleGenes) derived from network analyses of SCZ-DNMs with MAGI and GeneMANIA. We used 120 genes (SCZ-commonGenes) from a recent GWAS study. To compare gene expression between pre-/postnatal stages, we used BrainSpan and BrainCloud datasets with a linear model. To quantify the involvement of each gene in prenatal brain development, we calculated a fetal effect score (FES), which is defined as $-\beta/se$, where β is the linear model's coefficient for *stage* (pre- or postnatal) and *se* is the standard error of β . SCZ-neuroGenes and SCZ-moduleGenes were highly expressed in the prenatal stage and had higher FESs than SCZ-commonGenes. The average FESs for SCZ-commonGenes, SCZ -neuroGenes, SCZ -moduleGenes were -0.8, 4.8, 12.7 respectively in the whole brain, that for the dorsolateral prefrontal cortex were -0.5, 0.9, 3.5, and for hippocampus were -0.3, 0.8, 3.0. For cell-type expression analysis, we calculated the specificity indexes (SIs) using single-cell expression data from cerebral cortices in humans and mice. In human data, SCZ-neuroGenes and SCZ-moduleGenes had higher SIs in fetal replicating cells than that in the SCZ-commonGenes (the median SIs for SCZ-commonGenes, SCZ -neuroGenes, SCZ -moduleGenes were 0.04, 0.08, 0.15). In mouse data, SCZ-neuroGenes and SCZ-moduleGenes had higher SIs in undifferentiated cell types. The median SIs for SCZ-commonGenes, SCZ -neuroGenes, SCZ -moduleGenes were 0.013, 0.024, 0.027 for dopaminergic neuroblasts, that for neural progenitors were 0.017, 0.029, 0.040, and for neuroblasts were 0.022, 0.030, 0.032. Our results showed that SCZ-neuroGenes and SCZ-moduleGenes are expressed in an earlier period and fetal-specific cell types than SCZ-commonGenes, suggesting that expression patterns in specific cell types in early fetal stages have significant impacts on the risk of SCZ.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2076. Developmental programming of vitamin b12 deficiency model in zebrafish

Authors:

A. Verma, S. S. Nongmaithem, M. Challapalli, S. Ramachandran, G. R. Chandak; CSIR-CCMB, Hyderabad, India

Abstract Body:

Vitamin B12 (B12) is one of the key co-factors in One-Carbon Metabolism (OCM) pathway which influences various biological processes including DNA synthesis, amino acid homeostasis, epigenetic regulation, and redox defense. We have shown earlier that B12 deficiency leading to an altered OCM is highly prevalent in the Indian population and is associated with low birth weight and development defects including neural tube defects, congenital malformations, etc. Zebrafish have been used as a tractable model to study embryonic development due to their ability to produce large clutches of *ex utero* developing optically transparent embryos. Transcobalamin 2 (*tcn2*) is the only known B12 transporter in zebrafish and has a similar structure and B12-binding affinity with human TCN2. This study aimed to understand the role of B12 in early fetal developmental programming by generating a zebrafish model of B12 deficiency using the CRISPR-Cas9 system. We generated two *tcn2* knockout (KO) alleles by targeting the exon 7 resulting in premature stop codons at the end of exon 7. Complete disruption of *tcn2* in the zebrafish was established by qRT-PCR and immunoblotting. The F2 generation *tcn2*^{-/-} zebrafish grew normally and did not show any overt phenotype. However, in the F3 generation, nearly 70% of *tcn2*^{-/-} fish died during embryonic development. Majority of *tcn2*^{-/-} embryos showed delayed somitogenesis as compared to WT embryos, remained largely unhatched, and showed pericardial edema, curved or deformed tails, abnormal yolk extension, and yolk sac edema. These phenotypes in F3 were partially ameliorated by the addition of 10 μM B12 (cyanocobalamin) to the embryo water throughout the developmental stages. We performed RNA sequencing analysis at one-cell stage in KO and wild type (WT) and at 24 hpf stage in three groups - KO, KO treated with B12 (KO_B12), and WT to identify differentially expressed genes (DEGs) between these groups. The gene set enrichment analysis of the DEGs showed changes in key pathways including cell cycle, MAPK signaling, calcium signaling, cellular senescence, apoptosis, etc. The opposite direction of the effect of majority of DEGs between KO/WT and KO_B12/WT indicates their potential role in the rescue of *tcn2*^{-/-} phenotype by B12 treatment. Our study on the *tcn2*^{-/-} zebrafish model delineates the importance of B12 in different developmental stages. At a molecular level, we report B12 deficiency-mediated dysregulation of several developmental processes which may explain its role in various human phenotypes. Survival of nearly 1/3rd of *tcn2*^{-/-} embryos in the F3 generation suggests a possible novel mechanism of B12 transport and needs to be investigated further.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2077. Developmental trajectory analysis of differentiating mouse sensory interneurons to further improve stem-cell differentiation protocols.

Authors:

E. Heinrichs^{1,2,3}, S. Gupta^{1,3}, R. Kawaguchi⁴, B. Novitch^{1,3}, S. Butler^{1,3}; ¹UCLA Neurobiology, Los Angeles, CA, ²UCLA Human Genetics, Los Angeles, CA, ³Eli and Edythe Broad Ctr. of Regenerative Med. and Stem Cell Res. at UCLA, Los Angeles, CA, ⁴UCLA Psychiatry, Los Angeles, CA

Abstract Body:

The developing dorsal spinal cord is populated with different classes of interneurons that integrate and relay sensory information including pain, itch, touch and heat. Loss of these modalities can be emotionally and physically devastating for patients. However, current research seeking to regenerate the spinal cord has focused on the recovery of coordinated movement, rather than somatosensation. Nonetheless, a full recovery from spinal cord injuries will require the restoration of both motor and sensory function. Towards this goal, we have recently published directed differentiation protocols capable of making heterogenous dorsal interneuron (dI) populations. While these protocols are a key step forward, protocols that direct individual dI types have remained elusive. Here we present further analysis of our published mouse dI differentiation single cell transcriptomic dataset to provide new insights into the genetic factors driving the early stages of dI differentiation. Using a combination of Seurat and Monocle3 packages in R, we plot the developmental trajectories of differentiating dIs and perform pseudotemporal analyses to identify candidate genes that direct specific developmental bifurcations. Analyses of known marker genes show that dorsal progenitor expression peaks at lower pseudotime values than post-mitotic dI genes, validating the approach. By examining differential gene expression between trajectories, we have identified novel marker genes including *Smoc1*, which is highly temporally restricted to the earliest stages of dI5 differentiation. We have also observed novel developmental hierarchies by plotting these trajectories in three-dimensional UMAP space. Specifically, dI1 and dI5 bifurcate to become distinct populations very early during the differentiation process, while dI2-dI4 have a common origin, and only become different from each other later in differentiation. Collectively, these results reveal the developmental logic by which spinal somatosensory neurons are generated, which can be recapitulated in vitro using pluripotent stem cells to enable disease modeling and therapeutic discovery.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2078. Discordances among NIPT, QF-PCR and cytogenetic results in prenatal cases.

Authors:

F. Thuriot^{1,2}, A. Gomez^{1,2}, S. Farhan^{1,2}, A. Ruchon^{1,2}, G. Tchakarska^{1,2}, J. Lavoie^{1,2,3}; ¹Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, ²McGill Univ. Hlth.Ctr., Montreal, QC, Canada, ³Dept. of Pathology, McGill Univ., Montreal, QC, Canada

Abstract Body:

Non-invasive prenatal testing (NIPT) is being increasingly used as a first-tier screening test in Canada and around the world. However, despite the advantages of NIPT, it is not a diagnostic test and invasive prenatal testing must be used to confirm all cases of screen positive NIPT results. Quantitative fluorescent polymerase chain reaction (QF-PCR) can be used as a rapid aneuploidy detection test. Based on these results and depending on the clinical indication, chromosomal microarray (CMA) and karyotype analysis may be ordered.

Here, a few prenatal cases where discordances have been observed among NIPT, QF-PCR and cytogenetic testing (CMA and/or karyotype) will be illustrated.

The first case showed a NIPT result of monosomy X. QF-PCR was performed on the amniotic fluid and chromosome Y material was detected. Karyotype analysis was subsequently performed to further characterize these findings. A mosaic karyotype with a cell line with a pseudodicentric chromosome Y and a cell line with a 45,X karyotype was detected. The second case had a normal QF-PCR result; however, given the abnormal ultrasound findings, CMA was ordered and a partial terminal deletion of the long arm of chromosome 13 was observed. Importantly, the QF-PCR markers present in the long arm of chromosome 13 were non-informative as only one peak corresponding to the mother's genotype was present. The third case had an elevated risk for trisomy 18 with NIPT. QF-PCR results were inconclusive because of maternal cell contamination. Karyotype analysis was subsequently performed on the amniotic fluid and a mosaic karyotype including a cell line with an isochromosome 18p was detected. Fluorescent in situ hybridization (FISH) studies later confirmed the results.

These few cases demonstrated that despite the high accuracy of these tests, discrepancies can occur due to technical and/or biological limitations. A standardized and comprehensive testing algorithm is therefore needed to ensure that the appropriate testing takes place.

Despite a universal healthcare system in Canada, disparities in prenatal testing algorithms are apparent. National consensus on how the different genetic tests should be utilized in the prenatal setting is needed.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2079. DNA damage and antioxidant capacity as prenatal markers of placental epigenetic aging clock.

Authors:

F. Tekola-Ayele¹, C. Zhang², R. Biedrzycki³, K. Grantz¹, J. Grewal¹; ¹NIH, Bethesda, MD, ²Natl. Univ. of Singapore, Queenstown, Singapore, ³Glotech, Inc, Bethesda, MD

Abstract Body:

Accelerated aging of the placenta determined by epigenetic clocks in delivered placenta samples has been linked to obstetric and fetal complications such as preterm birth, fetal growth restriction, and preeclampsia. Therefore, identifying non-invasive prenatal markers of placental epigenetic aging holds a potential for developing diagnostics before onset of adverse clinical outcomes, but such data are scarce. The goal of the present study was to examine whether longitudinally measured 8-hydroxydeoxyguanosine (8-OHdG; a DNA damage marker) and total antioxidant capacity (TAC; a measure of cumulative antioxidant defense) in maternal plasma during pregnancy are associated with epigenetic aging of placenta. Maternal blood samples were collected during pregnancy at four consecutive research visits (10-14, 15-26, 23-31, and 33-39 gestational weeks) as part of the NICHD Fetal Growth Studies. 8-OHdG and TAC were measured in plasma (n=301), and the ratio of 8-OHdG and TAC was computed to estimate the degree of DNA damage relative to total antioxidant capacity. Epigenetic age of delivered placenta was estimated using previously developed clocks that use 62 and 396 methylation CpG sites (Clock62 and Clock396). Placental age acceleration was defined to be the residual upon regressing estimated epigenetic age on gestational age at delivery. The association of 8-OHdG, TAC, and 8-OHdG/TAC ratio with placental age acceleration was tested using covariate-adjusted linear regression models, with interaction terms between each biomarker and maternal age and between each biomarker and gestational week at blood draw. As gestation advanced, 8-OHdG showed a rise (18% rise in mean 8-OHdG from 89.6 to 105.5 nM between 10 to 40 gestation weeks), whereas TAC declined (6% decline from 89.5 to 83.7 nM). Higher 8-OHdG/TAC ratio was associated with increased placental age acceleration (Clock62 adjusted- β (S.E.) = -0.55 (0.29) per each SD increase in 8-OHdG/TAC ratio; P=0.049), whereas higher TAC was associated with lower placental age acceleration (Clock62 adjusted- β (S.E.) = -0.61 (0.25) per each SD increase in TAC; P=0.01). These data suggest that circulating levels of maternal DNA damage that outweigh antioxidant buffer during pregnancy may be one mechanism related to accelerated aging of placenta. The potential value of these biomarkers as a non-invasive monitoring tool for early detection of placental epigenetic aging and pregnancy complications warrants future studies.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2080. Down Syndrome therapeutics: The integrated stress response and restoration of protein synthesis.

Authors:

J. Korenberg, M. Gandelman; Univ. of Utah, Salt Lake City, UT

Abstract Body:

Neuronal protein synthesis is essential for learning and memory, and its attenuation caused by the activation of the integrated stress response (ISR), triggered by phosphorylation of eIF2 α has been found to have a role in DS and Alzheimer's disease, among others. Downstream increase of the transcription factor ATF4 is the main effector of the ISR. ATF4 then coordinates a transcriptional program that promotes cell survival during acute stress, although it has the potential of becoming maladaptive under severe or chronic stress conditions. Initial reports have found activation of the ISR in the hippocampus of Ts65Dn mice and in post-mortem brain tissues from patients with DS but the causes, consequences and druggability of the ISR in DS remain largely unexplored.

We utilized skin fibroblasts derived from identical twins discordant for DS to study activation of the ISR and protein synthesis rates. We found elevated levels of phosphorylated eIF2 α in DS fibroblasts, associated with downstream activation of ATF4 and a decrease in protein synthesis rates. Because DYRK1a is located in a chromosome region associated with deficits of learning and memory and has a role in translational control, we investigated whether its knockdown would prevent activation of the ISR and restore protein synthesis. We found that DYRK1a/dyrk1a levels were increased in DS fibroblasts as expected, and its knockdown decreased p-eif2 α and ATF4 levels, but protein synthesis was not restored. We next utilized ISRIB and trazadone, known modulators of the ISR, and found that both successfully increased protein synthesis rates in DS fibroblasts, indicating these drugs hold potential for reversing alterations at the cellular level that can lead to cognitive and memory deficits. In conclusion skin fibroblasts from twins discordant for DS recapitulate DS brain ISR activation and decreased protein synthesis that is not restored by rescue of DYRK1a but is restored by ISR inhibitors. The results also provide a cellular druggable mechanism for the multiple effects of Dyrk1a in DS. Finally, this establishes individuals' cell lines as a valuable model for studying individuals' ISR, for testing interventions, and for personalized treatment to reverse alterations at the cellular level that can lead to cognitive and memory deficits in DS.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2081. Explore SOX9 haploinsufficiency in neural stem cells.

Authors:

S. Chan, Z. Wang, K. Miu, H. Cheung, W-Y. Chan; The Chinese Univ. of Hong Kong, Hong Kong, Hong Kong

Abstract Body:

Campomelic dysplasia (CD) is a rare congenital disease caused by *SOX9* haploinsufficiency and only 5-10% of patients survived. Complications of CD patients include sex reversal, abnormal development of the skeletal system, and mild to moderate learning difficulties. *SOX9* is a relatively well-known gene in skeletal and gonadal development and has recently been associated with cancer as a metastasis marker, it also induces neural stem cells (NSC) and mediates neurogenic-to-gliogenic cell fate switch in the central nervous system (CNS). However, the underlying mechanism of *SOX9*'s role in the NSC induction remains unclear.

With CRISPR/Cas9 techniques, a *SOX9* edited hiPS cell line was generated according to the medical literature of a CD patient. The cells were differentiated into NSC. qPCR result suggested that loss of *SOX9* altered the expressions of key genes, such as *SOX1*, *SOX2*, *Nestin*, and *PAX6*, in iPSC-derived NSC. Further analysis was performed by stranded-specific RNA sequencing. Based on the transcriptome level, several pathways such as MAPK, p53, and Wnt signaling pathways were found altered in NSC. Further analysis including western blot, immunostaining, and qPCR analysis will be performed to validate RNA sequencing results. NSC functions, such as neurogenesis and gliogenesis, and conductivity of the resulting neurons will also be tested by continuous differentiation and high-density microelectrode array analysis. Collectively, the current study aims to identify the *SOX9*'s role in NSC and the subsequent effect in the CNS to explain the learning difficulties reported in CD patients.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2082. Exploring roles of distal enhancers in *SOX9* haploinsufficiency in sex determination

Authors:

F. Ou¹, H-C. Suen¹, H. Sun², Z. Wang¹, K. Miu¹, P. Shi², J. Liao¹, W-Y. Chan¹; ¹The Chinese Univ. of Hong Kong, Hong Kong, China, ²City Univ. of Hong Kong, Hong Kong, China

Abstract Body:

Disorders of sex development (DSD) are congenital anomalies involving discordance between genotype to phenotype in gender identity. People with DSD suffer from ambiguous genitalia, impaired steroid hormone production, reduced or null fertility, and partial to complete sex reversal. Various mutations of the dosage-sensitive haploinsufficient *SOX9* gene (sex-determining region Y-related high-mobility group box 9) were identified in male to female sex reversal patients. *SOX9* regulates the production of the anti-Müllerian hormone (AMH) in Sertoli cells to inhibit the creation of a female reproductive system in an XY embryo. A single copy of the wild-type allele of the heterozygous *SOX9* is insufficient to maintain normal male reproductive tract development. Duplication and deletion of distal enhancers of *SOX9* found in DSD patients suggest the potential roles of enhancers in *SOX9* haploinsufficiency in sex determination. But the underlying genetic mechanism of the haploinsufficiency is still largely unknown. With the advances in high throughput sequencing, gene editing technologies, stem cell, and spatial ncRNA assay, we aim to decipher the chromatin topological interaction complex at *Sox9* enhancers across the sexual developmental time window and elucidate the functional roles of the enhancer complex in *Sox9* haploinsufficiency. Interactions between enhancers, promoters, and cell type-specific transcription factors will be identified by CAPTURE. By integrating scRNA-Seq and scATAC-Seq data, we have profiled temporal chromatin landscapes across gonadal somatic cell types during sex development and identified 3 putative distal enhancers of *Sox9*. The dynamic expression of non-coding enhancer RNA (eRNA) that may involve in the topological interaction for *Sox9* transcription dosage regulation will be further evaluated with a newly developed spatial ncRNA assay on fresh mouse gonads. Co-localization of protein complexes and eRNA would be validated by conducting rISH-PLA. Human iPSC models with disruption to components of the enhancer complex will be differentiated to Sertoli cells for validation. Our project will provide a valuable prototype for studying haploinsufficiency from 3D chromatin topological interactions and non-coding RNA regulation.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2083. Fetal genetic autopsy: A five-year retrospective audit monocentric audit

Authors:

J. Martinovic^{1,2}, F. Gubana¹, A. Bonnin¹, C. Egoroff¹, F. Razavi¹, A. Receveur¹, J. Saada¹, A. Bazin³, D. Buzas³, E. Julien⁴, R. Stoeva⁴, J. Grevoul-Fesquet⁵, E. Azria⁶, M-V. Senat⁷, J. Melki⁸, A. Vivanti¹, A. Benachi¹, E. Leyne¹; ¹Antoine Beclere Univ. Hosp., AP-HP, Clamart, France, ²Inst. Natl. de la Santé et de la Recherche Médicale (Inserm), UMR-1195, Université Paris Saclay, Kremlin Bicêtre, France, ³René Dubos Hlth.Ctr., Pontoise, France, ⁴Le Mans Hlth.center, Le Mans, France, ⁵Sud Francilien Hlth.Ctr., Corbeil-Essonnes, France, ⁶Saint Joseph Hosp., Paris, France, ⁷Kremlin Bicêtre Univ. Hosp., AP-HP, Kremlin Bicêtre, France, ⁸Inserm and Univ. Paris Saclay, Kremlin Bicetre, France

Abstract Body:

Normal 0 21 false false false FR X-NONE X-NONE /* Style Definitions */ table.MsoNormalTable {mso-style-name:"Tableau Normal"; mso-tstyle-rowband-size:0; mso-tstyle-colband-size:0; mso-style-noshow:yes; mso-style-priority:99; mso-style-parent:""; mso-padding-alt:0cm 5.4pt 0cm 5.4pt; mso-para-margin-top:0cm; mso-para-margin-right:0cm; mso-para-margin-bottom:10.0pt; mso-para-margin-left:0cm; line-height:115%; mso-pagination:widow-orphan; font-size:11.0pt; font-family:"Calibri",sans-serif; mso-ascii-font-family:Calibri; mso-ascii-theme-font:minor-latin; mso-hansi-font-family:Calibri; mso-hansi-theme-font:minor-latin; mso-fareast-language:EN-US;} **Objectives** To assess the efficiency and the pertinence of oriented genetic fetopathological examinations at a tertiary maternity laboratory. We were particularly interested in the rate of new macroscopic and/or microscopic data brought up by fetal and placental examinations (re)orienting further molecular studies to come. Additionally, the feasibility and the results of genetic post-natal tests performed during fetopathological examinations. **Methods** We conducted a 5-year retrospective audit that included 1710 fetuses and their placentas. All of these cases were referred to the foetopathology laboratory Antoine Beclere, APHP, Paris Saclay University Hospital, between January 2013 and December 2017, including medical terminations (TOPs), stillbirths and miscarriages. **Results** In 51.8% of the reports the fetopathological examination added an important finding (ie considerably modifying a final diagnosis), which was missing on the prenatal follow-up. Among those, the highest level of the new data was observed in the deceased newborns: 69.7%, and 48.3% in TOPs. The non-systematic, hence examination-oriented postnatal karyotypes and aCGH sample requests resulted in 9.2% and 25.9% of pathological results respectively. The highest level of aneuploidies was observed in stillbirths (54.2%), while the aCGH was most frequently unbalanced in TOPs (83.3%). The global failure rate of karyotypes and aCGH performed after fetopathological examinations was respectively 31.7% and 15.1%. Furthermore, the predictive factors of these failures on fetal /placental tissue in immediate post-natal collection are discussed to enhance the yield of better results in the future. An accurate diagnosis following fetopathological examinations and genomic tests directed by the fetopathological team was established in 51.2% of the cases: 69.9% in the cohorte of miscarriages, 56.2% among TOPs, ad 42.2% in stillbirths.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2084. Gaucher disease: systematic review.

Authors:

G. Gómez¹, H. Giron Osorio¹, L. Moreno Giraldo²; ¹Univ. Santiago de Cali, palmira, Colombia, ²Univ. del Valle, cali, Colombia

Abstract Body:

Abstract: An updated bibliographic review on Gaucher disease is carried out, in which emphasis is placed on its recognition as an orphan disease for Colombia, due to its low prevalence, chronicity, high complexity, burden of morbidity and mortality and importance of early diagnosis in order to establish timely, targeted treatments that impact its natural history. Are addressed aspects Historical, epidemiological, clinical, genetic-hereditary, pathophysiological, diagnostic, therapeutic, follow-up guidelines, ethical and legal framework relevant to understanding of disease. In addition, the clinical spectrum of the pathology is thoroughly described, recognizing the multisystemic nature and the wide variability of phenotype, and possible correlations with the genotype, highlighting the importance of the implementation of public health strategies for the early identification of this pathology either through the implementation of neonatal screening, creation of diagnostic-therapeutic algorithms, routes of action, education on red flag signs and symptoms, as well as recognizing the importance of genetic counseling as a key prevention mechanism in the disease. **Methods and materials:** A systematic review was carried out. The bibliographic search was carried out in English and Spanish and included scientific reference articles on Gaucher disease dating from the year 2015-2022, from the databases: PUBMED, MEDLINE, SCIENCEDIRECT, GOOGLE SCHOLAR, where the search for keywords: Gaucher Disease, Sphingolipidosis, orphan disease, neonatal screening, lysosomal storage diseases, errores innatos del metabolismo Inborn Errors of metabolism. **Conclusion:** There are currently a large number of therapeutic alternatives available that, when used in the initial stages, could have repercussions on the natural history of the disease and on the patient's prognosis, hence the importance of knowledge of GD by health personnel, recognizing the character multisystemic disease and the wide variability of the phenotype, and possible correlations with the genotype, in addition to the importance of implementing public health strategies for the early identification of this pathology, therapeutic algorithms and care routes.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2085. Genetic Characterization of VACTERL Association-Like Congenital Malformations

Authors:

H. Pang, M. Hsieh, W. Shi, B. Sapkota, S. Li, D. DeMarzo, Y. Enchautegui-Colon, A. C. Tsai; Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK

Abstract Body:

VACTERL association is a nonrandom occurrence of multiple congenital abnormalities, including vertebral defects, anal atresia, cardiac defects, tracheal-esophageal fistula, renal and radial abnormalities, and limb malformations. Evidence has shown that genetic and epigenetic factors, teratogenic influences, and maternal medical conditions may contribute to VACTERL association. In particular, blastocyst developmental field defect and mutations in heterogeneous genes associated with early organogenesis (e.g. genes involving Sonic Hedgehog (SHH) and NOTCH signaling) may contribute to clinical features of VACTERL association. A few monogenic genetic disorders like Fanconi anemia and CHARGE syndrome also have overlapping features with VACTERL association; pathogenic variants in genes related to these conditions have been found in some patients. Our studies used conventional cytogenetics, chromosomal microarray analysis (CMA), and next-generation sequencing (NGS) to identify possible etiologies for 23 pediatric patients presenting with clinical features of VACTERL association. FISH studies, karyotypes, or both methods were also performed on 14 patients to identify chromosomal abnormalities, to which none were detected. In addition, CMA was performed on 19 patient samples, revealing changes in three patients— a 1q21.1 duplication in the first, 7q31.3 deletion in the second, and a 1.4 Mb 17p12 deletion in the third. The 17p12 deletion was consistent with the patient's clinical diagnosis of hereditary neuropathy with liability to pressure palsies. NGS (exome slicing and/or whole-exome sequencing) was also performed on all 23 patient samples. NGS detected pathogenic variants in the *SALL1*, *GDF1*, and *CHD7* genes in 3 (13%) patients. These findings indicated the presence of monogenic genetic disorders with similar features to VACTERL association. 14 (60%) patients possessed variants of uncertain significance (VUSs) in multiple genes, including *ZIC3*, *SALL4*, *RECQL4*, *CHD7*, *FGF8*, *HAAO*, *FANCD2*, and *FANCI*. Notably, identical 18-bp deletions in *FGF8* were found in two unrelated patients. Different VUSs in *ZIC3* and *FANCD2* were also detected in more than one patient. Exome slicing was ordered for 6 (26%) remaining patients who presented with no reportable sequence variants. Our results showed that multiple genes may play a role in developing VACTERL association. Genetic testing, especially NGS, should be routinely used to identify underlying causes in patients displaying features of VACTERL. Overall, our results expand the list of genes that may contribute to developing the clinical features of VACTERL.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2086*. Genetic deconvolution of fetal and maternal cell-free DNA in maternal plasma enables next generation non-invasive prenatal screening

Authors:

J. Zhang^{1,2}, J. Li^{2,3}, C. Xu¹, H-F. Huang¹; ¹Fudan Univ., Shanghai, China, ²BioBiggen, Beijing, China, ³Heristar, Houston, TX

Abstract Body:

Current non-invasive prenatal screening (NIPS) tests analyze circulating fetal cell-free DNA (cfDNA) in maternal peripheral blood for selected chromosomal abnormalities. Many genetic disorders are refractory to NIPS largely because the maternal genetic material constitutes most of the total cfDNA present in the maternal plasma which hinders the detection of fetus-specific genetic variants. The inability to reconcile different genetic cues in a single NIPS assay prevents its expansion for the concurrent screening of different types of genetic disorders and improvements on test performance. In this study, we developed an innovative sequencing method, termed Coordinative Allele-aware Target Enrichment Sequencing (COATE-seq), followed by multidimensional genomic analyses of sequencing read-depth, allelic fraction, and linked single nucleotide polymorphisms (SNPs), to accurately separate the fetal genome from maternal background. Analytical confounders including multiple gestations, maternal copy number variations and absence of heterozygosity were successfully recognized and precluded for fetal variant analyses. In addition, fetus-specific genomic characteristics, including the cfDNA fragment length, meiotic error origins, meiotic recombination, and the recombination breakpoints were identified which reinforced the fetal variant assessment. In 1,129 qualified pregnancies tested, 54 fetal aneuploidies, eight microdeletions/microduplications and eight monogenic variants were detected with 100% sensitivity and 99.3% specificity. Importantly, using the comprehensive cfDNA genomic analytical tools developed, we found that 60.3% of aneuploidy samples had aberrant meiotic recombination providing important insights into the mechanism underlying meiotic non-disjunctions. To the best of our knowledge, this is the first report on aneuploidies associated with aberrant homologous recombination identified by cfDNA studies. Altogether, we show that the genetic deconvolution of the fetal and maternal cfDNA enables thorough and accurate delineation of fetal genome which paves the way for the next-generation prenatal screening of essentially all types of human genetic disorders.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2087. Genetic determinants of gestational length in Indian women: A two-stage genome-wide association study.

Authors:

E. Bhattacharjee¹, *. Ayushi², GARBH-Ini study group, R. Thiruvengadam^{2,3}, S. Bhatnagar², P. P. Majumder¹, A. Maitra¹; ¹Natl. Inst. of BioMed. Genomics, Kalyani, India, ²Translational Hlth.Sci. and Technology Inst., Faridabad, India, ³Pondicherry Inst. of Med. Sci., Puducherry, India

Abstract Body:

Background: Preterm birth (PTB), a major public-health burden globally, is associated with high neonatal and infant mortality and long-term health problems. LICs and LMICs contribute more than 60% of PTBs annually; India accounts for 23.4% of PTBs. Although genomic factors enhance PTB risk, estimates of genomic risk factors in Indian and other South Asian women are unavailable. **Hypothesis:** Maternal genotypes at specific loci are associated with gestational length (GL), the determinant of PTB.

Methods: We genotyped 6211 women enrolled in a prospective pregnancy cohort, GARBH-Ini, developed by us in Haryana, India, using Global Screening Array v3.0 with multi-disease drop-in panel (0.7M loci). The study participants, each with spontaneous singleton live birth, after quality control, were randomly grouped into two equal subgroups for discovery and validation. After adjusting the GL (weeks) for maternal age, BMI, parity and occupation, genotype association testing was performed using linear regression. SNPs found to be significantly associated in discovery subgroup were retested for association in the validation subgroup.

Results: Three SNPs were discovered to be significantly associated with GL, of which, rs67151250, an intronic SNP in *ARHGAP26* in chromosome 5, was validated. Population substructure was not detected ($\lambda = 1.02$). Phasing and imputation of chromosome 5 using data of 1000 Genomes Phase 3 and Genome Asia Pilot project as reference, led to the discovery and validation of additional 6 intronic SNPs in *ARHGAP26*. Joint analysis revealed, rs67151250 and the 6 imputed SNPs were associated with GL at the genome-wide significance level ($p < 5e-8$). TT genotype at rs67151250 lowered GL compared to CC and CT (Kruskal-Wallis $p = 0.0057$; Wilcoxon p : TT vs CC = 0.0043, TT vs CT = 0.0021, CC vs CT = 0.16), which we further confirmed by testing under a recessive model ($p = 1.7e-8$, Beta = -2.09, 95% CI = -2.81 to -1.36). Moreover, logistic regression under recessive model, revealed association of the T allele at rs67151250 with PTB ($p = 2.8e-6$, OR = 6.76, 95% CI = 3.04 - 15.03). *ARHGAP26* is a Rho-GTPase that inhibits RHOA. GTP-bound RHOA was previously reported to be significantly increased in the myometrium of preterm delivering women compared to those delivering at term and possibly results in early activation of myometrial contraction.

Conclusion: This first report on genetic determinants of gestational length from South Asia, identified novel maternal *ARHGAP26* variants that provided new insights into susceptibility to PTB. This may help in the identification of women at-risk for PTB and implementation of measures to prevent PTB, thus reducing public health burden.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2088*. Genome-wide analysis of preeclampsia identifies novel loci and loci previously associated with blood pressure traits.

Authors:

C. Greene¹, E. A. Jasper², J. Hellwege², S. S. Verma³, B. Xiao³, G. P. Jarvik⁴, I. B. Stanaway⁵, K. A. Leppig⁶, G. Hayes⁷, O. Dikilitas⁸, I. J. Kullo⁸, A. Fedotov⁹, J. B. Harley¹⁰, M. Lee¹¹, Y. Zhang^{12,11}, B. Namjou-Khales¹³, I. Holm¹⁴, T. Edwards², D. R. Velez-Edwards²; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Univ. of Pennsylvania, Philadelphia, PA, ⁴Univ Washington Med Ctr., Seattle, WA, ⁵Univ. of Washington, Seattle, WA, ⁶Kaiser Permanente of Washington, Seattle, WA, ⁷Northwestern Univ., Chicago, IL, ⁸Mayo Clinic, Rochester, MN, ⁹Columbia Univ., New York, NY, ¹⁰US Dept. of VA, Cincinnati, OH, ¹¹Geisinger Genomic Med. Inst., Danville, PA, ¹²Princeton Univ., Princeton, NJ, ¹³CCHMC, Cincinnati, OH, ¹⁴Boston Children's Hosp., Boston, MA

Abstract Body:

Preeclampsia (PE) is a hypertensive disorder of pregnancy, characterized by new-onset high blood pressure, proteinuria, or signs of organ damage occurring after 20 weeks of pregnancy. This condition affects approximately 5% to 7% of pregnant women worldwide and is a leading cause of maternal morbidity and mortality. PE is also associated with subsequent adverse health outcomes, including long-term increased risk of cardiovascular disease. Greater understanding of genetic susceptibility to PE is needed. To this end, we conducted a cross-ancestry meta-analysis of new and publicly available PE genome-wide association study (GWAS) summary statistics. We incorporated previously published summary statistics (from BioBankJapan, UKBiobank, and FinnGen [Sakaue et al]) in addition to unpublished data from two electronic health record biobanks, BioVU and eMERGE. In total, we utilized data from 355,858 individuals (3,398 cases and 352,460 controls). We conducted a fixed effects meta-analysis with genomic control and a minor allele filter (MAF > 0.01), consistent across datasets. Prior to filtering, 563 of 30,088,180 SNP associations were initially statistically significant ($p < 5 \times 10^{-08}$). After, 15 of 20,500,321 SNP associations were statistically significant ($p < 5 \times 10^{-08}$). Of these 15 SNPs, 8 were mapped to distinct genes in dbSNP, some of which had been previously associated with related phenotypes such as pulse pressure and cardiovascular disease. The lowest p-value SNP was rs74958360, located in *GLYAT* ($p = 6.23 \times 10^{-10}$). An additional significant SNP was rs6748519, located in an intergenic region approximately 5 kb upstream of *QPCT* ($p = 2.93 \times 10^{-08}$). This tag SNP notably represents a region on chromosome 2 displaying linkage disequilibrium, presenting an opportunity for further association mapping. Enrichment analysis of loci achieving suggestive significance ($p < 5 \times 10^{-05}$, 1652 loci representing 544 unique genes) revealed an over-representation of ontology terms relating to the molecular function “calcium ion binding” (2.68-fold enrichment, $p = 3.64 \times 10^{-04}$) and to the cellular component “perikaryon” (5.18-fold enrichment, $p = 9.40 \times 10^{-04}$), among others. Through this work we have identified several SNPs warranting further investigation, adding to a growing body of knowledge surrounding the genetic risk of PE.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2089. Heteroplasmy dynamics and segregation in pre-implantation embryos: Rare insights from complete mitochondrial genome sequencing of clinical trophectoderm biopsy

Authors:

A. Aggarwal¹, A. Novoselska², R. Antes¹, C. Librach², S. Madjunkova²; ¹CReATe Fertility Ctr., Toronto, ON, Canada, ²Univ. of Toronto, CReATe Fertility Ctr., Toronto, ON, Canada

Abstract Body:

Objective: Heteroplasmy (Het) is cellular coexistence of mutated and normal mtDNA copies. Knowledge of maternal transmission of Het through early embryo development and its effects on human embryo is limited, due to restricted availability of human embryos for research and recent advancements in technology for low DNA input genomic analysis. Whole genome amplification (WGA) is integral to preimplantation genetic testing (PGT); however, it may introduce sequence errors and coverage bias, leading to false detection of low-level Mt-Het. We optimized analysis of human embryonic mtDNA from a trophectoderm (TE) biopsy and determined a threshold for detection of Het. Our objective here was to study patterns of heteroplasmic variation and transmission at the embryonic and fetal stage of early development employing the developed pipeline. **Materials & Methods:** We studied 43 samples from 11 cases, analyzing 20 maternal to embryo and 12 maternal to fetal transmissions. Each case included maternal DNA (blood), embryonic DNA (WGA from TE biopsy of blastocysts undergoing PGT), and fetal DNA of miscarried embryos (product of conception tissue). Illumina DNA Prep with Enrichment using whole MtDNA probe capture by ligation (Twist Biosciences) and Next Seq 550 was used for data generation. Optimized Het detection threshold of 5% was applied for variant analysis. **Results:** Mt-genome was sequenced at mean depth ~4300x (100% at 50x). 951 variants were identified at 249 sites. From 138 unique sites, 35 segregated in multiple cases which included haplogroup markers and ancestral variants. Based on gene length, D-loop and protein coding regions had more variants than expected ($p < 0.01$). 112 heteroplasmies (<80%, >5%) at 62 sites were identified. Changes in allele frequency between maternal and fetal tissues followed a normal distribution with mean=0. A weaker correlation of variant allele frequency between sibling embryos ($r=0.78$) than between a mother and fetus ($r=0.81$) was identified. We found a germline bottleneck of 18 Mt-units. We identified two variants: c.4216T>C and c.15928G>A, associated with repeated miscarriage on MitoMap. **Conclusions:** Our optimized pipeline allowed detection of low Het variants (>5%) in human embryonic mtDNA. High variant frequency present in the embryonic mt-genome is transmitted through a severe germline bottleneck of 18 Mt-units. The correlation data is suggestive of a potential germline bottleneck after the divergence of oocyte lineages. We are expanding the analysis on a larger sample size (~500 embryos). These findings may have key implications for understanding variants in early embryos and their significance in human embryo development.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2090. High-risk *APOLI* genotypes and infant growth restriction associations in Black women affected by preeclampsia.

Authors:

R. Durodoye¹, T. Azhibekov², A. Miller¹, R. Davis³, L. Bruggeman⁴, S. Williams¹; ¹Case Western Reserve Univ., Cleveland, OH, ²Metro Hlth.Med. Ctr., Cleveland, OH, ³Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, ⁴Cleveland Clinic, Cleveland, OH

Abstract Body:

Background

Preeclampsia (PE) is a multisystem hypertensive disorder of pregnancy contributing to disproportionately higher maternal mortality rates in Black women in the United States. Evidence indicates that risk variants for the gene Apolipoprotein L1 (*APOLI*) (termed G1 and G2) associate with elevated PE risk, a higher prevalence of preterm birth, and low birth weight in specific infant populations carrying high-risk (two *APOLI* risk variants) genotypes. These two variants are found almost exclusively in populations of sub-Saharan African descent. The mode of inheritance and associations between these variants and clinical outcomes remain unclear and understudied in this population.

Methods

We harmonized data from the University of Tennessee CANDLE database and the Ohio March of Dimes database to conduct an unmatched case-control study of 1,361 Black women consisting of PE cases (both term and preterm) and 893 term uncomplicated births as controls. This analysis was designed to (1) clarify the inheritance model for *APOLI*-associated PE risk using logistic regression modeling and (2) determine how *APOLI* genotypes affect infant health outcomes.

Results

After adjustments for study location, both the recessive (OR 1.8, 95% CI: 1.1, 2.8) and additive (OR 1.3, 95% CI: 1.0, 1.7) inheritance models associated significantly with PE risk, but only in cases with preterm delivery. PE term cases showed significantly lower birth weights compared to controls ($p < 0.0001$). All control infants (low- and high-risk genotypes), and low-risk genotype term cases showed positive associations between birth weight and gestational age (all $p < 0.0001$), but there was no association observed in high-risk genotype term PE cases. Additionally, term cases with high-risk genotypes were over three times more likely to be small for gestational age compared to term cases with low-risk genotypes (OR 3.5, 95% CI: 1.2, 10.3).

Conclusion

These findings indicate that high-risk *APOLI* genotypes contribute to phenotypes of placental insufficiency and altered fetal growth. Our data best fit a recessive model of PE risk, although the additive model was also significant. These findings underscore the heterogeneity of PE phenotype and the importance of *APOLI* risk variants in conditions that exhibit higher morbidity and mortality in Black women in the United States.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2091. Host genomics and human milk exposure are associated with sex-specific changes in the gut microbiota of infants with asthma in the CHILD Cohort Study.

Authors:

S. A. Stickley¹, Z. Y. Fang¹, A. Ambalavanan¹, Y. Zhang², C. Petersen^{3,4}, D. Dai^{3,4}, B. Robertson⁵, C. Yonemitsu⁵, P. J. Mandhane⁶, E. Simons⁷, T. J. Moraes⁸, M. R. Sears⁹, S. E. Turvey^{3,4}, P. Subbarao⁸, L. Bode⁵, M. B. Azad^{7,10}, Q. Duan^{1,2}; ¹Dept. of BioMed. and Molecular Sci., Queen's Univ., Kingston, ON, Canada, ²Sch. of Computing, Queen's Univ., Kingston, ON, Canada, ³Div. of Allergy and Immunology, Dept. of Pediatrics, Univ. of British Columbia, Vancouver, BC, Canada, ⁴Dept. of Pediatrics, Child and Family Res. Inst. and British Columbia Children's Hosp., Vancouver, BC, Canada, ⁵Dept. of Pediatrics and Larsson-Rosenquist Fndn. Mother-Milk-Infant Ctr. of Res. Excellence (MOMI CORE), Univ. of California San Diego, La Jolla, CA, ⁶Dept. of Pediatrics, Univ. of Alberta, Edmonton, AB, Canada, ⁷ Dept. of Pediatrics and Child Hlth., Univ. of Manitoba, Winnipeg, MB, Canada, ⁸ Dept. of Pediatrics, Hosp. for Sick Children and Univ. of Toronto, Toronto, ON, Canada, ⁹Dept. of Med., McMaster Univ., Hamilton, ON, Canada, ¹⁰Manitoba Interdisciplinary Lactation Ctr. (MILC), Children's Hosp. Res. Inst. of Manitoba, Winnipeg, MB, Canada

Abstract Body:

Asthma is the most prevalent chronic disease among children and a leading cause of hospitalizations and school absenteeism, with higher prevalence in males than females during childhood. Although numerous risk factors have been reported for this multifactorial condition, the mechanisms underlying the sex-specific differences in asthma remain poorly understood. Previous studies suggest that the early-life gut microbiota may play a crucial role in asthma pathogenesis through bidirectional crosstalk along the gut-lung axis, however, more research is needed to investigate the role of sex differences. In this study, we integrate host genomics and early-life exposures, such as breastfeeding practices and human milk composition, to study sex-specific effects on the gut microbiota and susceptibility to asthma and recurrent wheeze. Specifically, we leverage existing datasets from the CHILD Cohort Study, including gut microbiota sequenced using 16S rRNA from stool samples collected at two ages: 3 months (N=747) and 1 year (N=766). We also utilize genomic profiles obtained from the Illumina HumanCoreExome Bead Chip, breastfeeding practices reported by repeated questionnaires, and quantifications of the 19 most abundant human milk oligosaccharides (HMOs) from breastmilk samples collected 3-4 months postpartum using high-performance liquid chromatography. Using generalized linear models, we identified that increased abundance of *Bifidobacterium longum* in males at 3 months was associated with decreased risk of recurrent wheeze (ages 2-5 years; P=0.013). *B. longum* was also found to be increased at 3 months in males who were breastfed at the hospital compared to those who were not (P=4.0×10⁻³). In addition, increased abundance of *Klebsiella oxytoca* at 1 year was associated with increased risk of asthma (diagnosed by age 5) in both males and females (P=0.026, 0.045). Our genome-wide association study of *K. oxytoca* at 1 year identified associations with single nucleotide polymorphisms on Chromosome 20 in females only (rs12481476, P=2.4×10⁻⁹). Furthermore, *K. oxytoca* abundance was increased in males with high genetic risk of recurrent wheeze when they were exposed to increased concentrations of the HMO fucosyllacto-N-hexaose in human milk (P=0.036). Our findings suggest that host genomics and exposure to human milk impact the infant gut microbiota, which in turn may influence asthma and wheeze susceptibility in a sex-specific manner. A better understanding of the sex-specific differences in gut dysbiosis and asthma may help identify early-life biomarkers, preventative strategies, and therapeutic targets.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2092. Human engineered heart tissues uncover *MYH7* alleles associated with a frail sarcomere and ventricular noncompaction

Authors:

T. Monroe, D. Fullenkamp, M. Roy-Puckelwartz, E. McNally, L. Castillo; Northwestern Univ., Chicago, IL

Abstract Body:

Background: Pleiotropy, wherein genetic variation in the same gene can cause multiple phenotypes, is a feature of some inherited cardiomyopathies. *MYH7* variants most commonly associate with hypertrophic cardiomyopathy (HCM) and are less commonly found in dilated cardiomyopathy (DCM). Most recently, studies from large human cohorts as well as individual families indicate *MYH7* truncations and missense variants disrupting the first 600 amino acids of the beta myosin heavy chain head also associate with left ventricular noncompaction cardiomyopathy (LVNC). During heart development, the left ventricle normally loses its trabecular invaginations due to proliferation of the compact cardiomyocyte (CM) layer and concomitant quiescence of the trabecular layer. Deficient compact CM proliferation or excessive trabecular CM proliferation converge on a similar presentation, collectively referred to as noncompaction. We hypothesized that reduced contractility from specific *MYH7* head domain missense or truncating variants disrupts cardiomyocyte terminal maturation, leading to excessive trabeculation.

Results: We identified N-terminal variant carriers who had LVNC often in the setting of dilated cardiomyopathy, similar to what has been described with *MYH7* truncations. We engineered *MYH7* truncations in human induced pluripotent stem cells (hiPSCs) which were differentiated into cardiomyocytes and engineered heart tissues (EHTs) to model the developing heart. In EHTs with *MYH7* truncations contractility was reduced by 50%, EHT size was increased by 17%, and CM proliferation was increased by 40% in EHTs heterozygous for *MYH7* truncations (N=6 each).

Conclusions: Heterozygous truncation of *MYH7* results in impaired differentiation, reduced contractility, and decreased compaction in engineered heart tissues, and these findings are consistent with phenotypes that produce LVNC. We propose a general model where diminished contractility delays sarcomere maturation, extending the proliferation time of trabecular cardiomyocytes and resulting in excessive trabeculae.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2093. Identifying genes associated with preeclampsia using whole exome sequencing data from participants in the Penn Medicine BioBank.

Authors:

B. Xiao¹, S. Lee², Y. Jung², S. Verma¹, Regeneron Genetics Center, Penn Medicine BioBank, M. D. Ritchie¹, D. Kim¹; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Abstract Body:

Preeclampsia is a complex hypertensive disorder that affects women during pregnancy and is a leading cause of maternal and infant mortality. The heritability of preeclampsia from the maternal side has been estimated to be around 35%, a percentage suggesting that genetic factors play key roles in preeclampsia risk. However, common variants do not predict preeclampsia risk well by themselves. There have been few studies examining the contribution of rare variants to preeclampsia risk in larger cohorts, and there are no widely accepted risk genes for preeclampsia. Thus, we used SAIGE-GENE+ to conduct a gene burden rare variant analysis for preeclampsia from whole exome sequencing data from 3006 participants (905 cases, 2101 controls) in the Penn Medicine BioBank (PMBB), using different binning strategies to define rare variants (varying minor allele frequency (MAF) threshold and annotating variants by functional importance). We also analyzed the effects of rare variants in European (342 cases, 900 controls) and African (490 cases, 963 controls) ancestry participants. Numerous genes were significantly ($p \leq 0.05$) associated with preeclampsia, although none were significant after Bonferroni correction. The top genes associated with preeclampsia varied depending on the binning strategy and the ancestry of the participants. Some genes found to be most associated with preeclampsia showed high gene expression in preeclampsia or pregnancy-related tissues from the Genotype-Tissue Expression project. For example, SPON2, which was associated with preeclampsia when using rare missense variants in European ancestry participants, was more highly expressed in tissues from the cervix and uterus. FADS3 was also associated with preeclampsia when using all rare variants with $MAF < 0.01$ in African ancestry participants and showed higher expression in tissues from arteries. Other genes, such as PON1 and MDH2, have been linked to preeclampsia in previous studies. To test the predictive power of rare variants for preeclampsia, we divided PMBB into training and validation datasets randomly multiple times. In the training dataset, we used rare predicted loss-of-function (pLOF) variants to identify genes most associated with preeclampsia. In the validation set, we generated polygenic rare variant scores by summing the number of top associated genes that contained pLOFs. Adding the rare variant score to a model predicting preeclampsia using only common variants (polygenic risk scores) increased the AUC from an average of 0.554 to 0.572. Future work will focus on replication of these results in other cohorts to better identify genes truly associated with preeclampsia.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2094. Improving rare conditions diagnostic rates by standardizing practice and offering preclinical testing.

Authors:

E. Delot^{1,2}, S. Bhattacharya¹, D. A. Spencer¹, H. Barseghyan¹, E. Vilain^{1,2}, The DSD-Translational Research Network (DSD-TRN); ¹Children's Natl. Res. and Innovation Campus, Washington, DC, ²George Washington Univ., Washington, DC

Abstract Body:

Background: Differences of Sex Development (DSD) are an umbrella of conditions with overlapping phenotypes and a multitude of demonstrated etiology. As a result accurate diagnosis is rarely achieved in clinical practice, resulting in less than optimal outcomes for patients. With currently 14 sites in the US, the DSD-Translational Research Network (DSD-TRN) seeks to standardize practice across providers in all clinical disciplines, with emphasis on early diagnosis prior to management decision-making. The DSD-TRN provides standardized forms to support the network's standard of care and CME-accredited clinical case conferences, documents team activity, and supports a registry and biobank for translational and genomics research. **Methods:** Analysis of registry data and clinic activity reports. Biobank samples were subjected to exome or genome sequencing or reanalysis (on the Franklin platform) and to optical genome mapping (OGM, Bionano platform). **Results:** At least 43 different etiologies were reported, with a diagnostic yield of 43%, achieved through either endocrine (9%) or genetic (26% sex chromosome aneuploidies, 9% autosomal deletions, 56% single gene) testing. Diagnostic yield varied greatly by clinical site. Undiagnosed biobank samples contributed to the delineation of the new genitourinary and/or brain malformation syndrome (GUBS, OMIM #618820). Short-read sequencing identified 23 new diagnoses, including 4 in very established DSD genes (*AR*, *SRD5A2*) missed by clinical panel testing, allowing disambiguation of conditions with similar phenotypes but very different management. Manual analysis of genome sequence BAM files revealed a compound heterozygous 6.5 kb single-exon deletion in *LHCGR* (complementing a missense variant), missed by automated algorithms. In a proof of concept project, OGM was able to diagnose 7 classic DSD chromosomal etiologies, including 45,X/46,XY mosaics. OGM also uncovered additional complexity not reported by clinical testing. For example a 9p terminal deletion identified by microarray was revealed to be a translocation, exchanging telomeric regions between 9p and 5p, with loss of material on both chromosomes. OGM was also able to distinguish between different sizes of Y chromosome translocations to the X chromosome in 46,XX testicular DSD. Resolution was however insufficient to disambiguate the rearrangements at the *CYP21A2* locus implicated in Congenital Adrenal Hypoplasia. **Conclusions:** Combining systematic short read sequencing and optical mapping facilitate the diagnostic of DSD, and needs to be accompanied by a prioritization of genetic testing at each clinical site.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2095. Investigating uptake and impact of genetic and genomic evaluation following a perinatal demise.

Authors:

E. D'Orazio¹, K. Hall¹, V. E. Kimonis^{1,2,3}, F. Quintero-Rivera^{3,1}; ¹Div. of Genetics and Genomic Med., Dept. of Pediatrics, UC Irvine, Irvine, CA, ²Dept. of Neurology, UC Irvine, Irvine, CA, ³Dept. of Pathology & Lab. Med., UC Irvine, Irvine, CA

Abstract Body:

Despite public health efforts to mitigate stillbirth and neonatal death over the 20th and 21st centuries, the rate of decline in perinatal death has plateaued. Genetic etiologies, especially those implicated in undiagnosed causes of perinatal death, are thought to contribute to this trend. Ample literature has investigated the diagnostic yield of genetic testing in the case of stillbirth and neonatal demise. However, little research has explored the frequency of involvement of trained genetics specialists, such as medical geneticists and genetic counselors, in perinatal death cases from prenatal ascertainment of anomalies to demise. The current study examined retrospective demographic and clinical data from 111 perinatal demise cases and their gestational parents associated with attendance and uptake of prenatal genetic counseling, post-delivery genetics consult, genetic/genomic testing, and autopsy investigation at a large university-affiliated medical center between November 1st, 2017, and December 1st, 2021. Furthermore, this study investigated the potential diagnostic yield of genetic testing in the presence and absence of genetics specialist involvement providing evaluation and testing recommendations. Finally, this study appraised the degree of patient education in genetic post-counseling documented by genetics specialists versus non-genetics specialists. Through univariate analysis, genetic specialist involvement in perinatal cases was found to be associated with significant increases in genetic testing uptake ($p=0.007$), abnormal genetic testing results ($p<0.001$), positive results and results of uncertain significance that have a potential to contribute to disease), and increased degree of documentation of patient education outcomes through genetic post-test counseling compared to those services rendered by non-genetics providers ($p<0.001$). The findings of this study underscore the importance of active integration of genetics healthcare professionals into the process of perinatal postmortem investigation and allocating the practice of genetics evaluation and genetic testing selection to healthcare professionals with relevant genetics training.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2096. Knowledge and attitude of pregnant women in the Kingdom of Saudi Arabia toward Noninvasive prenatal testing: A single center study

Authors:

M. Akiel; King Saud bin Abdulaziz for Hlth.Sci. (KSAU-HS), Riyadh, Saudi Arabia

Abstract Body:

Background: Non-invasive prenatal testing (NIPT) is a screening tool for chromosomal aneuploidies. Prior knowledge of NIPT is an inherent factor in the decision-making process. We assessed the knowledge and attitude of pregnant women related to prenatal testing with a particular focus on NIPT. **Methods:** A prospective cross-sectional study, using a culturally validated questionnaire, was conducted with 342 pregnant women of whom 74.9% consented for prenatal screening. Mean age and gestational weeks \pm standard deviation was 31 ± 5 and 26 ± 11 , respectively. **Results:** A positive/very positive attitude was observed to ultrasound, followed by FCT, NIPT and lastly to CVS. More than half of the participants (56.1%) had no previous knowledge of NIPT. A reaching significance association was detected between education and knowledge of NIPT. Significant association was detected between risk for aneuploidy and knowledge of NIPT. The majority (74%) indicated their willingness to perform the test. The effect and value of society on the pregnant women to make a decision regarding NIPT was negligible. **Conclusion:** The pregnant women in the current study displayed a lack of knowledge and awareness regarding prenatal screening, particularly the NIPT. We recommend that pregnant women receive adequate counseling regarding prenatal screening to increase their awareness and knowledge of prenatal testing, including NIPT.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2097. Large-scale genome-wide association study meta-analysis of Hyperemesis Gravidarum confirms the nausea and vomiting hormone gene *GDF15* is the greatest genetic risk factor and identifies additional risk loci

Authors:

M. Fejzo¹, P. Mullin¹, N. Pujol Gualdo², T. Laisk², E. Biobank Research Team², K. W. MacGibbon³, X. Wang¹, N. Mancuso¹; ¹Univ. of Southern California, Los Angeles, CA, ²Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, ³Hyperemesis Ed. and Res. Fndn., Clackamas, OR

Abstract Body:

Most women experience nausea and vomiting of pregnancy and 18% require medication. The most severe form, Hyperemesis Gravidarum (HG), occurs in 2% of pregnancies and is associated with weight-loss and undernutrition. HG accounts for ~400,000 emergency department visits in the US annually and can cause serious maternal morbidity, suicidal ideation and PTSD. HG is associated with increased risk of preterm birth, neurodevelopmental delay, and autism. It is highly heritable. Herein we report results from the largest ever genome-wide association study (GWAS) of HG, comprised of 7,197 cases and 178,953 controls. We performed a meta-analysis of 4 independent studies: a multi-ancestry whole-exome sequencing study of 926 HG cases and 660 controls from the US, a GWAS of 23andMe customers of European descent comparing 1,306 cases to 15,756 controls with no NVP, an Estonian Biobank GWAS of 3,536 cases and 32,113 controls, and a FinnGen GWAS of 1,429 cases with excessive vomiting during pregnancy and 130,424 controls. We identified genome-wide significant ($P < 5E^{-8}$) variants at *GDF15* (rs1058587, $P = 1.745E^{-26}$), *IGFBP7* (rs4865233, $P = 2.737E^{-12}$), and *PGR* (rs10895102, $P = 5.11E^{-10}$), replicating our previous studies of HG. Of note, all 3 genes are known to be activated in early pregnancy, consistent with a putative role in HG risk. FUMA analysis suggests possible enrichment in stomach. Genetic loci in hCG and its receptor, were not associated with HG in any of the 4 studies, nor the meta-analysis, providing no support for the historical hypothesis that the pregnancy hormone is the cause. Future research efforts should focus on the genes identified in this study, which implicates the nausea/vomiting hormone *GDF15*-the most significant locus in all 4 cohorts- as well as the placenta and appetite protein *IGFBP7* and the progesterone receptor *PGR*. Replication studies are needed to confirm whether additional novel variants in/near genes that were significant using a relaxed P -value threshold of 5×10^{-6} in individual cohorts (ie *GDF15* co-receptor gene *RET* in the Estonian GWAS and the thyroid stimulating hormone receptor gene *TSHR* in the Finnish GWAS) are valid associations and/or generalizable to other populations. Overall, this study contributes to our understanding of the biology of nausea and vomiting and may provide new treatment avenues. Of note, drugs targeting the *GDF15* pathway have shown great promise in mitigating weight loss, loss of appetite, and vomiting in animal models and are currently in clinical trials in cancer cachexia, a disease with similar symptoms to HG. The strong link to this pathway in HG suggests these drugs, if safe, may hold great promise for treating HG in the future.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2098. Lethal genes and Mendelian disorders

Authors:

P. Cacheiro, A. Retnakumar, D. Smedley; Queen Mary Univ. of London, London, United Kingdom

Abstract Body:

Essential genes can be defined as those required for organism development. Existing resources that collect and characterise these genes are based on either cell proliferation assays, viability in model organisms, intolerance to variation metrics or predicted lethal genes. Resources like the Online Mendelian Inheritance in Man (OMIM) and the Human Phenotype Ontology (HPO) document phenotypic annotations for Mendelian disorders, yet the evidence on prenatal/neonatal lethal phenotypes is either limited or not easy to retrieve.

Here we queried OMIM for terms related to lethality and classified all Mendelian genes into lethality categories according to the earliest age of death reported using HPO age of death terms, from prenatal death to death in adulthood and non-lethal genes. We also explored how these categories correlate with evidence on mouse viability from the International Mouse Phenotyping Consortium.

After manual curation and exclusion of ambiguous entries, we found that 56% (1955/3517) of OMIM genes are not associated with disorders with clinical records of lethality, 33% (1174/3517) are only associated to disorders with records of lethal phenotypes, and 11% (388/3517) are linked to both lethal and non-lethal phenotypes. With regards to lethality categories, 917 genes, that represent 59% of all lethal genes and 26% of all disease genes, have records of prenatal, neonatal or infant death (pre-infant-lethal) as opposed to post-infant-lethal, where the earliest reported age of death ranges from childhood to adulthood. By associated mode of inheritance, 76% (695/917) of pre-infant-lethal genes are autosomal recessive, compared to 62% (402/645) of post-infant-lethal genes and 52% (1012/1955) of non-lethal genes. When we explored data on viability for the corresponding mouse orthologs - assessed at weaning for the homozygous knockouts - 75% of the pre-infant-lethal genes are lethal/subviable in the mouse, being this percentage similar for the other two groups, 52% and 50% respectively. The former percentage is slightly higher (82%) if we exclude infant death records. Further investigation of the abnormal phenotypes identified in these viable lines could help explain this discrepancy.

Monogenic forms of this extreme phenotype are likely underrepresented in current disease databases. Given the increasing number of prenatal and molecular autopsy sequencing studies, an additional literature search for prenatal and perinatal lethal phenotypes identified 94 novel genes with candidate pathogenic variants. We are currently building a Lethality Gene Portal to catalogue and showcase human lethal genes and support fetal sequencing studies.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2099. Leveraging duplex DNA sequencing to investigate germline mutation rates in the context of sperm selection and paternal age.

Authors:

J. Kunisaki¹, S. Lulla¹, K. Aston¹, J. Hotaling², A. Quinlan¹; ¹Univ. of Utah, Salt Lake City, UT, ²Univ. of Utah Sch. of Med., Salt Lake City, UT

Abstract Body:

BACKGROUND: Germline de novo mutations drive human evolution and underlie human disease. Their rates, patterns, and underlying mechanisms are therefore essential to understand. Work from our group and others have genome sequenced blood from parent-child trios to find that the male contributes 80% of transmissible mutations with an average paternal germline mutation rate of $\sim 0.9 \times 10^{-8}$. However, pedigree analyses study mutations solely from sperm capable of facilitating normal child development: thus, the unbiased germline mutation rate is unknown. To overcome this survivorship bias, one must examine a bulk sperm population comprised of reproductively “fit” and “unfit” gametes which accumulate DNMs amidst continuous cell division of spermatogonial stem cells (SSCs).

METHODS: We hypothesize the germline mutation rate in bulk sperm would be greater than pedigree-derived estimates, reflecting the continuous division of SSCs in a man's lifetime. To test this, we used targeted TwinStrand Duplex Sequencing to analyze low-frequency mutations in bulk sperm across 93 genes involved in spermatogenesis and DNA repair. This technology distinguishes true mutations as those present on complementary DNA strands from errors that arise on single strands. The resulting per-base error rate falls from 1/1000 to 1/100,000,000. We applied TwinStrand Duplex Sequencing to cross-sectional (single timepoint, n=26) and longitudinal (multiple timepoints, n=9) bulk sperm samples from normozoospermic men (sperm concentration ≥ 30 M/mL of ejaculate) aged between 24 to 68 years.

RESULTS: Our sequencing of bulk sperm supports the hypothesis that pedigree estimates of male germline mutation rates suffer from a survivorship bias. We observe an order of magnitude higher mutation rate (1.2×10^{-7}) and an increase of 8 mutations/year in bulk sperm. We also measured elevated pathogenic mutation rates in paternal age effect genes ($p=0.0029$), including BRAF and KRAS, which drive spermatogenesis and autosomal dominant Mendelian diseases such as Noonan syndrome. More broadly, we observe a pathogenic mutation rate of 5.9×10^{-9} across all genes, with such mutations accumulating at a rate of 0.4 pathogenic mutations/year. Further, we find that an average of 6.25% of mutations are clonal (i.e., present in more than one SSC), and 7% of clonal mutations are pathogenic. Such clonal mutations confer a lifelong risk for the transmission of deleterious mutations.

CONCLUSIONS: The observed differences in germline mutation rates between bulk sperm and pedigree studies suggest positive and/or negative selective forces act upon human sperm during fertilization and/or embryogenesis.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2100*. Metabolomic analysis reveals altered omega-6 fatty acid processing as a marker of *FMRI* premutation carriers with FXPOI compared to premutation carriers without ovarian insufficiency

Authors:

E. Allen¹, K. Shelly¹, D. Nelson², P. Jin³; ¹Emory Univ., Atlanta, GA, ²Baylor Coll. of Med., Houston, TX, ³Emory Univ Sch Med, Atlanta, GA

Abstract Body:

Fragile X-associated Primary Ovarian Insufficiency (FXPOI) is the clinical diagnosis given to roughly 20% of women who carry an *FMRI* premutation (PM) allele. This disorder is caused by the expansion of a CGG-repeat tract in the 5' untranslated region (UTR) of the *FMRI* gene to 55-200 repeats. Studies of FXPOI are limited by the availability of primary tissue and sufficient patient numbers who meet the clinical criteria of elevated FSH levels and cessation of menses for 4-6 months prior to age 40. We performed non-targeted metabolomic profiling of patient plasma by LC/MS on two cohorts of women, the largest to date, a test cohort (40 FXPOI Cases, 34 PM Controls) and a validation cohort (22 FXPOI Cases, 58 PM Controls) to identify metabolic indices of FXPOI among *FMRI* premutation carriers. Differential abundance analysis using the R-package xmsPANDA and Metabolite Set Enrichment Analyses (MSEA) via Mummichog revealed altered abundance of compounds in omega-6 fatty acid (n-6 FA) metabolism and arachidonic acid (AA) formation between females with a FXPOI diagnosis compared to female PM carriers without POI across both cohorts. Pathways downstream of FA and arachidonate metabolism were also identified, including prostaglandin synthesis and formation of pro-inflammatory metabolites from AA. Evaluation of menopausal status in controls did not implicate the menopause transition as the driver of this enrichment at the pathway level. The menopausal state of the controls was an important determinant for the direction of change for individual metabolites in the pathway. When the post-menopausal cases were compared to pre-menopausal controls, n-6 FA processing was lower. Whereas, cases compared to post-menopausal controls had increased processing of compounds through the n-6 FA and AA pathways. Intriguingly, key enzymes required for the conversion of membrane phospholipids into free FA precursors and the prostaglandins generated downstream of AA were also seen to be altered in mouse ovaries expressing *FMRI* PM alleles. Together, these data suggest altered metabolism of n-6 fatty acids is consistent across human and mouse models of FXPOI and is likely to provide critical information about the mechanism of dysfunction in PM ovaries.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2101. Molecular dysregulation in *Sox2*-expressing pituitary stem cells lacking *Prop1*

Authors:

B. Masser, S. A. Camper, L. Y. M. Cheung; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Mutations in the pituitary-specific transcription factor Prophet of Pit-1 (*PROPI*) are the most common genetic etiology of combined pituitary hormone deficiency (CPHD). CPHD is associated with short stature, attributable to growth hormone deficiency and/or thyroid stimulating hormone deficiency, as well as hypothyroidism and infertility. Pathogenic lesions impair pituitary development and differentiation of endocrine cells. We performed single-cell RNA sequencing (scRNA-seq) on pituitary stem cells from juvenile P4 female *Prop1*-mutant and control mice and observed overexpression of several genes, such as Pou Domain, Class 3, Transcription Factor 4 (*Pou3f4*), SRY-Box Transcription Factor 21 (*Sox21*), and SPARC Related Modular Calcium Binding 2 (*Smoc2*). We further performed single-cell regulatory network inference and clustering (SCENIC), which predicted activation of the *Pou3f4* regulatory gene network in *Prop1*-mutant stem cells and both *Sox21* and *Smoc2* to be downstream targets of *Pou3f4*. Therefore, we sought to elucidate the role of *Pou3f4* during pituitary development through analysis of *Pou3f4*-deficient mice, and to determine the contributions of *Pou3f4* upregulation to pituitary disease by generating double-mutant mice lacking both *Prop1* and *Pou3f4*. Here, we have shown that *Pou3f4* has a specific developmental expression pattern but is not required for pituitary organogenesis and endocrine specification. Furthermore, the *Prop1*; *Pou3f4* double-mutant mice did not show a rescue of embryonic pituitary dysmorphology, endocrine specification defects, or overexpression of *Sox21* or *Smoc2*, despite the computational prediction of the regulatory network. Collectively, these analyses have taken the initial steps in understanding the role of *Pou3f4* in the complex genetic hierarchy of pituitary organogenesis and pathophysiology. Future studies will investigate the functional significance of other *Prop1*-dependent changes in stem cell gene expression.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2102. Molecular Epidemiological Status of Group B Streptococcus in Ile Ife South Western Nigeria

Authors:

J. Omololu-Aso¹, O. O. Omololu-Aso²; ¹Obafemi Awolowo Univ., Ile Ife, Nigeria, ²Univ. Coll. Hosp. (UCH), Ibadan, Nigeria

Abstract Body:

Studies in some sub-Saharan African countries like Zimbabwe, Malawi, Kenya and Gambia revealed that Group B *Streptococcus* (GBS) is emerging as the main cause of neonatal sepsis and meningitis. However, in Nigeria, information on GBS disease prevalence remains sparse. We sourced to isolate GBS from the rectovaginal and neonatal samples that were obtained from a tertiary hospital in a populated area of Osun state and give updated information on the antibiotic susceptibility patterns, using demographic and clinical parameters. One hundred and seventy samples were collected from consenting mothers and neonates from June 2016 to January 2017. Ninety-Eight (98) GBS isolates were recovered from the vaginal, and rectal of the pregnant woman at the point of labour and the Umbilical cord of the neonate within 24hrs of birth. cultures for the isolation and identification of Group B Streptococcus (GBS) were carried out using the CDC recommended microbiological methods. The Kirby-Bauer disk diffusion method was employed to determine the antibiogram of GBS isolates in accordance with the Clinical and Laboratory Standards Institute (CLSI). The presence of resistant genes was examined using PCR. The prevalence rate of GBS maternal and neonatal colonization was 29.4% and 20.6% respectively while 4% of the colonized neonates had nosocomial GBS colonization. There was no significant association between GBS colonization status and age ($p > 0.05$), parity ($p > 0.05$), obstetric risk factors ($p > 0.05$) and sex of neonate. There was no incidence of GBS infection observed. Resistance to augmenting (88.8%), ampicillin (60.2%), penicillin (47%), tetracycline (34.7%), ceftriaxone (19.4%), clindamycin (13.3%), vancomycin (10.2%) and erythromycin (7.1%) were observed. one of the 8 representatives of the multidrug-resistant isolates harboured tetM gene while other resistant genes examined were negative in all MDR isolates. A high prevalence of maternal and neonatal GBS colonization has been established among pregnant women and neonates in the study area. Nosocomial infection was implicated in GBS colonization among neonates. However further research is called for using larger sample sizes and multiple curve studies for adequate extrapolation into the general population.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2103. Next-Generation Sequencing based pre-implantation genetic Testing in Nepal

Authors:

S. Khanal, S. Thapa, G. Joshi, A. Kunwar, N. Thakur; Kathmandu Ctr. for Genomics and Res. Lab., Lalitpur, Nepal

Abstract Body:

Preimplantation genetic testing reduces the risk of having a pregnancy with an affected fetus significantly. Screening for aneuploidy (PGT-A) is widely used as an add-on to standard IVF treatment of infertile couples around the world, though its effectiveness is heavily debated. We present our experiences in introducing NGS-based screening of embryos for aneuploidy. We emphasized on successful commencement of the NGS based screening of embryos for chromosomal aneuploidies. Baseline for analysis was created using 10 healthy male individuals. 3 days embryos from IVF were subjected to low-coverage whole genome sequencing. The analysis of the sequencing data was performed using the Ion Reporter software. The overall results from 177 total embryos analysis revealed around 28% of the embryos were euploid. However, most of the embryos possessed aneuploidies. Moreover, the results revealed complex mosaicism (mosaic in >3 embryos) being the most common anomalies in those embryos. PGT-A based on NGS is a reliable and effective technology for routine embryo screening, especially for couples who have had several miscarriages or failed embryo transfers.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2104. *Nlrp2* is a Maternal Effect Gene Required for both Placental and Embryonic Development.

Authors:

M. Sharif, Z. Anvar, I. Chakchouk, I. Van den Veyver; Baylor Coll. of Med.; Jan and Dan Duncan Neurological Res. Inst. at Texas Children's Hosp., Houston, TX

Abstract Body:

Pregnancies of women with bi-allelic inactivating mutations in NLRP2 are affected by a multi-locus imprinting disorder MLID and liveborn offspring with imprinting disorders, such as Beckwith-Wiedemann Syndrome (BWS) in children of women carrying these inactivating mutations. NLRP2 is a member of oocyte subcortical maternal complex (SCMC), required for embryonic development. We previously showed that mice with maternal loss of *Nlrp2* are subfertile with early embryonic loss and few liveborn offspring, which can have developmental defects and altered DNA methylation at selected imprinted loci. Our overarching goal is to molecularly characterize the mechanisms by which maternal loss of *Nlrp2* affects development of the placenta and embryo and genetic imprinting. To investigate this, we set up 5-14 timed matings each for the following crosses: WT x WT, maternal KO x WT, and WT x paternal KO. We used a novel application of X-ray microcomputed tomography (microCT) for comprehensive placental, embryonic and yolk sac characterization at 3 $\mu\text{m}/\text{voxel}$. We studied E11.5 conceptuses ($n=20-26$ per cross) for gross morphological anomalies, yolk sac vascularization and examined the placentas focusing on maternal decidua, junctional zone, and labyrinth. We found that placentas from *Nlrp2*-KO dams were found to be significantly smaller, had irregular lacunae, extensions and infoldings of decidua, abnormal junctional zone, and compact labyrinth with variations in vasculature. The conceptuses from *Nlrp2*-KO dams showed increased resorption ($n=25$, $P=0.0179$), abnormal yolk sac vasculature ($n=26$, 40.2%), irregular amniotic fluid volume ($n=26$, $P=0.0491$) and craniofacial anomalies ($n=22$, 37.6%) along with congenital heart defects (CHD's) ($n=22$, 32.7%). In conclusion, we show that the adverse reproductive outcomes due to maternal loss of *Nlrp2* are the result of a combination of both embryonic and extraembryonic defects. Future directions include detailed phenotypic characterization along with genomic profiling of embryos and placentas from *Nlrp2*-KO dams to study implantation (E4.5 & E5.5) and earlier stages of placentation (E9.5). This will reveal novel mechanisms linking regulation of imprinting to implantation, placentation, and embryonic development and will improve our understanding of rare imprinting disorders.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2105. Non-invasive prenatal testing (NIPT): A reliable accurate prenatal non-invasive diagnosis setting in Nepal

Authors:

S. Thapa, N. Thakur, A. Jang Kunwar, S. Khanal, G. Joshi; Kathmandu Ctr. for Genomics and Res. Lab. (KCGRL), Lalitpur, Nepal

Abstract Body:

Because of its enhanced accuracy, non-invasive prenatal testing (NIPT) for fetal aneuploidies utilizing cell-free DNA (cfDNA) has become widely used in clinical practice. The core goal of cell-free based prenatal testing is to provide minimally invasive, clinically accurate screening for fetal chromosomal aneuploidies in the early stages of pregnancy. The purpose of this study was to establish a validated NIPT workflow for cell-free fetal DNA (cffDNA) sequencing from maternal plasma for the detection of particularly autosomal trisomies (involving chromosomes 13, 18 and 21) and sex chromosome on an ion semiconductor sequencing technology. A total of ninety-one standard samples from healthy pregnant provided by Yourgene Health; their cfDNA library was prepared, perform emulsion PCR on One Touch 2 machine, enrich ISPs on One Touch ES machine, then load ISPs on an Ion 540 chip for sequencing. The sequencing output data was analyzed by using the bioinformatics pipeline of Yourgene Health. Eighty-one samples were successfully validated out of a total of 91 samples. However, 3% of samples did not meet quality, and 3% of sample libraries had low reads and failed validation. The method included a fetal DNA fraction calculation for all samples as well as additional quality indicators. After the establishment of Validated workflow, following an assessment of the continuing pregnancy's data, NIPT was offered to 174 pregnant women after the informed consent form. Maternal BMI, maternal age, gestational age along with the latest copy of ultrasound sonography (USG) was received and noted. Pre-test counseling was provided to all participants and addressed the main testing procedures, their benefits, limitations, turnaround time, and interpretation of the results. From the total clinical samples, 18 samples showed an abnormal result: 13 trisomy 21 samples, three trisomy 18 and two samples with sex chromosome abnormalities were reported. One case had to discontinue pregnancy due to complications in later pregnancy but unrelated to the NIPT results. 12 samples were needed to repeat due to the low fetal fraction (LFF) of cell-free DNA and later reported normal after the resampling and repetition of the test. During a post-test counseling session, the findings were explained. This research aims to introduce and set up Nepal's first NGS laboratory for NIPT, utilizing the Ion Torrent technology to provide a complete in-house solution for pregnant women in Nepal. The established laboratory standards for testing and reporting, as well as educational resources and counseling approaches, will contribute in the acceptance of NIPT in society and clinical practice.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2106. Ovarian subfertility genetic variants in females with unexplained infertility.

Authors:

R. Roman, L. Chorich, Z. Hawkins, M. Dougherty, M. Diamond, L. Layman; Med. Coll. of Georgia at Augusta Univ., Augusta, GA

Abstract Body:

Objective: Unexplained infertility is a common infertility diagnosis in the US, with up to 30% of patients having no identified cause after a standard evaluation. We hypothesize that abnormal ovarian follicular development may be a cause of unexplained infertility. Therefore, we sought to evaluate the overall prevalence of variants associated with primary ovarian insufficiency (POI) and ovarian dysgenesis (OD) in females with well-characterized unexplained infertility. **Methods:** Females with well-characterized unexplained infertility were recruited for the Assessment of Multiple Intrauterine Gestations from Ovarian Stimulation (AMIGOS) trial. De-identified DNA samples from 200 females were analyzed by whole exome sequencing (WES) to identify variants in twenty-nine genes associated with POI and OD listed in the Online Mendelian Inheritance in Man (OMIM) database. High quality reads were filtered to identify the prevalence of likely pathogenic variants including frameshift, splice-site, and stop-gained variants. Variant of uncertain significance (VUS) prevalence was studied, including missense variants with a Combined Annotation Dependent Depletion (CADD) score ≥ 20 and in-frame insertions/deletions. Only variants with allele frequencies < 0.01 in the gnomAD database were included. **Results:** One hundred and ninety-seven samples were adequate for WES. Five autosomal dominant likely pathogenic variants (four *ERCC6*, one *NR5A1*) were present in six patients (3%). Eighteen autosomal dominant VUS were identified in eighteen patients (9.1%), including sixteen missense variants with high CADD scores (one *BNC1*, seven *ERCC6*, four *ESR2*, one *FIGLA*, two *FOXL2*, and one *NOBOX* variant), one twenty-one base pair deletion (*ESR2*), and one three base pair deletion (*FOXL2*). Two X-linked VUS (one *BMP15* and one *DIAPH2* variant) were identified in two patients (1%). There were six autosomal recessive likely pathogenic heterozygous variants (one *FANCM*, one *GDF9*, one *HFMI*, one *MCM8*, one *XRCC1*, and one *ZSWIM7*) in eight patients (4.1%). Fourteen patients (7.1%) had two variants in genes associated with POI and/or OD. One patient (0.5%) had three variants associated with POI and OD. **Conclusion:** We have identified heterozygous likely pathogenic variants and VUS associated with ovarian subfertility in females with unexplained infertility. Future *in vitro* functional genomic studies will be needed to characterize their pathogenicity and to better understand the role of ovarian subfertility genetic variants in unexplained infertility.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2108*. Phenome and genomewide recessive scan of coding variants reveals novel associations for female infertility

Authors:

S. Ruotsalainen¹, J. Karjalainen^{2,1}, M. Kurki^{2,1}, H. Heyne^{3,1}, E. Widen¹, H. Laivuori^{1,4}, M. Aavikko¹, E. Lahtela¹, FinnGen, A. Palotie^{1,2}, M. Daly^{1,2}; ¹Inst. for Molecular Med. Finland, FIMM, Helsinki, Finland, ²Broad Inst., Boston, MA, ³Hasso Plattner Inst., Digital Hlth.Ctr., Potsdam, Germany, ⁴Dept. of Obstetrics and Gynecology, Tampere Univ. Hosp., Tampere, Finland

Abstract Body:

Identifying Mendelian diseases with recessive inheritance is challenging as most cases are caused by compound heterozygosity. Bottleneck events, such as in the Finnish population, enrich specific homozygous variants to higher frequencies and thus facilitate identification of disease associations. In the FinnGen project (n=377 277), we performed a recessive GWAS scan for 68 818 coding variants with at least 2 homozygous carriers against 2 274 phenotypes. We identified total of 570 homozygous associations (recessive p-value < 5E-8 and smaller than corresponding additive p-value) for 300 phenotypes and 229 genetic variants in 188 genes. Recessive inheritance to some phenotypes has been reported in OMIM for 71 (38 per cent) of these genes. We observed multiple recessive associations (both previously reported and not-reported in OMIM) for hearing-loss and eye diseases, including heterophoria. As a novel finding, we identified a highly Finnish enriched (over 40-fold compared to non-Finnish Europeans) stop-gained variant rs144313315 in *TBPL2* gene to be associated with female infertility in recessive GWAS (p-value from recessive model=4.75E-10, p-value from additive model = 0.289). A homozygosity in this gene has been reported in infertile female mice with anovulation. Homozygote females for this variant had significantly (p=6.33E-11) less offspring (mean=0.22) than heterozygotes or wild-type homozygotes (mean=1.73). None of the homozygous women had conceived without treatment for infertility. We identified another recessive association for female infertility for a splice donor variant rs17368310 in *PKHDILI* to have a clear recessive association with female infertility (recessive p=1.30E-18, additive p=1.34E-12). This variant is in high LD (r²= 0.79) with intronic variant rs1422 in *EBAG9* (Estrogen Receptor Binding Site Associated Antigen 9), which we hypothesize to be the actual causal gene in this case based on its more relevant biological connection to infertility as an estrogen-responsive gene.

In conclusion, our biobank-scale recessive scan of coding variants demonstrates the benefit of recessive GWAS and PheWAS scans. We found both known and novel recessive associations across a wide spectrum of phenotypes, including hearing-loss, eye diseases and female infertility, that are missed by the standard additive GWAS model.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2109. Phenomic and genomic analyses of endometriosis in the All of Us Research Program.

Authors:

D. Schlueter¹, C. Zeng², S. Goleva¹, J. Keaton³, T. Ferrara², O. Stubblefield¹, A. Williams², J. Dai¹, T. Cassini⁴, H. Mo⁵, J. Denny⁶; ¹Natl. Human Genome Res. Inst., Bethesda, MD, ²NIH, Bethesda, MD, ³Med. Genomics and Med. Genetics, Clarksburg, MD, ⁴NIH, Chevy Chase, MD, ⁵NHGRI, Bethesda, MD, ⁶NIH, Kensington, MD

Abstract Body:

Endometriosis is a disease characterized by endometrium-like tissue formation outside of the uterus, peritoneal irritation and adhesions. Endometriosis has a prevalence of roughly 10% in women of reproductive age. We leveraged electronic health records (EHRs) and patient provided information (PPI) to elucidate genetic and clinical correlates with this disease in participants derived from the All of Us Research Program. We identified cases and controls for endometriosis with two approaches and compare results from each: first through electronic health records and second through survey-based self-reported personal medical history. In our first, EHR-based approach, cases had at least one instance of endometriosis-related diagnostic codes (N=3738) and controls had none of these codes and had no record of prior laparoscopic excision or ablation procedure in their EHR (N=172366). In our second approach, we defined cases and controls using self-reported medical history at enrollment. Cases reported a prior diagnosis of endometriosis (N=4150) while controls who completed the past medical history module (excluding skip responses) and did not affirm endometriosis (N=53827). We performed phenome-wide association studies (PheWAS) for each case/control definition. A total of 374 phenome-wide significant associations were identified from the EHR-based PheWAS compared to 42 identified from the survey-based PheWAS. GWAS of the EHR-defined phenotype using whole genome sequencing (WGS) data (N=1390 cases, 5560 controls) a novel genome-wide significant locus at chr21:38534959 (OR= 5.60, P = 2.82E-8). One novel genome-wide significant result was found for the PPI-based GWAS at location chr11:15116317 (OR=1.57, P=1.12E-8) (N=2197 cases, 8788 controls). Results from the two sets of analyses are compared between the two representations of endometriosis.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2110. Placenta fraction of maternal blood cell-free RNA associates with pre-eclampsia during pregnancy

Authors:

Q. Yan¹, J. Wang²; ¹Columbia Univ. Irving Med. Ctr., New York, NY, ²Univ. of Pittsburgh, Pittsburgh, PA

Abstract Body:

Of adverse pregnancy outcomes, pre-eclampsia is the second-largest driver of maternal mortality worldwide. Although pre-eclampsia is likely a multifactorial disease, it is believed that placenta is responsible for the disease progression. However, it is challenging to monitor the health of placenta during pregnancy. Recently, a few pioneering studies showed that plasma cell-free RNA (cfRNA) from maternal blood offers an opportunity to non-invasively study the status of placenta and predict pre-eclampsia before the onset of symptoms. In this study, we hypothesize that the fraction of cfRNA released from placenta can be numerically estimated, the placenta fraction is associated with pre-eclampsia, and placenta fraction can improve the prediction of pre-eclampsia. Thus, first, we performed a tissue deconvolution of cfRNA from maternal plasma samples (GSE192902, discovery n=112 and validation n=87) with gene expression from whole blood, liver, and spleen from GTEx, and placenta from Gong et al. (2021) as references. There are four time windows, ≤ 12 weeks, 13-20 weeks, ≥ 23 weeks of gestation, and post-partum. We observed that the majority of cfRNA is derived from whole blood (0.86 ± 0.13 , mean \pm SD). The results show that placenta fractions are significantly higher during pregnancy (0.14 ± 0.12) than post-partum (0.02 ± 0.05 , $P = 6.65 \times 10^{-16}$). However, there is no difference between the three prenatal time windows. Second, we tested the association between placenta fraction and the risk of pre-eclampsia using the discovery group. The results show that placenta fraction at the second time window (13-20 weeks) is significantly associated with pre-eclampsia ($P = 4.77 \times 10^{-3}$). Third, we performed differential expression analyses of pre-eclampsia, and we identified 465 differentially expressed genes (DEGs) at the second time window, which replicated the results in Moufarrej et al. (2022). Then, we conducted prediction analyses of pre-eclampsia using support vector machine (SVM) and random forest with 10-fold cross-validations in the discovery group. SVM shows better prediction accuracy (AUC=0.98). The prediction model was further validated in the independent validation group (AUC=0.77). These results are similar to those reported by Moufarrej et al. (2022). Moreover, we added placenta fraction as an additional predictor, but the prediction accuracy is not improved, because 195 DEGs are correlated with placenta fraction ($FDR < 0.05$) and are already included as predictors. Next, we will leverage the genotype information in the deconvolution of plasma cfRNA as future work.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2111. Preeclampsia risk is influenced by apolipoprotein L1 genotype and micronutrient deficiencies.

Authors:

W. Bruner¹, R. Davis¹, C. Simpson²; ¹Univ. of Tennessee Hlth.and Sci. Ctr., Memphis, TN, ²Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN

Abstract Body:

Background: Preeclampsia (PE) is a hypertensive disorder of pregnancy that is associated with micronutrient deficiencies. Fetuses carrying two APOL1 high risk (HR) variants is also associated with PE risk. We hypothesized a potential additive effect between HR APOL1 genotype status and nutritional deficiencies would place individuals at a higher risk of developing PE. We assessed PE-associated risk of micronutrient deficiencies based on fetal APOL1 genotype in pregnant African American women.

Methods: This was a case-control study using a subset of African American mother and infant pairs collected from the Conditions Affecting Neurocognitive Development and Learning in Early Childhood cohort in Memphis, TN. We performed multiple logistic regression to examine the association of preeclampsia with 2nd and 3rd trimester plasma folate and 25-hydroxy vitamin D concentrations. Concentrations were dichotomized into high or low categories. Folate deficiency was defined as a concentration less than 15 ng/mL. 25-hydroxy vitamin D deficiency was defined as a concentration less than 20 ng/mL. Analyses assessed whether maternal or fetal apolipoprotein genotype status modified the association between folate and 25-hydroxy vitamin D association and preeclampsia. The reference group included individuals with both high folate or and 25-hydroxy vitamin D concentrations and low-risk apolipoprotein genotype.

Results: An increased risk for developing PE was found with individuals having both fetal HR APOL1 and low plasma folate concentrations in the 2nd trimester (OR 3.27; 95% CI 1.28, 7.65). Pregnancies with low 25-hydroxy vitamin D concentrations in the 3rd trimester were at an increased risk for preeclampsia (odds ratio 2.10; 95% confidence interval 1.09-4.12; P-value, 0.03). Risk for preeclampsia was greatest among pregnancies with fetal high-risk genotype and low vitamin D levels in the 2nd trimester (odds ratio, 2.79; 95% confidence interval, 1.06-6.83; P-value, 0.03) and 3rd trimester (odds ratio 6.40; 95% confidence interval 2.07-19.18; P-value, <0.01).

Conclusions: Our findings are suggestive of a potential additive effect between low 25-hydroxy vitamin D and fetal HR APOL1 genotype during the 3rd trimester. Similar potential additive effect was found between low folate and fetal HR genotype in the 2nd trimester.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2112. Prenatal and postnatal changes in chromatin accessibility and gene expression in mouse gut development for understanding Hirschsprung disease (HSCR).

Authors:

H. Chubaryov¹, A. Chakravarti²; ¹NYU, New York, NY, ²NYU Sch. of Med., New York, NY

Abstract Body:

Purpose: HSCR arises from developmental defects in the neuronal and glial cells of the enteric nervous system. Its successful formation relies on the proper migration, differentiation, and proliferation of enteric neural crest cells during embryonic development, genes for which are perturbed in HSCR. Poor outcomes in 30-50% of surgically resected HSCR patients suggests the role of postnatal developmental factors in its progression as well. In this study, the mouse was used to identify genes and their regulatory elements which are critical to prenatal and postnatal gut development. **Methods:** Previously published bulk RNA-seq whole gut data at embryonic stages E10.5, E12.5, E14.5 were combined with recently generated postnatal distal colon data at P0, P7, P21 and adult mice. Bulk ATAC-seq data was also generated for P0, P7 and adult timepoints; other timepoints are under completion. **Results:** Using maSigPro, genes were hierarchically clustered throughout development by gene expression (FPKM) across seven timepoints and functionally annotated using Enrichr. The clusters containing major genes for HSCR, *Ret* and *Ednrb*, were linked to functions involving neuronal migration and nervous system development, as expected, while the *Eno3* cluster, the only HSCR gene to increase its expression across postnatal development, was associated with mitochondrial ATP synthesis. Next, we used DESeq2, to conduct differential gene expression analyses between postnatal timepoints. Between earlier postnatal timepoints, P0 and P7, there were relatively few differentially expressed genes: 68 down- and 130 up-regulated in P7 versus P0 ($FDR \leq 0.001$ & $\text{Log}_2FC \geq 2$). The greatest change occurred between the earliest and latest postnatal timepoints: 1,572 down- and 2,241 up-regulated in adult versus P0. These changes involved upregulation of genes in the immune system and a consistent downregulation of genes in neurogenesis, in postnatal development. However, gene expression of *Ret* remained high through P7. These expression trends were mirrored for chromatin accessibility: 137 intervals had less and 319 more accessibility in P7 compared to P0 ($FDR \leq 0.01$ & $FC > 2$). The greatest changes were between P0 and adult with 20,880 greater and 21,224 with lesser accessibility in P0. In *Ret* open chromatin accessibility peaks were consistent between P0-P7, with an early peak disappearing and a new nearby peak appearing at the adult stage. **Conclusion:** This work will identify and integrate both the genes and their regulatory elements responsible for temporal gut development, eventually answering whether genetic variation in these elements contribute to postnatal disease progression in HSCR.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2113. Prenatal cytogenomic microarray analysis (CMA) reference model.

Authors:

M. Ghochani, M. Sweredoski, A. Padeganeh, A. Hajianpour, P. Rezaie, M. Moradian, R. Ramjit; Kaiser Permanente, Los Angeles, CA

Abstract Body:

Prenatal cytogenomic microarray analysis (PCMA) is used for genomic characterization of germline abnormalities in prenatal samples. In conjunction with traditional karyotyping, PCMA allows for higher resolution analysis to detect more clinically relevant genetic aberrations that can assist clinicians and families in determining the genomic health of the fetus and pregnancy. Prenatal sample types (i.e. amniotic fluid, chorionic villus sampling and products of conception) processed for PCMA have an inherent noise that is very different from the noise in postnatal blood samples. This noise could be caused by several reasons such as GC (Guanine/Cytosine) content and epigenetic differences, which may render the samples more sensitive to the wet lab processes such as fragmentation. For blood samples, the manufacturer-built postnatal reference model provided through industry benchmark (PostnatalRM) is applied, which sums up all the background noise specific to blood samples and subtracts it from all the probe signals in each sample. However, when using PostnatalRM for prenatal samples, prenatal-specific background noise remained, which resulted in artifacts that were not related to true genomic abnormalities in the fetus and could have otherwise been reported as potential false positive findings. To reduce this noise, a prenatal reference model (PrenatalRM) was built with an iterative approach. The model was built with 121 prenatal samples with normal CMA results; 25% of which had marginal quality control (QC) issues due to prenatal noise. With PrenatalRM, we reduced the number of artifacts in the PCMA results and the QC parameters showed statistically significant improvement. A comprehensive in-silico validation using 88 anonymized samples showed that > 90% of the samples that had previously failed QC with many artifact variant calls using the PostnatalRM, passed the QC when using the PrenatalRM. This improvement in result quality was because of the exclusion of artifacts and background noise from almost all PCMA samples. The validation demonstrated that the PrenatalRM had 100% sensitivity and 100% specificity in detecting real variants. Use of the PrenatalRM significantly reduced the time spent curating and signing out the PCMA cases, which ultimately resulted in improved turn-around-time and higher quality of the results. Different facilities could utilize our PrenatalRM to compare the sample-specific genomic dosages, and step-by-step procedure of this work can be used as a guideline to build reference models based on the different sample types for CMA such as prenatal samples, cancer specimens or FFPE (formalin-fixed paraffin embedded) specimens.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2114. Prenatal diagnosis of recurrent severe form of infantile galactosialidosis exemplifying a pitfall in parental carrier screening

Authors:

M. Araújo Castro, V. P. do Val, J. L. S. Duque, L. Pires, R. Honjo, G. Yamamoto, D. Bertola, C. Kim; HCFMUSP, Sao Paulo, Brazil

Abstract Body:

Galactosialidosis is an autosomal recessive inherited metabolic disease, caused by biallelic variants in the cathepsin A (*CTSA*) gene, which encodes the lysosomal protective protein, essential for the stabilization of the beta-galactosidase and neuraminidase; thus, its impaired activity leads to a phenotype with features of both GM1 gangliosidosis and sialidosis. Prenatal diagnosis is not unusual, as galactosialidosis is one of the most common inborn errors of metabolism among cohorts of non-immune fetal hydrops. Here we report one couple of Japanese and Brazilian origin with three babies affected by early infantile galactosialidosis and prenatally diagnosed only in the last gestation. Two older siblings died as newborns, both with hydrops fetalis and no etiologic diagnosis. Our patient is a female newborn, third-born child. Second trimester ultrasonography showed polyhydramnios and hydrops fetalis and thus the mother was submitted to an amniotic fluid drainage, subsequently evolving to oligohydramnios. Amniotic fluid karyotype was 46,XX. Whole exome sequencing was performed and identified two heterozygous variants: NM_000308.3(*CTSA*):c.308G>T:p.(Gly103Val) and c.1369g.A:p.(Gly457Ser). The baby was born through a c-section, at 36 weeks, and presented with respiratory discomfort and apnea. Physical examination revealed a coarse face with sparse eyebrows, palpebral fullness, anteverted nostrils and smooth, long philtrum with a short neck. Abdominal ascites and edema of the labia minora were present, as well as bilateral congenital club foot. Echocardiogram showed patent foramen ovale and patent ductus arteriosus, in addition to hypoplastic aortic isthmus and severe pulmonary hypertension. Abdominal ultrasonography showed hepatomegaly and moderate ascites with debris. Urinary sialiloligosaccharides chromatography was suggestive of Galactosialidosis. At 35 days of age the echocardiogram showed left ventricular systolic dysfunction, and thus supplemental oxygen and diuretics were introduced. She died at 65 days of age. Before this last pregnancy, parental carrier screening was performed, and only the c.1369g.A:p.(Gly457Ser) variant was reported in the father; we then analyzed the raw data and identified the c.308G>T:p.(Gly103Val) in the mother, confirming they are in trans. This case suggests that reporting variants of unknown significance (VUS) in carrier screening when a potentially deleterious variant in the same gene is identified in the partner is important to the genetic counseling and could possibly allow prenatal genetic testing for monogenic diseases in at-risk couples.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2115. Prenatal diagnostic testing in cases of cystic hygroma or increased nuchal translucency: A single center experience.

Authors:

N. Owen^{1,2}, L. Vossaert^{1,2}, R. Zemet Lazar¹, I. Van den Veyver¹, J. Smith^{1,2}, H. Dai^{1,2}, F. Xia^{1,2}, C. Eng^{1,2}, L. Meng^{1,2}, W. Bi^{1,2}, ¹Baylor Coll. of Med., Houston, TX, ²Baylor Genetics, Houston, TX

Abstract Body:

Objective: Routine prenatal diagnostic genetic testing in the setting of fetal anomalies includes chromosomal microarray analysis (CMA) and increasingly whole exome sequencing (WES). We aimed to review the diagnostic yield and spectrum for cases in which cystic hygroma or an increased nuchal translucency (NT) was observed. Traditionally these findings are associated with aneuploidies and Noonan syndrome.

Results: Retrospective review of prenatal CMA and WES cases was performed to identify cases with cystic hygroma or increased NT over a 4-year period. Cystic hygroma was identified in 69 CMA cases (of which 20 had additional fetal anomalies noted at the time of testing). Of these, 21 were positive for an aneuploidy: ten 45,X, six trisomy (T) 18, three T21, one T22 and one mosaic T13. Four cases had other pathogenic findings: a derivative chr4 (2.7 Mb 4q loss & 77.8 Mb 3q gain), a recombinant chr6 (10.7 Mb 6q loss & 1.2 Mb 6p gain), a mosaic 12p gain (Pallister-Killian), and a fourth case that was positive for both ~20% mosaic loss of X and a derivative chr6 (7.7 Mb 6q loss & 48.5 Mb 8q gain). The two cases with distal 6q deletions suggest a possible association between this region and cystic hygroma, though more cases are needed to correlate. In addition, 107 CMA cases had increased NT (29 had additional anomalies). Eight were aneuploidy-positive: three T21, two T18, one T22, one 45,X, and one mosaic T2. Other pathogenic findings were seen in three cases: a homozygous deletion of *HBA1/2*, a de novo deletion of *ARID1B*, and a de novo DGS deletion. Altogether, 29/36 (80.6%) positive cases were solved by an aneuploidy.

Prenatal WES identified 12 cases with cystic hygroma and 11 cases with increased NT. Causative variants were detected in 5 cystic hygroma cases (*CHRNA1*, *NF1*, *PTPN11*, *SMARCB1*, and *HSPG2*) and 4 increased NT cases (*NSD1*, *MYMK*, *NXK32*, and *PDHA1*). Only two (22.2%) of these diagnoses (*PTPN11* and *NF1*) fall into the Noonan spectrum of disorders.

For both CMA and WES, there was no correlation between the presence of additional anomalies (seen in 49 out of 168 total cases) and pathogenic findings: 10/29 total aneuploidies and 2/7 other pathogenic CNVs had additional fetal anomalies. For WES, 4/10 cases with isolated findings were solved versus 5/13 with additional fetal anomalies.

Conclusion: These data show that while aneuploidy and Noonan Syndrome are major genetic etiologies of cystic hygroma and increased NT, other diagnoses need to be considered indicating broad testing strategies may be beneficial. The presence of additional anomalies does not appear to affect the diagnostic yield in either CMA or WES.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2116. Prenatal onset of Spondyloepimetaphyseal Dysplasia SIK3 Type (SEMDK)

Authors:

D. Wachtell¹, M. Aubert-Mucca², J. Martin¹, A. Kot^{1,3}, J. Goldstein⁴, D. Cohn¹, S. Wong¹, J. Zieba¹, D. Krakow^{1,2,3,4}; ¹Dept. of Orthopedic Surgery, David Geffen Sch. of Med. at Univ. of California at Los Angeles, Los Angeles, CA, ²Dept. of Obstetrics and Gynecology, David Geffen Sch. of Med. at Univ. of California at Los Angeles, Los Angeles, CA, ³Dept. of Human Genetics, David Geffen Sch. of Med. at Univ. of California at Los Angeles, Los Angeles, CA, ⁴Dept. of Pathology, David Geffen Sch. of Med. at Univ. of California at Los Angeles, Los Angeles, CA

Abstract Body:

The *SIK3* gene [MIM*614776] encodes the salt-inducible kinase 3, an AMP-activated protein kinase. *SIK3* has recently been identified as responsible for an autosomal recessive spondyloepimetaphyseal dysplasia, SIK3 Type (SEMDK) [MIM#618162]. This rare, severe skeletal dysplasia shares phenotypic findings with Jansen's metaphyseal chondrodysplasia (MCDJ) [MIM#156400] which results from constitutively active heterozygous mutations in *PTH1R*. *SIK3* and *PTH1R* are involved in mTOR signaling through a mechanism by which *PTH1R* inhibits the action of *SIK3* in the growth plate. Findings common to both disorders include shortened long bones with abnormal mineralization and hypercalcemia with progressive osteosclerosis in the metaphysis of long bones, metacarpals, metatarsals, and the base of the skull. The characteristic *SIK3* phenotype includes a transverse hypoechoic gap in the long bones, intellectual disability with brain anomalies, and a severe immunodeficiency.

Here, we described the first prenatal case from a non-consanguineous union, where exome sequencing identified the following compound heterozygous variants in *SIK3*(NM_025164.6): c.1987G>A and c.3548C>T, predicting the missense protein changes p.Gly663Arg and p.Pro1183Leu, respectively. Prenatal ultrasound showed subjective poorly mineralized calvarium, shortened long bones (<1st percentile) with a distinct area of hypoechoic signal in multiple long bones. 23-week post-mortem x-rays showed undermineralization of the calvarium, irregular, widened, and flared metaphyses with an irregular ossification front, multiple transverse gaps in long bones, absence of pubic bone ossification, delayed ossification of the basal occipital bone, and increased density of the base of the skull.

A previous report on the cartilage growth plate of *Sik3*^{-/-} mice showed retained chondrocytes within bone. Human cartilage growth plate analysis showed a shortened and disrupted hypertrophic zone, cartilage bone nests, defective primary spongiosum with deficient trabeculae formation. In the human growth plate, there was marked increased and abnormal vasculogenesis in all zones, affecting column formation, not evident in *Sik3*^{-/-} mice.

These findings further our understanding of SEMDK, its prenatal presentation, and that *SIK3* alterations affect the cartilage growth plate. Prenatal exome analysis can identify rare phenotypes that may not be included in gene panel studies. Analysis of growth plates suggest differences between humans and mice due to *SIK3* mutations highlighting the importance of human studies when available.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2117. Rapid testing of congenital diaphragmatic hernia novel variants using CRISPR/Cas9 mouse embryo editing.

Authors:

E. Bogenschutz¹, S. Brown¹, C. Heffner¹, C. Carpenter¹, K. Palmer¹, K. Peterson¹, K. Snow¹, Y. Shen², F. High^{3,4,5}, P. Donahoe^{4,5}, W. K. Chung^{6,7}, S. A. Murray¹; ¹The Jackson Lab., Bar Harbor, ME, ²Dept. of Systems Biology, Columbia Univ. Irving Med. Ctr., New York, NY, ³Dept. of Surgery, Boston Children's Hosp., Boston, MA, ⁴Pediatric Surgical Res. Lab., Massachusetts Gen. Hosp., Boston, MA, ⁵Dept. of Surgery, Harvard Med. Sch., Boston, MA, ⁶Dept. of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY, ⁷Dept. of Med., Columbia Univ. Irving Med. Ctr., New York, NY

Abstract Body:

Congenital diaphragmatic hernia (CDH) is a common and severe structural birth defect arising in 1:3,000 live births. CDH is characterized by either a weakening or partial loss of diaphragm muscle accompanied by lung hypoplasia and pulmonary hypertension, driving the associated 30-50% mortality. Many novel, damaging alleles predicted to contribute to CDH have been identified in humans, but few are functionally validated *in vivo*. This is due in part to the requirement for a mammalian model such as mice, and the high cost and time needed to generate these models. To close this technological gap, we have taken advantage of the high efficiency of CRISPR/Cas9 to create edited "F0" mouse embryos, allowing for rapid interrogation into whether the genetic alteration impacts diaphragm and lung development. Using this discovery platform we can generate both precise orthologous missense mutations via homology directed repair (HDR) using an oligonucleotide donor and insertion/deletions through non-homologous end joining (NHEJ), often leading to gene-knockouts. Diaphragm and lung defects are then scored using high-resolution micro-CT 3D imaging. From initial work with the platform, design of CRISPR/Cas9 gene targeting components to edited embryo phenotype takes an average of one to two months. Thus far, we have screened six statistically significant, predicted damaging missense mutations residing in both novel and known CDH genes discovered from CDH patient whole gene sequencing. From these six genes we have found that an allelic series in the known CDH gene *Zfp2*, a missense mutation in *Cnot1*, and an exon two deletion in *Cdc42bpb* all lead to diaphragmatic hernias in late-stage embryos and postnatal mice. The three remaining missense mutations either showed no observable structural defects in embryonic day (E)18.5 embryos, or few edited embryos were recovered at the developmental timepoint. This later cause is likely due to the increased propensity of NHEJ repair, leading to gene disrupting insertion/deletions and early embryonic lethality due to early gene loss. To improve the rate of HDR and increase the percent of embryos harboring the desired missense mutation we have begun testing multiple HDR enhancing methods and conditions *in vivo*. Treatments that increase HDR will be implemented in the embryo gene-editing platform.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2118*. Reconsidering the “advanced” maternal age for invasive prenatal testing

Authors:

I. Maya¹, L. Sagi-Dain², L. Basel-Salmon³, L. Salzer³; ¹Recanati Genetic Inst., petah tikva, Israel, ²Carmel Med. Ctr., Haifa, Israel, ³Rabin Med. Ctr., Petach Tikva, Israel

Abstract Body:

Background: Advanced maternal age, traditionally defined as maternal age over 35 years, has been used for decades to select pregnancies at increased risk for Trisomy 21. However, in the era of widely used non-invasive prenatal screening (NIPS), as well as invasive fetal testing by microarray techniques and the negligible risk for invasive testing-associated pregnancy loss, this maternal age cut-off must be re-evaluated. The objective of this study was to examine the appropriate maternal age threshold for recommendation of invasive testing for CMA. **Methods:** This retrospective cohort study was performed using database of Rabin Medical Center chromosomal microarray analysis (CMA) laboratory. All prenatal microarray tests in pregnancies with normal fetal ultrasound and recorded maternal age were included. We calculated the rate of clinically significant microarray findings per each year of maternal age, and plotted a receiver operating characteristics (ROC) curves to estimate the optimal cutoff for maternal age value. **Results:** Of the 7,033 prenatal microarray analyses, 108 (1.53%) clinically significant results were noted, 17 of these detectable by NIPS for common autosomal trisomies. ROC curve analysis failed to define a specific maternal age cut-off, in the overall cohort as well following omission of NIPS-detectable common autosomal trisomies. Only one abnormal microarray finding was found in 247 pregnancies of mothers aged 20-28 years (0.4%), vs. 21/1703 (1.23%) for mothers aged 29-34 years, 69/4195 (1.64%) for mothers aged 35-40 years, and 17/888 (1.91%) for mothers aged 41-50 years. Thus, using maternal age cut-off of 29 years for consideration of invasive testing, the majority (99.1%) of abnormal CMA findings would be detected, leaving an overall residual risk of 0.01%, while the cutoff of 35 years would detect 80% of abnormal findings, associated with a residual risk of 0.38% (one in 263). **Conclusions:** No clear maternal age cut-off could be established for the optimal detection of clinically significant microarray findings. Nevertheless, as a cut-off of 29 years enables to detect 99% of abnormal results, this might serve as the optimal cut-off, rather than the historical 35-years threshold.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2119. Recurrent constitutional chromosome 5 inversion

Authors:

M. Doco-Fenzy^{1,2,3}, E. Landais⁴, N. Gruchy^{5,3}, M. Spodenkiewick⁴, L. Herissant⁶, O. Pichon⁷, B. Cogne⁷, S. Bezieau⁸, M. Sebai⁹, M. Essid⁹, Y. Elaribi⁹, L. Lissy¹, C. Poisier¹, D. Sanlaville¹⁰, N. Chatron^{11,3}; ¹Service de génétique, REIMS, France, ²CHU Nantes, Nantes, France, ³ACLF, Paris, France, ⁴Service de génétique, CHU REIMS, REIMS, France, ⁵Service de génétique, Caen, France, ⁶Service de génétique, CHU Reims, REIMS, France, ⁷Service de génétique, CHU Nantes, Nantes, France, ⁸CHU HOTEL-DIEU, Nantes, France, ⁹Hosp. Mongi Slim, Tunis, Tunisia, ¹⁰HCL, CBPE, BRON Cedex, France, ¹¹Service de Génétique, Hospices Civils de Lyon, Bron Cedex, France

Abstract Body:

Background The recurrent pericentric inversion inv(5)(p13q13) is classically considered a polymorphism in Human with no clinical nor reproductive consequences. Indeed this inversion is described and found inherited. Surprisingly pericentric inversion, same size, on other submetacentric chromosome is described in meiotic rearrangement resulting in segmental duplication / deletion. Because few genomic data are reported in literature we investigated human families with different ethnic origins in order to update knowledge of this inv(5). **Method** We report 4 non related families with inv(5)(p13q13). - First and second inv(5) was diagnosed during pregnancy, and paternally inherited. - Third inv(5) was also discovered during pregnancy in context of elevated T21 serum markers, and found inherited. - Forth family, inv(5) was diagnosed on constitutional blood karyotype in context of blood myelodysplasia diagnosis. Inv(5) was inherited from healthy maternal grand father. The breakpoints of inv(5) were mapped in the 4 families. using Genome (WGS), Sanger sequencing, and Optical Genome Mapping (OGM). **Conclusion** To our knowledge, this study is the first molecular characterization of the inv(5)(p13q13) finding similar breakpoints in at least 3 families. We are looking for additional samples. A better knowledge of this entity will be useful to justify the decisions and prenatal genetic counselling in case of its fortuitous prenatal detection. Indeed different practice might occur facing the detection of this inv(5) in the karyotype of a foetus and its parents.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2120. Rescue of limb hemorrhage in a cohesinopathy mouse model with severe limb reduction

Authors:

A. Strasser, X. Zhou, A. Gonzalez-Reiche, M. Wu, A. Borges, B. Zhang, H. van Bakel, E. Wang Jabs; Icahn Sch. of Med. at Mount Sinai, New York City, NY

Abstract Body:

The Establishment of Sister Chromatid Cohesion N-Acetyltransferase 2 (ESCO2) acetylates the SMC3 subunit of the cohesin complex, which tethers sister chromatids together until their separation during anaphase, and regulates chromatin looping by mediating interactions between enhancers and promoters to modulate transcriptional outputs. Dysregulation of one or both processes underlie the molecular mechanism of Roberts syndrome (RBS), an autosomal recessive condition with profound growth deficiency and malformations, most notably phocomelia or severe limb reduction, which lend to its alias, pseudothalidomide syndrome. However, when and how these molecular processes are dysregulated in the developing limb to cause the RBS phenotype has yet to be elucidated. To investigate these causes, we generated an *Esco2* conditional knockout mouse line using *Prrxl-Cre* that recapitulates a human RBS loss-of-function mutation. Using bulk RNA-seq, single-cell RNA-seq, gene co-expression network, confocal microscopy, and immunohistochemistry analyses, we compared normal and mutant limb development from E9.5 to E13.5. Our results reveal a severe morphological and vascular phenotype culminating in skeletal reduction and hemorrhage of mutant limb buds by E12.5. These skeletal and vascular phenotypes are preceded by two transcriptional signatures at the onset of limb development in mutant limbs at E9.5: 1) a p53 pathway signature related to cell cycle arrest, DNA damage response, and cell death, and 2) an immune signature involving the inflammatory leukotriene signaling pathway. Given these unique, upregulated transcriptomic signatures at E9.5, we reasoned that mechanistically reversing these transcriptional signatures might in turn prevent the onset of the RBS limb phenotype. To predict potential drug candidates, we applied the Ensemble of Multiple Drug Repositioning Approaches (EMUDRA), a drug repositioning approach that matches transcriptomic signatures with drug-induced signatures. The analysis identified pifithrin-alpha, an FDA-approved p53 inhibitor, as a potential therapeutic agent. Administration of pifithrin-alpha successfully rescued the hemorrhage in mutant limbs, but not the skeletal phenotype. Taken together, our data indicate that 1) the vascular limb defects are linked to the p53 pathway, and that 2) partial morphological rescue suggests *ESCO2* might function in an additional transcriptional capacity. Further investigation of these mechanisms will significantly advance our understanding of normal and compromised limb development

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2121. Sex-specific variation in the commonly shared mothers' milk and infant gut microbiota are associated with sex disparity in childhood asthma risk.

Authors:

Z. Fang¹, S. A. Stickely¹, A. Ambalavanan¹, Y. Zhang², K. Fehr^{3,4}, S. Moossavi^{3,4}, C. Petersen^{5,6}, R. J. de Souza⁷, P. J. Mandhane⁸, E. Simons³, T. J. Moraes⁹, M. R. Sears¹⁰, S. E. Turvey^{5,6}, P. Subbarao⁹, M. B. Azad^{3,4}, Q. Duan^{1,2}; ¹Dept. of BioMed. and Molecular Sci., Queen's Univ., Kingston, ON, Canada, ²Sch. of Computing, Queen's Univ., Kingston, ON, Canada, ³Dept. of Pediatrics and Child Hlth., Univ. of Manitoba, Winnipeg, MB, Canada, ⁴Manitoba Interdisciplinary Lactation Ctr. (MILC), Children's Hosp. Res. Inst. of Manitoba, Winnipeg, MB, Canada, ⁵Div. of Allergy and Immunology, Dept. of Pediatrics, Univ. of British Columbia, Vancouver, BC, Canada, ⁶Dept. of Pediatrics, Child and Family Res. Inst. and British Columbia Children's Hosp., Vancouver, BC, Canada, ⁷Dept. of Hlth.Res. Methods, Evidence, and Impact, Faculty of Hlth.Sci., McMaster Univ., Hamilton, ON, Canada, ⁸Dept. of Pediatrics, Univ. of Alberta, Edmonton, AB, Canada, ⁹Dept. of Pediatrics, Hosp. for Sick Children and Univ. of Toronto, Toronto, ON, Canada, ¹⁰Dept. of Med., McMaster Univ., Hamilton, ON, Canada

Abstract Body:

Gut microbiota has been known to play a critical role in childhood asthma although the mechanisms remain poorly understood. Recent studies have identified bacteria that co-occur in mothers' milk and their infants' stools, which suggest that human milk microbiota could contribute to infant gut microbial colonization. To the best of our knowledge, this is the first study to investigate the impact of microbes commonly shared between mothers' milk and infants' gut on long-term lung health of human milk-fed children. Moreover, we hypothesize that sex-related variations of microbial composition could in part explain for higher prevalence of asthma and wheeze outcomes among boys compared to girls. Sequencing of 16S rRNA from 885 human milk samples of the CHILD Cohort Study were applied to assess milk microbial composition. Genome-wide single nucleotide polymorphisms (SNPs) were genotyped in these mothers using the Illumina HumanCoreExome BeadChip. Infant genetic risk score (GRS) was previously calculated using a published genome-wide association study (GWAS) of asthma based on 44 SNPs. Gut microbial composition was determined by 16S rRNA sequencing of stool samples from human milk-fed infants at 3 months (N=323). Linear regression analysis stratified by infant sex identified the main and interaction effects of the commonly shared microbes and GRS on recurrent wheeze (ages 2-5 years). Our results indicate that male infants with a high GRS who were exposed to increased *Bifidobacterium longum* in their mothers' milk and own gut had lower recurrent wheeze prevalence (FDR-adjusted P=0.06). This trend was not found in female infants. Instead, among female infants with a high GRS, our results indicate that increasing both the milk and gut microbe *Streptococcus* spp. were independently associated with increasing prevalence of recurrent wheeze (FDR-adjusted P=0.02). Lastly, we assessed the interaction between maternal SNPs and environmental exposures on the milk microbes. Decreased abundance of *Streptococcus* spp. were significantly associated with interactions between loci in chromosome 4 and prenatal whole grain consumption (e.g., rs66976123, P = 3.4 × 10⁻⁸). This interaction result suggests that prenatal whole grain consumption could protect against infant recurrent wheeze development by modulating *Streptococcus* spp. abundance. Our study integrated human milk and infant gut microbiota while considering infant sex, and provided insight into the underlying mechanisms of sex disparity in asthma and determinants of asthma-related microbe synthesis. These results have potential to guide early life sex-specific prevention strategies for childhood asthma.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2122. Single-nucleus chromatin accessibility and RNA sequencing reveal impaired brain development in prenatally e-cigarette exposed neonate's brain

Authors:

Z. Chen¹, W. Chen¹, Y. Li¹, M. Moos², D. Xiao¹, C. Wang¹; ¹Loma Linda Univ., Loma Linda, CA, ²U.S. Food and Drug Admin., Silver Spring, NH

Abstract Body:

Although emerging evidence reveals that vaping alters the function of the central nervous system, the effects of maternal vaping on offspring brain development remain elusive. Using a well-established in utero exposure model, we performed single-nucleus ATAC-seq (snATAC-seq) and RNA sequencing (snRNA-seq) on prenatally e-cigarette-exposed rat brain. We found that maternal vaping distorted neuronal lineage differentiation in neonatal brain by promoting excitatory neurons and inhibiting lateral ganglionic eminence-derived inhibitory neuronal differentiation, resulting in excitatory to inhibitory neuron ratio change (E/I imbalance). By integrating gene expression and chromatin accessibility data, we recaptured the neurogenesis pseudotime course and identified the developmental widows from where e-cigarette exposure shifted the excitatory and inhibitory neuron differentiation trajectory. In addition, canonical pathway enrichment analysis revealed that synapse formation and synaptic transmission were the major biological processes effected by prenatal e-cigarette exposure. Moreover, maternal vaping disrupted calcium homeostasis, induced microglia cell death, and elevated susceptibility to cerebral ischemic injury in the developing brain of offspring. In sum, our results suggest that the aberrant calcium signaling, diminished microglial population, and impaired microglia-neuron interaction may all contribute to the underlying mechanisms by which prenatal e-cigarette exposure impairs neonatal rat brain development. Our findings raise the concern that maternal vaping may cause adverse long-term brain damage to the offspring.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2123*. Sperm mosaicism predicts transmission of *de novo* mutations to human blastocysts.

Authors:

M. Breuss¹, X. Yang², V. Stanley², J. McEvoy-Venneri², X. Xu², A. J. Morales³, J. G. Gleeson^{2,4}; ¹Univ. of Colorado, Sch. of Med., Aurora, CO, ²Univ. of California, San Diego, La Jolla, CA, ³Fertility Specialists Med. Group, San Diego, CA, ⁴Rady Children's Inst. for Genomic Med., San Diego, CA

Abstract Body:

De novo mutations underlie individually rare but collectively common pediatric congenital disorders. Some of these mutations can also be detected in tissues and from cells in a parent, where their abundance and tissue distribution can be measured. We previously reported that a subset of these mutations is detectable in sperm from the father. In addition, every male carries up to dozens of mosaic mutations in their sperm, and we predicted that ~1 in 15 males harbors such a mutation that is likely deleterious if transmitted. However, it is unclear how these detected mutations transmit to the next generation based on their measured abundance. Here, in three independent couples undergoing *in vitro* fertilization, we first identified male gonadal mosaicism through deep whole genome sequencing. We then confirmed variants and assessed their transmission to supernumerary preimplantation blastocysts (32 total) through targeted ultra-deep genotyping. Across 55 gonadal mosaic variants, 15 were transmitted to blastocysts for a total of 19 transmission events. This represented an overall predictable but slight under-transmission based upon the measured mutational abundance in sperm. We replicated this conclusion in an independent, previously published family-based cohort. Therefore, unbiased preimplantation genetic testing for gonadal mosaicism may represent a feasible approach to reduce the transmission of potentially harmful *de novo* mutations. This—in turn—could help to reduce their impact on miscarriages and pediatric disease.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2124. Successful implementation of a short turn-around time prenatal exome sequencing process in a perinatal centre.

Authors:

C. Barnett^{1,2}, L. De Jong³, A. Rogers¹, W. Waters⁴, L. Rawlings⁵, K. Simons³, S. Gao³, J. Soubrier³, R. Kenyon⁵, M. Lin⁶, R. King⁶, D. Lawrence⁶, S. Le Blanc¹, L. McGrgeor¹, S. Sallevelt¹, J. Liebelt¹, T. Hardy^{3,7}, H. Scott^{3,8}, K. Kassahn^{3,8}; ¹Paediatric and Reproductive Genetics Unit, Women's and Children's Hosp., North Adelaide, Australia, ²Faculty of Hlth.and Med. Sci., Univ. of Adelaide, Adelaide, Australia, ³Genetics and Molecular Pathology, SA Pathology, Adelaide, Australia, ⁴Genetics and Molecular Pathology, SA Pathology/Women's and Children's Hosp., North Adelaide, Australia, ⁵Genetics and Molecular Pathology, SA Pathology, North Adelaide, Australia, ⁶ACRF Cancer Genomics Facility, SA Pathology, Adelaide, Australia, ⁷Repromed, Dulwich, Australia, ⁸Ctr. for Cancer Biology, An alliance between SA Pathology and the Univ. of South Australia, Adelaide, Australia

Abstract Body:

Background: In 2020 we introduced prenatal exome trios for real-time clinical decision making into our diagnostic and clinical services. This testing is offered with reporting prior to 23 weeks gestation. *Aim:* We reviewed the outcomes of testing including its impact on pregnancy management for the first 18 months of offering this testing. *Methods:* An audit of the number of referrals, diagnostic outcomes and clinical outcomes was performed and we are presenting the results of this testing. *Results:* We reported 38 prenatal exome trios with an average turn-around time of 12 days from receipt of request. Diagnostic rate was 37%. Although likely pathogenic and pathogenic variants matching phenotype were generally only reported, in one pregnancy we reported suspicious compound heterozygous variants of uncertain significance in *EXOSC3* after consultation with the referring clinician. Notable diagnoses included nemaline myopathy (*NEB*), Skraban-Deardorff syndrome (*WDR26*), Smith-Lemli-Opitz syndrome (*DHCR7*), thanatophoric dysplasia (*FGFR3*), Sotos syndrome (*NSDI*), and congenital myasthenic syndrome (*RAPSN*). Noonan syndrome was also commonly diagnosed with *de novo* variants in *PTPN11* and one case with a maternally inherited *PTPN11* variant. We report on the impact of early diagnosis prior to 23 weeks gestation on pregnancy counselling and early decision-making regarding perinatal care. *Conclusion:* We have successfully expanded our prenatal genomic care to offer early trio testing to assist in real-time pregnancy counselling with a workable turnaround time of 12 days. This was possible within the context of a single-centre by streamlining the clinical referral and diagnostic testing processes without the need for on-demand sequencing, thereby demonstrating its potential for broader adoption in tertiary care hospitals and pathology laboratories.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2125. Tensor decomposition on electronic health records identifies latent subtypes of preterm birth that associate with polygenic risk scores

Authors:

S. Cruz Gonzalez¹, A. Abraham², J. Capra¹; ¹Bakar Computational Hlth.Sci. Inst., Univ. of California San Francisco, San Francisco, CA, ²Vanderbilt Univ., Nashville, TN

Abstract Body:

Phenotypic and clinical heterogeneity are hallmarks of many human diseases. Refining disease categorization by explicitly modeling phenotypic and longitudinal heterogeneity has the potential to reveal mechanistic and therapeutic insights.

To illustrate this potential, we leverage rich longitudinal phenotypic data from two electronic health record (EHR) databases to study the heterogeneity of preterm birth in over 30,000 deliveries obtained from two distinct medical systems. Preterm birth is a syndromic disease with substantial heritable risk, but limited understanding of its genetic architecture and few treatment options. We apply tensor decomposition to EHR-derived phenotypes from 28,954 individuals in the Vanderbilt EHR. The tensor decomposition identifies individuals with distinct disease traits and temporal patterns across multiple latent factors. In both white and black cohorts, we discover interpretable latent factors representing clinical subtypes dominated by morbidity along distinct disease axes, such as reproductive, mental health, and endocrine systems. The longitudinal trajectories of latent factors ranged from acute pre-pregnancy to chronic post-pregnancy. Using each individual's combination of latent factors, we are able to train accurate boosted decision tree models to predict risk of preterm birth. Integrating genetic data available for a subset of the cohort, we find that several latent factors are associated with polygenic risk scores for relevant risk factors, such as diabetes and BMI, while others do not have strong genetic associations. We then compare the latent factors derived in a separate EHR cohort of 5,978 deliveries from UCSF. Our work demonstrates that tensor decomposition can parse phenotypic heterogeneity and identify disease subtypes with distinct phenotypic and temporal signatures, enable risk prediction, and discover novel genetic associations. This approach has the potential to identify heterogeneous disease processes underlying many complex diseases.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2126. The abnormal expression of has-miR-196b-5p, FAS, and FAS-AS1 genes in Iranian patients with recurrent pregnancy loss

Authors:

S. Ghaderian, M. Ahmadi; Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of

Abstract Body:

Recurrent pregnancy loss (RPL) is recognized as the most common pregnancy complication. Several factors and underlying mechanisms have been proposed; however, the exact etiology remained unclear and needed to be elucidated. Apoptosis is a vital process in the normal development of the fetus, and any aberrant expression of apoptotic genes during development probably takes part in the occurrence of RPL. Based on the comprehensive literature review, bioinformatic assessments, and the importance and crucial role of the FAS and FAS-AS1 genes and hsa-miR196b-5p during the apoptosis pathway in RPL, the present study has investigated the expression levels, correlation, and diagnostic merit of them in the blood samples of women with a history of RPL in comparison with those women with healthy fertility as the control group (n = 50, age < 35). The obtained data revealed that while the expression level of the has-miR-196b-5p was significantly decreased (P = 0.0001), the FAS (P = 0.0001) and FAS-AS1 (P = 0.01) genes were significantly upregulated in RPL women compared to the control samples. Besides, a significant positive correlation between the expression levels of FAS and FAS-AS1 (r = 0.55, P = 0.0001) and a remarkable negative association between FAS and has-miR-196b-5p (r = -0.35, P = 0.01) were disclosed. However, the observed relationship between the expression levels of has-miR-196b-5p and FAS-AS1 was not significant (r = -0.2, P = 0.19). Moreover, the results demonstrated that among these genes, the highest diagnostic value for distinguishing RPL samples from normal ones was related to the mRNA expression of the FAS gene (AUC = 0.74, P = 0.0001). In conclusion, the findings of the current report have suggested that the aberrant expression of FAS, FAS-AS1, and hsa-miR196b-5p genes in blood specimens of patients with RPL compared to their control counterparts can be considered promising biomarkers in blood specimens for the diagnosis, prognosis, and targeted therapy of the women suffering from RPL.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2127. Understanding caudal developmental abnormalities using single-nucleus multi-omics data from wild type and Danforth's short tail mouse E9.5 tailbuds

Authors:

C. Zajac¹, R. D. Albanus^{2,1}, N. Manickam¹, E. Curka¹, P. Thomas¹, C. Keegan¹, S. C. J. Parker¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Washington Univ., St. Louis, MO

Abstract Body:

Caudal embryo development is important for multiple organ systems arising from all three germ cell layers, including the gastrointestinal system, genitourinary system, and the lower spinal cord. Embryonic malformations associated with caudal birth defects affect 1 in 10,000 human live births. The Danforth's short tail (Sd) mouse exhibits cessation of the vertebral column at the lumbar level and malformations of urogenital and gastrointestinal organs and provides a model to study the etiology of human caudal birth defects. The Sd mutation is caused by an endogenous retroviral (ERV) insertion 12.5 kb upstream of the *Ptfla* gene in the promoter region of the lncRNA gene *Gm13344*, a region orthologous to a human pancreatic developmental enhancer. To better understand the cell-specific regulatory landscape of the Sd mouse, we performed single nucleus chromatin profiling (snATAC-seq) from control (WT) and Sd E9.5 tailbuds that went under stringent quality control using the *ataqv* package followed by clustering. We then performed single nucleus multiome (ATAC+RNA on the same nucleus) profiling from WT E9.5 tailbuds. For quality control, we matched barcodes assigned to each modality and performed clustering based on the RNA modality. We obtained 10,428 and 10,422 high-quality ATAC profiled nuclei for WT and Sd tailbuds respectively from the first data set. Joint clustering analyses of these nuclei discovered 18 distinct clusters that represent diverse cell types, some of which have differential chromatin accessibility at key caudal embryonic developmental gene markers, including *Ptfla*. Most recently, we have generated data using single nucleus multiome (ATAC+RNA on the same nucleus) data from WT E9.5 tailbuds where initial QC revealed 7,203 high-quality nuclei that were used for clustering. This dual-modality data set generated 11 clusters at the RNA level, grouping together cells expressing important gene markers for embryonic development such as Sonic Hedgehog (Shh) and lncRNA 9030622022Rik. Preliminary results provide an insight to cluster characterization and identified transcription factors potentially responsible for changes in chromatin accessibility. The next steps in this project involve further clustering exploration and integration of chromatin accessibility modality to identify transcription factors perturbed as a result of the Sd mutation. We expect that these integrative analyses will improve our understanding of the causal cell types and regulatory networks that influence the Sd phenotype and, consequently, caudal developmental disorders in humans.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2128. Understanding the burden of clonal and non-clonal sperm mosaicism by bulk and single-cell genome sequencing

Authors:

X. Yang^{1,2}, M. Breuss³, C. Barrows¹, K. Kennedy⁴, J. Blackinton⁴, G. Harton^{4,5,6,7}, J. Gleeson^{1,2}; ¹Univ. of California, San Diego, La Jolla, CA, ²Rady Children's Inst. for Genomic Med., San Diego, CA, ³Univ. of Colorado, Sch. of Med., Aurora, CO, ⁴BioSkrbyb Genomics Inc., Durham, NC, ⁵Bryant Univ., Smithfield, RI, ⁶Eastern Virginia Med. Sch., Norfolk, VA, ⁷Univ. of Kent, Canterbury, United Kingdom

Abstract Body:

Mutations during embryonic development, aging, cellular metabolism, and environmental exposure are permanently recorded in the genomes of each cell and its daughters. Depending upon whether the mutations can be detected in regular next-generation sequencing, they are recognized as clonal or non-clonal in nature and present different features. Elucidating the patterns of these mutations and their potential to transmit to offspring is a key to understanding congenital *de novo* mutation (DNM) disorders and genetic variability across human generations. As parents age, the number of DNMs observed in the offspring increases, and with this, an increased risk of disease is caused in offspring. Although age-related DNM risks have been reported in large populations, our understanding of how paternal-specific clonal and non-clonal mosaicism contribute to offspring is still limited due to technical challenges. We recently demonstrated that a considerable portion of DNMs in children with neurological and psychiatric disorders arise from clonal mosaic mutations in the sperm and the life-long transmission risk from clonal sperm mosaicism does not increase with age (Yang and Breuss et al. Cell. 2021), resulting in congenital disorders or miscarriage in 1/300 concepti. We further aim to unravel the feature of the non-clonal gonadal mutation burden at the single-cell level. We compared the 30x genomic sequences obtained from primary template-directed amplification workflow from 14 fluorescent sorted live single human sperm cells collected from 2 healthy young men, to establish a control background. More than 80% of the single-sperm reached >1x coverage for >85% genome, with a 10% false-negative rate and 1/2 base substitution rate compared with conventional methods, allowing us to study SNVs in single sperms. The GC contents and replication timing matched the patterns in transmitted *de novo* mutations after our variant calling and filtering pipelines. We established that each sperm harbors 380 +/- 120 sperm-detectable mosaic mutations, >90% of them are non-clonal. These preliminary results help us to understand the non-clonal mosaic mutational burden in human sperm. This result will enable us to discover the non-clonal age-related genetic impacts of human sperm.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2129. Utility of preconception carrier screening in couples with unexplained infertility.

Authors:

S. Yatsenko¹, P. Ghosh¹, M. Schmitz², J. Settino³, M. Aarabi⁴, J. Sanfilippo⁵, A. Rajkovic⁶; ¹Univ Pittsburgh, Magee-Womens Hosp., Pittsburgh, PA, ²Univ. of Toronto, Toronto, ON, Canada, ³Magee Womens Res. Inst., Pittsburgh, PA, ⁴Univ Pittsburgh Med Ctr, Pittsburgh, PA, ⁵Univ Pittsburgh, Dept. of Obstetrics, Gynecology and Reproductive Sci., Pittsburgh, PA, ⁶Univ. of California San Francisco, San Francisco, CA

Abstract Body:

Diagnosis of unexplained infertility is given to ~30% of infertile couples. Patients intending to proceed with an IVF cycle frequently undergo a preconception screening for autosomal recessive and X-linked conditions associated with moderate to severe childhood manifestations, ~40% of which are also associated with neonatal and prenatal lethality. While most births after assisted reproductive technology (ART) are uncomplicated, pregnancies following ART also been reported to have an increased risk of miscarriage and congenital structural abnormalities. We hypothesize that a subset of couples with unexplained infertility may carry damaging variants in such genes. Exome sequencing was performed on DNA from peripheral blood samples of 19 couples with a diagnosis of unexplained infertility. Variants in 112 preconception screening genes recommended by the American College of Medical Genetics and Genomics (ACMG) (Tier-3) were classified according to ACMG guideline for variants interpretation. The probabilities of both partners to carry variants in the same gene were calculated based on the allele frequencies reported in the Genome Aggregation Database (gnomAD v2.0.2). Aggregate gene carrier frequencies were used to calculate the expected disease prevalence according to Hardy-Weinberg Equilibrium and compared to the observed disease prevalence in the general population. We detected 11 pathogenic, 3 likely pathogenic, and 65 heterozygous single nucleotide variants of uncertain significance (VUS) in 40 genes and three copy number variants. In four (4/19, 21%) couples, both partners were carriers of a variant in the same gene: *TNX8* (Ehlers-Danlos-like syndrome); *DHCR7* (Smith-Lemli-Opitz syndrome); *G6PC* (Glycogen storage disease type IA); and *CYP21A2* (congenital adrenal hyperplasia due to 21-hydroxylase deficiency). The observed proportion of affected couples was much higher than the highest expected 4% rate in the general population. We identified a subset of developmentally constrained genes with a much lower prevalence of affected live born individuals than expected, and a high frequency of heterozygous VUS alleles. Our findings indicate an enrichment for variants in genes associated with severe childhood conditions among individuals affected by infertility. Rare genetic variants may also contribute to infertility among heterozygote carriers due to embryonic/fetal lethality or metabolic disturbances. Further studies are necessary to determine the significance of rare coding variants in a set of preconception screening genes with respect to their functional significance for embryonic development, fetal loss and infertility.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2130. Variant reclassification in expanded carrier screening brings challenges for clinical management: experience of a Hong Kong prenatal diagnosis centre.

Authors:

Y. Zheng¹, Y. Cao¹, T. SHAM¹, A. KWAN¹, B. CHAN², C. LAU¹, J. CHONG¹, M. CHAU¹, T. LEUNG¹, K. Choy¹; ¹The Chinese Univ. of Hong Kong, Hong Kong, China, ²Hong Kong Maternal & Fetal Med. Clinic, Hong Kong, China

Abstract Body:

Expanded carrier screening (ECS) utilizing next-generation sequencing could simultaneously evaluate hundreds of genes associated with autosomal recessive (AR) and X-linked conditions. Besides reporting pathogenic (P)/ likely pathogenic (LP) variants classified by current knowledge, it also detects variants with uncertain significance (VUS) or conflicting interpretations, which have witnessed reclassification based on new evidence after the initial report. However, its clinical impact on the screened couples was unclear. Therefore, we investigated the frequency and evidence of ECS variant reclassification and evaluated its clinical impact. From April 2019 to May 2022, cases requesting ECS (covering 302 genes) with complete clinical records were recruited. Of all 978 participants, 364 were couples, 219 were single females and 31 were single males. Particularly, 44.4% (259/583) of female cases had an ongoing pregnancy at recruitment. Overall, 11.1% (109/978) of cases received reclassification alerts regarding 81 variants from 56 genes. The duration between initial and amended reports ranged from 1 to 35 months. Re-curated genes accounted for 18.5% (56/302) of the ECS panel, and 48.2% (27/56) of the genes were in the “metabolic newborn screening” catalog. Among all re-curated variants, 81.5% (66/81) were missense variants. In comparison with the variant curation in ClinVar on June 1st, 2022, 51.9% (42/81) of the re-curated variants were still of conflicting interpretations. 75.3% of reclassified variants (61/81) were upgraded from benign (B)/likely benign (LB)/VUS to P/ LP mainly due to updated patient data, *in silico* prediction data, or population data. For all four null variants upgraded from VUS to P, new evidence was derived from the establishment of a loss-of-function molecular mechanism for the gene. There were 23.5% (19/81) of variants upgraded from LP to P. Additionally, 1.2% of variants (1/81) were downgraded from LP to VUS. In consequence, additional counseling or management was needed for 3.3% (32/978) cases: 1) unscreened partners of 27 cases were recommended for ECS; 2) phasing the newly and priorly reported variants for diagnosis of an AR condition was suggested for 3 cases, and 3) 2 cases were recommended to investigate their ongoing pregnancies or offspring since their reclassified variants were for AR conditions in which their partners were also carriers. Variant reclassification was common in ECS and brought challenges to post-reclassification counseling and management of patients. These findings emphasize the importance of pre-test counseling for notifying the possibility of variant reclassification and a change in risk status.

Prenatal, Perinatal, and Developmental Genetics Posters - Thursday

PB2131. Whole genome sequencing and analysis of 4,053 individuals in trios and mother-infant duos from the Born in Guangzhou Cohort Study

Authors:

S. Huang; Guangzhou women and children's Med. center, Guangzhou Med. Univ., Guangzhou, China

Abstract Body:

Large-scale birth cohorts that recruit trios/duos families are essential resource for determining the genetic and environmental contribution to maternal-infant's health. While large-scale genomic studies of randomly chosen individuals are carried out around the world, study based on birth cohorts is scarce especially among the Chinese and the East Asian population. Here, we present the Phase I genomic study of the Born in Guangzhou Cohort Study (BIGCS) where we conduct whole-genome sequencing (~6.63x) and analysis of 4,053 Chinese participants in trios or mother-infant duos living in South China. We identify 56.23 million high-quality genetic variants, including 32% novel ones and construct a reference panel with 4,490 haplotypes that enables more accurate genotype imputation for individuals of Chinese ancestry. We characterize the population genetic structure and dissect the genetic relationship of 10 Chinese dialects in details utilizing the unique geographical and cultural composition of BIGCS participants. We investigate genome-wide associations for multiple maternal and infant traits and reveal novel genetic associations with total bile acid, gestational weight gain and lipid metabolism level in maternal peripheral or fetal cord blood. Using inter-generational mendelian randomization, we have dissected the maternal and fetal genetic effect underlying the observed associations between maternal phenotypes and fetal growth. Our findings fill the gap of the missing diversity in human genetics and demonstrate the great value of genomic study of birth cohort in advancing the worldwide maternal-infant health and genetic knowledge.

Prenatal, Perinatal, and Developmental Genetics Posters - Wednesday

PB2132. Whole-genome sequencing analysis in families with pregnancy loss

Authors:

T. Workalemahu; Univ. of Utah, Salt Lake City, UT

Abstract Body:

Background: One to two percent of couples suffer recurrent pregnancy loss and over 50% of the cases are unexplained. Whole genome sequencing (WGS) analysis has the potential to identify previously unrecognized causes of pregnancy loss, but few studies included DNA from families, including parents, losses, and live births.

Methods: We conducted a pilot WGS study of four families with recurrent pregnancy loss, including parents, losses and healthy live births. DNA was extracted from parents' saliva, buccal swabs from ten live-births, and formalin-fixed paraffin embedded placenta from eight unexplained pregnancy losses, which included an embryonic loss (<10 weeks' gestation), fetal deaths (10-20 weeks' gestation) and stillbirths (≥ 20 weeks' gestation). We used the Illumina platform for WGS and state-of-the-art protocols for DNA restoration, variant detection, and quality control. We utilized variant prioritization methods and online databases, and calculated probability of being loss-of-function intolerant (pLI) scores to evaluate pathogenicity of single nucleotide variants (SNVs), including de novo variants, and variants following autosomal dominant, compound heterozygous and X-linked recessive modes of inheritance.

Results: In samples that passed quality control (n=3 families), we identified 87 SNVs involving 75 genes in embryonic loss (n=1), 370 SNVs involving 228 genes in fetal death (n=3), and 122 SNVs involving 122 genes in stillbirth (n=2). Of these, 22 de novo, 6 autosomal dominant and an X-linked recessive SNVs were pathogenic (pLI>0.9), impacting known genes (e.g., DICER1, FBN2, FLT4, HERC1, and TAOK1) involved in embryonic/fetal development and congenital abnormalities. Further, we identified missense compound heterozygous SNVs impacting genes (e.g., VWA5B2) in two fetal death samples that were absent from live births and population controls, providing evidence for haplosufficient genes relevant to recurrent pregnancy loss.

Conclusions: In this pilot study, we identified SNVs that may be relevant to unexplained pregnancy loss, providing the justification for conducting WGS using larger numbers of families to identify embryonic/fetal-lethal variants. Our findings also warrant validation by targeted sequencing of additional families and losses. Elucidating genes causing pregnancy loss will facilitate the development of risk stratification strategies and novel therapeutics.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2133. "Grown up" minors: Recontact after withdrawal or attrition

Authors:

D. Patrinos, B. M. Knoppers; Ctr. of Genomics and Policy, McGill Univ., Montreal, QC, Canada

Abstract Body:

The participation of minors in biomedical research is critical for generating meaningful research outcomes which can translate into positive health outcomes. Many pediatric diseases are rare and heterogenous. Moreover, minors have distinct health and physiologic needs from adults. However, the participation of minors in research raises distinct ethical challenges. One such challenge is the issue of re-contact. Re-contacting minors enrolled in a research project upon their attaining the age of majority or medical maturity to seek their informed consent to continue their participation is considered an ethical requirement. This is especially relevant in the context of genomic research, where data and biological samples collected from minors may be stored, shared, and used for many years. While the issue of re-contacting minors has received significant attention in the literature, one aspect that has not received much attention has been the question of re-contacting a minor who has been withdrawn from a research project or who has been lost to follow-up, once they reach the age of majority. Indeed, the ethical permissibility of re-contact in these circumstances has not been thoroughly addressed. Our presentation will therefore analyze the ethical permissibility of re-contacting minors whose participation in research has ended, once they have reached the age of majority or maturity. Our approach identifies 3 scenarios in which the participation of a minor in a research project may end: (1) parental (or guardian) withdrawal from the research project, (2) the minor themselves asks to be withdrawn from the research project, and (3) loss to follow-up (attrition). We consider the ethical permissibility of re-contact in each scenario, based on a review of international and national research ethics guidelines and policies, as well as the relevant scholarship. We argue that re-contact is ethically permissible if it was the parent/guardian who chose to withdraw the minor or in cases where the minor was lost to follow-up. Respect for autonomy - the recognition of now-adult minor's right to choose - is the basis of this conclusion. However, where the minor themselves had asked to be withdrawn during childhood, we argue that re-contact is not ethically permissible, based on respect for their wishes and their right to be heard. Finally, we conclude with an overview of the ways in which informed consent may be sought from the now-adult participant.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2134. “Frankly, I thought this was one and done”: Patient and provider perspectives on return of reclassified variants in clinical oncology

Authors:

S. Makhnoon¹, Z. Zhao², T. Mendoza¹, R. Volk¹, S. Shete³, B. Arun⁴, S. Peterson⁵; ¹UT MD Anderson Cancer Ctr., Houston, TX, ²Univ Texas HSC Houston, Houston, TX, ³Univ Texas MD Anderson Cancer Ctr, Houston, TX, ⁴MD Anderson Cancer Ctr., Houston, TX, ⁵Univ of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract Body:

Background: Genetic variant reclassification can clarify a variant’s clinical significance and is increasingly facilitated by the availability of updated information about normal human genomic diversity. A primary challenge in genomic medicine today is accurate reinterpretation of identified variants and relaying that information to patients. Despite the prevalence of variant reclassification, the clinical and socio-behavioral consequences of reclassification for patients, or how clinicians return and manage reclassifications remains unknown. **Methods:** We conducted a qualitative study with a sample of 26 participants in two stakeholder categories: 19 patients with reclassified variants in cancer susceptibility genes, including 16 (84%) with downgraded variants and 3 (16%) with upgraded variants; and 7 key informant oncologists and genetic counselors. Participants were recruited from three geographically dispersed cancer hospitals in Texas, New Jersey, and Ohio. Semi-structured interviews were iteratively coded and analyzed for themes related to experience of receiving and returning reclassified variants, challenges, and suggested improvements. **Results:** News of reclassification, although unexpected, were received favorably by patients with understanding of the evolving evidence base in genetics and appreciation for being informed. Patients with downgraded variants reported not being able to recall receiving a reclassified result. Patients rarely communicated news of downgrades to their relatives as they believed that their relatives would not be interested in this development. Patients with upgraded variants emphasized the desire for more personalized return system including scheduled phone calls, telegenetics appointments, and in-person meetings. Key informant provider interviews revealed endorsement of the return of all variants including downgrades. However, due to the large number of downgrades and their clinical insignificance, low-touch automated methods of return were suggested to reduce workload including mailed letters and MyChart updates rather than phone calls. Rules under the 21st Century Cares Act designed to prevent information blocking was reported as facilitator of more timely return of reclassified results. **Conclusion:** Stakeholder perspectives regarding the return of reclassified genetic variants suggested several actionable changes to the return-of-results system. In the absence of established guidelines on the optimal way to recontact patients with new valuable genetic information, this work provides much needed empirical evidence from the perspectives of patients and providers.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2135. “I just wanted more”: Hereditary cancer syndromes patients’ perspectives on the utility of circulating tumour DNA testing for cancer screening

Authors:

M. Clausen¹, E. Adi-Wauran^{1,2}, S. Shickh^{1,2}, A. R. Gagliardi^{2,3}, A. Denburg^{4,2}, L. E. Oldfield⁵, J. Sam¹, E. Reble¹, S. Krishnapillai^{1,2}, D. A. Regier^{6,7}, N. N. Baxter^{8,1,2}, L. Dawson⁹, L. S. Penney¹⁰, V. Carile¹¹, M. Basik^{12,11}, S. Sun^{6,13}, K. A. Schrader^{6,13}, A. Karsan^{6,13}, A. Pollett¹⁴, T. J. Pugh^{5,15}, Y. Bombard^{1,2,15}; ¹St. Michael's Hosp., Toronto, ON, Canada, ²Univ. of Toronto, Toronto, ON, Canada, ³Toronto Gen. Hosp. Res. Inst., Toronto, ON, Canada, ⁴The Hosp. for Sick Children, Toronto, ON, Canada, ⁵Princess Margaret Cancer Ctr., Toronto, ON, Canada, ⁶BC Cancer Agency, Vancouver, BC, Canada, ⁷Univ. of British Columbia, Vancouver, ON, Canada, ⁸Univ. of Melbourne, Melbourne, Australia, ⁹Mem. Univ., St. John's, NL, Canada, ¹⁰IWK Hlth.Ctr., Halifax, NS, Canada, ¹¹Jewish Gen. Hosp., Montreal, QC, Canada, ¹²McGill Univ., Montreal, QC, Canada, ¹³Univ. of British Columbia, Vancouver, BC, Canada, ¹⁴Mount Sinai Hosp., Toronto, ON, Canada, ¹⁵Ontario Inst. for Cancer Res., Toronto, ON, Canada

Abstract Body:

Background: Hereditary cancer syndromes (HCS) predispose individuals to a higher risk of developing multiple cancers in their lifetime. However, current cancer screening strategies are limited in their ability to screen for all cancer risks. Circulating tumour DNA (ctDNA) detects DNA fragments shed by tumour cells in the bloodstream and has the potential to detect cancers early, advancing tailored clinical management. ctDNA is being integrated into care. Yet, there is a lack of evidence on patient perspectives on using ctDNA in their care.

Aim: To explore patients’ perspectives on the utility of ctDNA in their care to help inform its clinical adoption and implementation.

Methods: A qualitative interpretive description study was conducted using semi-structured phone interviews preceded by a brief background video on ctDNA with adult HCS patients recruited from CHARM, a Canadian research consortium. Participants were purposively sampled until thematic saturation was reached. Data were thematically analyzed using NVivo. Strategies such as iterative analysis, constant comparison, and double coding were used to ensure quality and rigour.

Results: Thirty HCS patients were interviewed (n=19 women, n=25 self-reported as white, age range 20s-70s). Participants were highly concerned about their risk of developing cancers, particularly those without reliable screening options for early detection. They “just wanted more” than what their current screening strategies can offer. Hence, participants were enthusiastic about ctDNA’s potential to be comprehensive (able to detect multiple cancers including those without reliable screening methods), predictive (able to detect cancers early), and tailored (lead to the personalization of their clinical management). They thought adding ctDNA as “another tool in the toolbox” would help fill gaps in current screening. That said, participants acknowledged potential limitations of ctDNA, including the risks of false positives, false negatives, and experiencing additional anxiety while awaiting results. However, they saw the potential benefits of ctDNA outweighing its limitations. Indeed, they were highly tolerant of a test that may result in further tests or procedures because of their fear of developing another cancer: “I will take anything because I know that my body can turn”.

Conclusion: Participants saw the utility of ctDNA in their care, as reflected by how their belief in ctDNA’s potential to fill gaps in current screening and improve their care overshadowed its limitations. These findings may be used to inform ctDNA’s clinical adoption and implementation.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2136*. A New System for Triageing Patients in a Medical Genetics Clinic Saves Time and Improves Quality.

Authors:

L. Randolph, C. E. F. Miers; Children's Hosp. Los Angeles, Los Angeles, CA

Abstract Body:

Background: We receive 150 referrals in Medical Genetics /month. It is time-consuming to triage them. In the past, they were printed for an MD's review. Although we strove for consistency, there was variation. This took ~7 minutes per referral, or 18 hours/month. **Methods:** We began an algorithmic approach to electronic triaging with a genetic counselor (GC) in charge. An algorithm was developed for each indication which covers which additional studies are required; the urgency; and which MD is appropriate. Each triage now takes ~4 minutes. We have developed 33 algorithms. As fragile X and CMA are first-tier tests for a child with isolated LD/ID/autism, we require those studies prior to being scheduled. If studies are declined by the insurer, we see the patient without them. If results are positive, the referral is triaged under a different algorithm (e.g., "abnormal microarray"), which allows them to be seen sooner. Sometimes, free parental testing is offered by the lab. In those cases, we require such testing prior to scheduling. This allows for more informed genetic counseling of results. Often, patients are referred to discuss abnormal testing results, but the results aren't included in the records. Previously, we cancelled appointments due to lack of results during pre-clinic review. This left open slots in our schedule that were not fillable in the short time before the clinic day. This represents lost revenue (a new visit list charge is \$645) and missed opportunities to reduce our waitlist. **Results:** We have triaged ~3500 referrals since instituting this protocol. At 50% less time required per case compared to the prior system, we spend 8 hours/mo of a GC's time on initial triage. Most of these require a second partial review to determine if the correct results or requested records have been received before scheduling can occur. Requiring prior testing in the common LD/ID/autism category of referrals has saved the clinic, patient and insurer an extra patient visit. The workup begins, and the clinic appointment occurs sooner than it would have. It also saved our MDs' and staff's time in ordering and trying to obtain authorization for these tests. Scheduling fewer visits for these patients leaves slots open for other patients and less rescheduling of appointments due to lack of test results. Our referring providers have accepted the system. **Summary:** Standardization of triaging is a fair way of booking patients with better throughput, saving patients and insurers time and money and improving patient care. A GC performs this function well with occasional input from a physician. We recommend this method to other medical offices.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2137. A Sociotechnical Analysis of Returning Genomic Informed Risk Assessments in Primary Care Pediatrics

Authors:

D. Karavite¹, S. Terek¹, J. Connolly², M. Harr³, M. Behr¹, N. Muthu¹, R. Grundmeier¹, H. Hakonarson³; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Children s Hosp. of Philadelphia, Philadelphia, PA, ³Children S Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Background: The Electronic Medical Records and Genomics (eMERGE) network is a National Institutes of Health funded consortium conducting research in genomics, including discovery, clinical implementation, and public resources. eMERGE IV is focused on implementing a program based on polygenic risk scores and will recruit 25,000 individuals across 10 clinical sites, all of whom will receive a genomic-informed risk assessment (GIRA) that combines monogenic and polygenic determinants with social determinants, patient, and family history to delineate if participants are at high risk for any of 10 conditions, 4 of which affect children (asthma, obesity, type I diabetes, type II diabetes).

In a pediatric setting, the primary care pediatrician who receives GIRAs will play an important role in responding to caregiver questions and concerns, communicating with genetics and other specialties, and managing relevant recommendations. In developing tools and processes to support the return of results workflow, we applied a sociotechnical model, Safety Engineering in Patient Safety, to our requirements analysis to represent the human, technical, environmental, and other factors essential to the evaluation and design of information technology.

Methods: We performed a sociotechnical analysis to inform the design of systems and processes to support the return of GIRA results in a pediatric primary care network. We developed an interview format with seven phases including a review of example GIRA reports and the result workflow, 5 questionnaires (demographics, genetics experience, low-risk results, high-risk results, support systems), and semi-structured interview questions.

Results: Interviews with 20 participants revealed several important challenges. While pediatric primary care providers feel they have an important role in helping to manage genetic testing results like the GIRA, they require additional education, knowledge of evidence supporting results and recommendations, collaboration and participation of genetics, and highly efficient processes and tools to streamline communication and workflow in the highly time stressed environment of primary care.

Conclusion: Sociotechnical models provided important requirements for systems and processes that go beyond the technical requirements of the electronic health record.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2138. A systematic review of health policy guiding the identification, analysis, and management of genomic secondary findings

Authors:

S. Majeed¹, C. Johnston¹, C. Mighton¹, V. Rokoszak², S. Grewal², V. Aguda², D. Malkin³, Y. Bombard²; ¹Univ. of Toronto, Toronto, ON, Canada, ²St. Michael's Hosp., Toronto, ON, Canada, ³The Hosp. for Sick Children, Toronto, ON, Canada

Abstract Body:

Introduction: An explosion in the utilization of genomic sequencing for precision medicine over the past decade has yielded an analogous abundance in secondary findings (SFs). Existing practices within the continuum of SF identification, analysis and management are numerous, inconsistent, and sometimes contradictory across health conditions and geographic regions at a local, regional, and international levels. This widespread practice variation can lead to inconsistent care and subsequent downstream impacts on SF management and patient outcomes. As the frequency of genomic profiling continues to rise, synthesis of the SF health policy landscape is of utmost importance in order to ascertain current needs, gaps and limitations across the SF policy continuum. **Methods:** We carried out a systematic review to catalogue and appraise the current guidance established internationally directing the identification, analysis, and management of SFs for participants receiving genomic sequencing. A comprehensive search in MEDLINE, Embase and Cochrane databases was conducted. Articles were reviewed independently by two reviewers using Covidence for screening, data extraction and quality assessment. AGREE-II was used for critical appraisal and a narrative synthesis was conducted. **Results:** We screened 710 unique records and found 31 relevant studies producing guidance for SF interrogation based on our eligibility criteria, most of which were developed in North America. Preliminary results indicate that most policies focus on the SF management (97%; n = 30), but fewer provide guidance on the SF analysis (58%; n = 18) or identification (55%; n=17). Further, the most frequent topic in managing SFs is patient consent (42%, n=13), followed by the return of results (39%; n = 12) and medical actionability (n=9). Appraisal of policies revealed a low level of quality particularly for rigour of development (i.e., strength of evidence used), stakeholder involvement, and editorial independence. **Conclusions:** Our results highlight key gaps in health policy related to specific SF genes or classes used for identification and best practices for SF analysis methods. Moreover, although SF management processes like informed consent and disclosure are well described, best practices for other procedures following disclosure (such as surveillance, treatment, etc.) remain unknown. Lastly, many studies did not have adequate supporting evidence from which to base guidance, and this remains a significant limitation of the applicability, generalizability, and usefulness of these policies. Future work should focus on filling these policy gaps and supporting evidence-based practice.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2139. Access to Clinical Genetics Service and Testing for Patients with Retinal Dystrophies: a Quality Improvement Study

Authors:

Z. Shao¹, B. G. Ballios², H. Faghfoury³; ¹Univ. of Toronto, Toronto, ON, Canada, ²Dept. of Ophthalmology and Vision Sci., Univ. of Toronto, Toronto, ON, Canada, ³Dept. of Med., Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Retinal Dystrophies (RD) is a group of genetic disorders that leads to vision impairment in both pediatric and adult populations. They can be isolated or syndromic depending on the genetic etiology. Over 300 genes, both nuclear and mitochondrial, cause RD. Gene therapy for RD associated with RPE65 was approved in USA since 2017. Additional gene therapies for isolated and syndromic RD are in their final stage of clinical trials, increasing the demand for early molecular genetic diagnosis for RD. Limited clinical genetic resources are creating bottlenecks in patient care and potentially limiting access to therapeutic options for RD patients. We used quality improvement (QI) methods to identify factors impacting genetic care of RD patients. We conducted an audit of RD referrals to a tertiary hospital adult genetics clinic between September 2017 to December 2019 to investigate wait times from referral to initial assessment and time to results disclosure. Of the 316 total referrals, the time from referral to initial assessment ranged from 10 months to 24 months (mean 14 months) and the time from initial assessment to results disclosure was stable at approximately 6 months. A pilot model to mainstream genetic testing during initial assessment by ophthalmology was trialed between January-June 2022. Of 200 new referrals to an RD expert clinic, 122 were seen within this 6-month period, with approximately 80% offered a comprehensive RD genetic panel based on a clinical diagnosis of non-syndromic RD. Within this timeframe 49 genetic tests were sent and 42/49 results were disclosed to patients, with several confirmed diagnoses resulting in alterations to patient management. Challenges associated with this model include a lack of resources for coordination of sample collection and results tracking, as well as challenges in securing genetic counsellor support for patient counselling and cascade testing in family members. Further investigations are ongoing to merge the unique strengths of both clinical genetic testing models. We believe this study serves as an example of QI for optimization of clinical care for other treatable genetic conditions.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2140. Advancing Health Equity in Genomics: Reflections and Recommendations for Future Research Directions from an NHGRI Workshop

Authors:

L. Hindorff¹, E. Madden¹, A. Jackson¹, T. Akintobi², K. Baker³, R. Begay⁴, E. Burchard⁵, J. Carpten⁶, N. Cox⁷, V. Di Francesco¹, D. Dillard⁸, F. Fletcher⁹, M. Fullerton¹⁰, N. Garrison¹¹, E. Green¹, C. Hammack-Aviran¹², J. Hildreth¹³, V. Hiratsuka¹⁴, C. Horowitz¹⁵, C. Hughes Halbert⁶, M. Inouye^{16,17}, R. Kittles¹⁸, L. Landry^{19,20}, J. Leek²¹, N. Limdi²², N. C. Lockhart¹, E. Ofili², E. Pérez-Stable²³, N. Risch⁵, M. Sabatello²⁴, L. Saulsberry²⁵, L. Schools¹, J. Troyer¹, B. Wilfond^{26,10}, G. Wojcik²⁷, V. Bonham¹, S-J. Lee²⁸, J. Cho¹⁵; ¹NHGRI, NIH, Bethesda, MD, ²Morehouse Sch. of Med., Atlanta, GA, ³Whitman-Walker Inst., Washington, DC, ⁴Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ⁵Univ California San Francisco, San Francisco, CA, ⁶Univ. of Southern California, Los Angeles, CA, ⁷Vanderbilt Univ Med Ctr., Nashville, TN, ⁸Southcentral Fndn., Anchorage, AK, ⁹Baylor Coll. of Med., Houston, TX, ¹⁰Univ. of Washington, Seattle, WA, ¹¹Univ. of California, Los Angeles, Los Angeles, CA, ¹²Vanderbilt Univ. Sch. of Med., Nashville, TN, ¹³Meharry Med. Coll., Nashville, TN, ¹⁴Univ. of Alaska Anchorage, Anchorage, AK, ¹⁵Icahn Sch. of Med. at Mount Sinai, New York, NY, ¹⁶Wellcome Trust Sanger Inst, Cambridge, United Kingdom, ¹⁷Univ. of Cambridge; Baker Heart and Diabetes Inst., Cambridge, United Kingdom, ¹⁸City of Hope, Duarte, CA, ¹⁹Dana Farber Cancer Inst., Boston, MA, ²⁰Univ. of Pennsylvania, Philadelphia, PA, ²¹Fred Hutchinson Cancer Res. Ctr., Seattle, WA, ²²Univ. of Alabama at Birmingham, Birmingham, AL, ²³NIMHD, NIH, Bethesda, MD, ²⁴Columbia Univ., NY, NY, ²⁵The Univ. of Chicago, Chicago, IL, ²⁶Seattle Children's Res. Inst., Seattle, WA, ²⁷Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ²⁸Columbia Univ., New York, NY

Abstract Body:

Although advances in genomics have fueled the application of genomics to improve understanding of human health and disease, not all groups have benefitted equitably. Lack of diverse representation in genomic research and the genomic workforce impedes equitable access to and application of genomics and thus may increase health disparities disproportionately. Without improvement in research, resource allocation, and clinical practice, inequities in the benefits of genomic research and its applications will persist. The National Human Genome Research Institute (NHGRI) held a virtual public workshop in April 2022 to identify opportunities in genomics and health equity research, with over 300 attendees. The workshop objectives were to: 1) identify areas of genomic research that are important to advance health equity; 2) identify research and partnerships needed to address structural factors that impact health equity in genomics; and 3) define how success is measured within genomics and health equity.

Recommendations identified from keynote speakers, panelists, breakout groups, and interactive Zoom chats will be presented. In summary: first, genomic research and the workforce need to continue to diversify. Next, additional data are needed on how lack of diversity in genomic research impacts health inequities at various stages of translation from scientific discovery to clinical and public health implementation. Research is needed on the effects of inappropriate use of race or other socially constructed traits and the underutilization of genomic markers within algorithms used for laboratory reference ranges and clinical diagnoses. To build trust and conduct equitable and ethical research, researchers should nurture long-standing relationships with communities that have been underrepresented in and harmed by biomedical research. Outreach to groups representing socially constructed traits in addition to race, such as sexual and gender minorities or individuals with disabilities, was encouraged. Metrics of health equity, such as measures of access to genomic testing, should be applied across genomic studies. The top recommendation from attendees was to support genomic workforce development by targeting Minority Serving Institutions. The next highest-ranked recommendation was to allocate

ASHG 2022 Annual Meeting Poster Abstracts

sufficient time and equitable resource distribution for community engagement. Supporting research to improve health equity in genomics aligns with NHGRI's guiding principles to strive for global diversity in all aspects of genomics research and to maximize both the usability of and the opportunities to benefit from genomics for all members of the public.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2141. Anxiety is a prominent neurobehavioral feature in Kabuki syndrome that significantly affects quality of life.

Authors:

A. Kalinousky¹, T. Rapp², H. Hijazi¹, J. Johnson³, H. Bjornsson⁴, J. Harris³; ¹Johns Hopkins Univ., Baltimore, MD, ²Univ. of North Carolina, Chapel Hill, NC, ³Kennedy Krieger Inst., Baltimore, MD, ⁴Univ. of Iceland, Reykjavik, Iceland

Abstract Body:

Background: Kabuki syndrome (KS) is a Mendelian Disorder of the Epigenetic Machinery caused by loss of function variants in either of the two genes, *KMT2D* (34-76%) or *KDM6A* (9-13%), both which are involved in the regulation of histone methylation. Previously, representative neurobehavioral deficits of KS were recapitulated in a *Kmt2d*-mouse model, emphasizing the role of KMT2D in brain development, specifically in hippocampal neurogenesis in the granule cell layer of the dentate gyrus. Interestingly, anxiety, a phenotype that is well-associated with decreased hippocampal neurogenesis, has been anecdotally reported in individuals with KS. Methods: In this study anxiety and behavior were assessed in a cohort of 63 individuals with molecularly confirmed KS diagnosis, as well as in their 25 biologically unaffected siblings, via clinically-designed questionnaires (SCARED/GAS-ID and CBCL/ABCL). Participants' age ranged from 4 to 44 years old, with 86% of participants having a pathogenic variant in *KMT2D*, and the rest having variants in *KDM6A*. In addition, data was collected on adaptive function and quality of life in participants with KS using appropriate surveys including ABAS-III and PROMIS. Survey scores were analyzed across different age groups of affected individuals and in contrast to their unaffected siblings using two-tailed paired and unpaired t-tests as well as correlation analyses. Results: Individuals with KS were found to have significantly higher anxiety scores and total behavior problem scores than their unaffected siblings ($p=0.023$, $p=0.022 \times 10^{-5}$). Moreover, a large proportion of affected individuals (23.9% children and 64.7% adults) have clinically significant anxiety that does not correlate with impaired adaptive function while negatively correlating with the quality of life ($p=0.000166$). Executive dysfunction was found to be the most prominent neurobehavioral phenotype, aside from anxiety, amongst individuals with KS, especially in younger children. Conclusions: These findings indicate that anxiety is a prominent neurobehavioral feature of KS that increases with age and is correlated with decreased quality of life. Providers should therefore carefully screen individuals with KS for anxiety in order to allow for prompt intervention. Neurobehavioral anxiety measures may also be an important outcome measures for clinical trials in KS.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2142. Assessing the performance of the Clinician-reported Genetic testing Utility InDEx (C-GUIDE): Further evidence of inter-rater reliability

Authors:

R. Hayeems^{1,2}, S. Luca¹, L. Chad^{1,2}, N. Quercia^{1,2}, A. Hossain¹, M. Meyn³, E. Pullenayegum^{1,2}, W. J. Ungar^{1,2}; ¹The Hosp. for Sick Children, Toronto, ON, Canada, ²Univ. of Toronto, Toronto, ON, Canada, ³Univ. of Wisconsin Sch. of Med., Madison, WI

Abstract Body:

Background: Genome sequencing (GS) outperforms other rare disease diagnostics, but standardized approaches to assessing its clinical utility are limited. In previous work, we developed and validated a novel outcome measure, the Clinician-reported Genetic testing Utility InDEx (C-GUIDE). C-GUIDE is a 17-item measure that operationalizes this concept from providers' perspective. It captures the utility of genetic testing related to understanding diagnosis and prognosis, informing medical management, awareness and actionability of reproductive and health risks for patients/family members, and patient/family well-being. Preliminary evidence of its inter-rater reliability was obtained through a clinical vignette study. The purpose of this study was to further assesses its inter-rater reliability using actual clinical cases.

Method: One genetic counsellor and one medical geneticist independently completed C-GUIDE Version 1.1 after genetic test results were disclosed to a shared set of pediatric patients or their family members. Raters also completed a case description questionnaire, including information about the patient's age, indication for testing and type of test chosen. Inter-rater reliability was assessed by comparing the raters' C-GUIDE scores using analysis of variance (ANOVA) to generate intra-class correlation coefficients (ICC).

Results: Of the 42 patients rated on C-GUIDE, the most common indications for testing were hearing loss (n = 18) or craniosynostosis (n = 11) and the most common tests ordered were gene panels and microarrays. Test results were diagnostic or partially diagnostic for 11 patients, potentially diagnostic for 14 patients, or non-diagnostic for 17 patients. The ICC for C-GUIDE was 0.61 (95% CI: 0.56, 0.65). When examining individual C-GUIDE items, the ICC between raters was >0.7 for 10 items, 0.5-0.7 for 1 item, and <0.5 for 6 items. Low-moderate ICCs were identified for items related to reduced diagnostic impact/medical management, the identification of future health risks, and psychosocial well-being. ICCs also varied by result type (i.e. diagnostic, potentially diagnostic, non-diagnostic). Agreement on the global item was 0.57 (95% CI: 0.33, 0.74).

Conclusion: Rater instructions for item completion will be modified to improve consistency of item interpretation. While further assessments of reliability are warranted following modifications, these findings provide additional evidence that C-GUIDE demonstrates acceptable inter-rater reliability and can be used as a strategy for measuring the value of genetic testing, as perceived by clinicians.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2143. Assessment of cancer patients using a distress thermometer to evaluate perceived and calculated stress in order to identify individuals requiring additional psychological support.

Authors:

T. Bhende¹, Q. Hasan¹, M. Srinivas¹, C. Rao²; ¹Kamineni Hosp., Hyderabad, India, ²Renovo Soumya Hosp., Hyderabad, India

Abstract Body:

Assessment of cancer patients using a distress thermometer to evaluate perceived (PS) and calculated stress (CS) in order to identify individuals requiring additional psychological support. The diagnosis and treatment of cancer drastically affects the life of individuals. The psychological stress differs from person to person and its assessment will help in identifying the type of psychological support required. To assess this, NCCN has recommended the use of Distress Thermometer (DT) as a screening measure for identifying psychological distress among cancer patients. The current preliminary study was carried out on 300 cancer patients who were undergoing treatment in two different oncology units. The distress interview was facilitated by a genetic counselor. Results indicated that Females vs males (PS= 39.66% vs 14 %) (CS= 25.7% vs 15.11%) perceive high stress & same was observed with the calculated stress. But, as the duration from time of diagnosis increased, the level of both PS & CS decreased, irrespective of gender. It was observed that, younger patients had higher stress. To our surprise, the percentage of individuals with high perceived stress was least in colorectal maybe because other categorized types were female cancers, reiterating above mentioned results. Among the 300 individuals 24% were suspected to be at a higher risk of Hereditary Cancer Syndrome and thus genetic counseling was provided to them, and it was observed that these patients had higher stress in comparison to non-hereditary, because it was compounded with their concerns about other family members, specially children. The observations of the present study indicates the important role of a genetic counselor at an oncology unit in identifying both, those who require support from psycho-oncologist and others who require genetic testing & post-test counseling.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2144. Assessment of reading comprehension levels of interpretation sections of genetic testing reports

Authors:

M. Gillentine, J. Narayanan, C. Paschal; Seattle Children's Hosp., Seattle, WA

Abstract Body:

With advances in technology and awareness, the number of individuals and families receiving genetic testing is increasing. Additionally, the 21st Century Cures Act allows access to test results in as little as 2 business days after completion, often without guidance from genetic counselors or clinicians. While there are efforts to increase genetic literacy in the public, diagnostic labs are responsible for providing reports with interpretation that is accessible to patients. To assess if we are meeting such a need, pediatric chromosomal microarray (CMA) and Next Generation Sequencing (NGS) panel reports were pulled from electronic medical records in Epic from 2020 to May 2022, during which time there were 3 lab directors. Nine NGS panels analyzing 10-50 genes were assessed, focused on diabetes, vascular anomalies, interstitial lung disease, differences in sex development, craniosynostosis, Rett/Angelman syndrome, and intestinal dysmotility. We analyzed 149 CMA report interpretation sections (n = 80 with pathogenic/likely pathogenic results, n = 56 with VUS, n = 8 with autosomal recessive disease risk, n = 4 with long stretches of homozygosity, n = 3 with likely benign results) and 225 NGS panel reports (n = 202 with pathogenic/likely pathogenic results, n = 28 with VUS) from 2020 to 2022 with eight different reading level tools (Flesch Reading Ease formula, Flesch-Kincaid Grade Level, Gunning Fog Scale, SMOG Index, Coleman-Liau Index, Automated Readability Index, and Linsear Write Formula). CMA and NGS reports were both above the national average reading level (7th-8th grade), with CMA reports (reading level = college level on average) written at a significantly higher grade level than NGS reports (reading level = 10th-11th grade on average). No differences were observed by sequencing panel. For NGS reports, those with pathogenic results were written at a significantly higher reading level than those with VUS. These results highlight the need to increase accessibility to genetic testing reports. We are interested in identifying strategies to provide more comprehensible reports, such as using plain language when possible and providing resources for individuals and families receiving results, especially when clinicians and genetic counselors may be unavailable at the time of test results being released into patients' charts. Future studies may investigate additional report types, such as those from exome sequencing. Increasing the comprehension of genetics reports will require input from all levels of the testing process including genetic counselors, variant scientists, lab directors, and clinicians.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2145. Awareness about rare diseases through Indian cinema

Authors:

A. Nanda; Kamineni Hosp., New Delhi, India

Abstract Body:

It is estimated that 6-8% of the world's population are affected by Rare diseases. Despite this staggering number, public awareness is limited. The aim of the present study was to analyze how RDs are depicted in popular Indian movies, given its large audience and if they are accepted by them.

In the present study, a movie search was conducted using IMDB to find multilingual movies addressing RDs, as well as a digital survey was conducted to understand the impact of movies on public understanding of RD's . From 75,000 movies made since 1913, only 49 Indian movies depicting a RD received popularity. This includes 70% Hindi, 6% Marathi, 2% Kannada, 2% Bengali, 8% Malayalam and 6% Telugu.

Analysis was confined to Hindi movies because of its popularity. Following are the movies and respective RD's. 35% of movies showed different cancer types, with none discussing hereditary cancer syndromes, an important RD, Congenital deafness (Khamoshi, 1972), Blindness (Sparsh, 1980) and speech impairment (Black, 2005) , Alzheimer's (U me aur Hum, 2008), Multiple Sclerosis (Guru, 2007), Cerebral Palsy (Margarita with a Straw, 2014), Tourette syndrome (Hichki, 2018), Autism and Asperger's syndrome (My Name is Khan, 2010), Aplastic Anemia (Rockstar, 2011) and Progeria (Paa, 2009). Some of these movies were box office hits, suggesting that people were willing to appreciate such films. Films depicting RDs don't give sufficient information about the cause, clinical implications, prevention and management strategies to educate the audience, however the survey indicated that it can be used as a medium to improve their understanding of RDs.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2146. Barriers to genetic testing when indicated at diagnosis for autism spectrum disorder.

Authors:

A. Wang, I. D. Little, C. Gunter; Natl. Human Genome Res. Inst., Bethesda, MD

Abstract Body:

Multiple professional organizations recommend conducting genetic testing at the time of a diagnosis for autism spectrum disorder (ASD), but as few as 20% of individuals ever receive genetic testing for this indication. We have taken a mixed-methods approach to assess barriers to genetic testing. First, we questioned whether genetic literacy of families and patients might be a significant factor. Genetic literacy is defined as the knowledge of genetic principles and the ability to apply this information to everyday life. Higher genetic literacy is associated with greater comprehension of genetic test results and higher participation in genetic healthcare services. Previously, we developed a survey to measure three facets of genetic literacy: familiarity with genetic terms, accuracy in genetic knowledge, and skills to apply genetic information to health. We distributed this survey first to a general population sample (N=2050), allowing us to compare it to a similar sample from 2013. After adjusting for education, the 2021 sample scored significantly higher in term familiarity, but not in knowledge, although scores still fell in the “moderate” range for both facets. We then surveyed participants in SPARK (N=2023), a large genetic research study in ASD. Compared to the general population, participants in SPARK scored significantly higher in all three facets, although still in the moderate range.

To identify additional challenges with ASD genetic testing, we report on a second study interviewing healthcare providers who actively diagnose ASD to assess the needs and barriers in the process. Qualitative analysis reveals that multiple factors, including a lack of formal genetics education, parental concerns, insurance barriers, and an inability to make genetic testing referrals directly, all contribute to low rates of genetic testing usage for ASD. Overall, our findings suggest that low genetic testing uptake may be due to only moderate genetic literacy among the general population and numerous practical barriers for healthcare providers. Future research should focus on raising genetic literacy among families to help them make informed decisions, and to streamline the incorporation of genetic testing into the ASD diagnosis process.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2147*. Benefits, harms and costs of newborn genetic screening for long QT syndrome: estimates from the PreEMPT Model

Authors:

K. Christensen¹, P. M. McMahon², J. M. Yeh³, N. K. Stout⁴, C. Y. Lu¹, S. Stein², R. J. Hyland⁵, A. C. Wu⁶; ¹Harvard Pilgrim Hlth.Care/Harvard Med. Sch., Boston, MA, ²Harvard Pilgrim Hlth.Care Inst., Boston, MA, ³Boston Children's Hosp./Harvard Med. Sch., Boston, MA, ⁴Harvard Pilgrim Hlth.Care Inst., Boston, MA, ⁵Boston Children's Hosp., Boston, MA, ⁶Boston Children's Hosp. / Harvard Med. Sch., Boston, MA

Abstract Body:

Background: Long QT syndrome (LQTS) is a potentially lethal and highly-actionable condition that the American College of Medical Genetics and Genomics recommends for secondary findings disclosure during clinical exome or genome sequencing. Population genetic screening for LQTS is feasible, but whether its benefits would outweigh harms and justify costs is unclear, particularly among newborns. Methods: The Precision Medicine Policy and Treatment (PreEMPT) Model uses microsimulation modeling to examine the impact of newborn genetic screening. As part of the PreEMPT Model, we developed a module to compare genetic screening for LQTS against usual care in a U.S. birth cohort of 3.7 million newborns. Newborn genetic screening included surveillance of individuals with pathogenic/likely pathogenic variants or a family history of disease. Usual care included surveillance of individuals with a family history of LQTS. Outcomes included numbers of diagnosed and undiagnosed individuals; reductions in sudden cardiac death; number of individuals receiving implantable cardioverter defibrillators (ICDs); and incremental cost-effectiveness ratios. Results: Through age 20 years, newborn genetic screening would reduce the number of undiagnosed individuals from 1,412 to 620 (difference: 792, 95% uncertainty interval (UI): 266 to 1487) and the number of sudden cardiac deaths associated with LQTS from 49 to 17 (difference: 32, 95% UI: 17 to 42) but increase the numbers of children receiving ICDs from 18 to 132 (difference: 114, 95% UI, 70 to 556). Compared to usual care, newborn genetic screening costs \$91,000 per life-year saved (95% UI, \$67,000 to \$179,000 per life-year saved). Conclusions: Results suggest newborn genetic screening for LQTS would reduce undiagnosed cases and prevent deaths, but would greatly increase the number of children who receive treatment and also increase use of surgical intervention. The cost-effectiveness of newborn genetic screening for LQTS was borderline at conventional willingness-to-pay thresholds. This study demonstrates how modeling can provide insights into the tradeoffs between benefits and costs that will need to be considered as newborn genetic screening is more widely adopted.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2148. Cancer patients' experience of receiving variant of uncertain significance results: An Asian perspective.

Authors:

N. Ishak, T. Shaw, S-T. Li, J. Yuen, H. Goh, Z. Chua, F. V. F. Que, Z. Zhang, J. Chiang, J. NGEOW; Natl. Cancer Ctr. Singapore, Singapore, Singapore

Abstract Body:

Objective: Due to a lack of ancestry-matched, functional, or segregation data, Asians have a higher rate of having a variant of uncertain significance (VUS) reported, with up to 50% of patients receiving VUS results following panel testing. The management of VUS results has been challenging, as it has been met with increased anxiety and distress among cancer patients undergoing genetic testing. Furthermore, misinterpretation of VUS results as pathogenic has resulted in misdiagnoses and incorrect clinical management. This study aims to investigate cancer patients' experience of receiving a VUS result as an exploratory study among the Asian population.

Methods: A qualitative, semi-structured interview study was conducted to explore patients' views of receiving a VUS result. Participants who had undergone genetic testing and received a VUS result were recruited through the Cancer Genetics Service of the National Cancer Centre Singapore. Twenty participants agreed to partake. Participants' responses were transcribed verbatim and analysed using thematic analysis to identify significant themes.

Results: Thematic analysis revealed five major themes, namely: (1) VUS results are interpreted as an uncertain outcome; (2) a VUS result provides relief and encourages positive behavioural adjustments (3) fatalism and religion are used as coping mechanisms to deal with the uncertainty of the result; (4) genetic counsellors, family and the community provide reassurance and support; and (5) patients value updates to variant classifications for future management.

Conclusion: This novel study has provided unique insight into Asian patients' perspectives on receiving a VUS result. It highlights how patients manage their VUS results and uncertainty, which could improve the pre-and post-test counselling practice in Asia. Emphasis must be placed on clear communication of VUS results to dispel the possibility of misconceptions, misdiagnosis, and mismanagement in cancer care.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2149*. Challenges in incorporating polygenic risk scores in the electronic health record.

Authors:

H. Bangash, A. Miller, S. Nigbur, J. Petrzela, P. Caraballo, R. Freimuth, I. Kullo; Mayo Clinic, Rochester, MN

Abstract Body:

Purpose: In its current phase, the electronic MEDical Records and GENomics (eMERGE) Network is returning polygenic risk scores (PRSs) to participants and integrating the related discrete data into the electronic health record (EHR). We describe challenges encountered in incorporating PRS results into the EHR at Mayo Clinic Rochester, Minnesota—an eMERGE IV site.

Methods: The study team worked on 5 domains to integrate PRS data into the EHR: i) develop detailed workflows for participant recruitment, and return of PRS results supplemented with monogenic causes and family history; ii) identify relevant resources needed for the study—a development team with the expertise and capacity to build PRS data ‘parsers’ to facilitate data transfer into the EHR was assembled after interactions with multiple institutional information technology (IT) teams; iii) apply for multiple institutional committee approvals; iv) build a prototype clinical decision support (CDS) to trigger in the EHR for patients with a high PRS for any 1 of 10 diseases of interest; and v) develop PRS-related educational content to serve as a resource for clinicians.

Results: Over a 2-year span, the study team engaged with 30 different committees (e.g., clinical decision support subcommittee, clinical communications subcommittee), departments (e.g., community internal medicine, IT), and groups (e.g., Mayo Center for Individualized Medicine) across the institution to integrate PRSs into the EHR. During this time, the team encountered a number of challenges, both logistical and financial including the need to: i) obtain multiple institutional committee approvals, which delayed downstream work related to CDS development; ii) develop custom data ‘parsers’ to interface with enterprise APIs, requiring significant financial investment; iii) build a new genomic test order in the EHR to enable PRS results reports to be viewed by both patients and clinicians; iv) contract additional IT support due to limited institutional resources, which led to higher costs; v) use new technologies (e.g., Epic Registry, Ticket Scheduling) to incorporate PRSs into study workflows; and vi) harmonize institutional workflows and timelines with eMERGE Network goals.

Conclusion: We identified multiple challenges while developing workflows and processes to integrate PRS results and discrete data into the EHR. Many of the challenges were related to embedding a federally funded translational genomics research project into existing clinical, EHR and IT workflows. The lessons learned can inform the integration of PRSs into the EHR in diverse settings to enable their use in primary care.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2150. Clinical cancer and direct-to-consumer genetic test result sharing behavior among US adults: Findings from HINTS 2020

Authors:

S. Shete, S. Peterson, R. Yu, S. Makhnoon; Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract Body:

Introduction: Sharing genetic test results with different stakeholders including family members, healthcare providers, spouse, and friends is a health behavior of clinical importance in genomic medicine. Sharing behavior is perhaps most salient for results of high-risk cancer testing, which should be shared with family members as the information can inform relatives' health risk, cascade genetic testing. In addition, sharing different types of direct-to-consumer (DTC) genetic tests results can be important to raise public awareness of genetic testing. Population representative data on result sharing is sparse and limited by their lack of differentiation among types of genetic tests as sharing have been largely studied as secondary outcomes of genetic research studies. Methods: We use nationally-representative population-based data collected from the Health Information National Trends Survey (HINTS 5, cycle 4) to report prevalence and correlates of genetic test result sharing behavior for high-risk cancer tests, genetic health risk tests, and ancestry tests; for four different groups of individuals: healthcare providers and genetic counselors (HCP/GC), first degree relatives (FDR), spouse/partner, and friend/other. Results: Overall, 89.9% of those who underwent high-risk cancer genetic testing shared their results with HCP/GC. Sharers were more likely to be younger (50.1 vs 59.1 years, $p=0.018$), have \leq high school education (80.1% vs 19.9%, $p=0.036$), have a personal history of cancer (92.8% vs 7.2%, $p<0.001$), and reside in an urban area (80.1% vs 19.8%, $p=0.023$). In the multivariable analyses, adjusted for sex, race/ethnicity, income, family and personal history of cancer, females were 9 times more likely to share than males ($p=0.006$) and those with personal history of cancer were less likely to share than those without ($OR=0.025$, $p<0.001$). Of the genetic health risk test takers, 66.5% shared with HCP/GC, 38.7% with a FDR, 66.6% with a spouse/partner, 12.8% with a friend, and 14.1% did not share results with anyone. Of those who underwent ancestry testing, very few shared results with HCP/GC (2.6%), whereas modest sharing was reported with FDR, spouse/partner, friend (41.3%, 30.4%, and 20.4% respectively). 5.2% of ancestry testers did not share results with anyone. Discussion: These population representative data of sharing different types of genetic test results with various stakeholders is clinically relevant as mediators of outcomes such as proband-mediated cascade genetic testing as well as a metric for public engagement with genetic testing in the US.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2151. Clinical utility of returning all clinically relevant secondary findings from genomic sequencing: A systematic review

Authors:

C. Mighton^{1,2}, S. Dhaliwal^{1,3}, V. Rokoszak^{1,2}, S. Majeed^{4,2,1}, A. Sebastian², V. Aguda^{1,5}, S. Shickh^{2,1}, S. Grewal¹, D. Lightfoot⁶, Y. Bombard^{7,1}; ¹Genomics Hlth.Services Res. Program, St. Michael's Hosp., Toronto, ON, Canada, ²Univ. of Toronto, Toronto, ON, Canada, ³Western Univ., London, ON, Canada, ⁴The Hosp. for Sick Children, Toronto, ON, Canada, ⁵Cardiff Univ., Cardiff, United Kingdom, ⁶Hlth.Sci. Library, St. Michael's Hosp., Toronto, ON, Canada, ⁷Inst. of Hlth.Policy, Management and Evaluation, Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Introduction: Genomic sequencing (GS) is an effective diagnostic test, but secondary findings (SFs) complicate its clinical adoption. Most SF disclosure policies focus on medically actionable SFs. However, in practice, SF disclosure is extremely variable, and laboratories may offer a broader range of results. Furthermore, the majority of policies used to guide decisions on SF disclosure are not evidence-based, as evidence was limited at the time of their development. The field is now at a critical juncture; evidence on the outcomes of SF disclosure has accrued over the past decade, and policies will likely be updated in consideration of this evidence. There is a need for rigorous evidence synthesis and critical appraisal to guide practice and policy.

Aim: To systematically review and synthesize the literature to assess the clinical utility of returning all clinically relevant SFs across all patient populations receiving GS. This will be assessed through: 1) the yield of SFs overall and for major reporting categories (medically actionable, non-medically actionable Mendelian conditions, carrier status, pharmacogenomic results, polygenic risk), 2) clinical outcomes of SF disclosure (e.g., changes in management); and 3) the proportion of patients with phenotypic features or a family history concordant with the condition associated with their SFs.

Methods: A systematic review is being conducted. The review has been registered in PROSPERO (CRD42021254662). A search was conducted in Medline, EMBASE, and Cochrane Databases using MeSH and keywords related to SFs and GS. All articles are screened for eligibility by two reviewers in duplicate. Two reviewers will extract data on bibliographic information, genetic test information (e.g., sequencing type, variant interpretation methods), patient population (e.g., age, disease indication), and study outcomes. Two reviewers will perform quality assessment with critical appraisal tools from the Joanna Briggs Institute. Data on each outcome will be reported narratively and will be combined statistically through meta-analyses if data permit.

Results: The search returned 4401 articles. Title and abstract screening are ongoing, and to date, 173 studies have proceeded to full text screening. Full text screening, data extraction, critical appraisal, and synthesis will be completed by September 2022.

Conclusions: This will be the most up-to-date and comprehensive systematic review of the clinical utility of all types of SFs. This review will provide high-quality evidence which is necessary to inform decisions about SF reporting, disclosure, and reimbursement as GS is increasingly used across medical specialties.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2152. Communicating genetic and socioeconomic factors underlying racial health disparities in medical education

Authors:

P. Orozco Scott, C. Seah, J. Catlett, S. Leisman; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Despite national commitments to anti-racism, a paucity of guidance exists on how to critically appraise genetic, pathophysiological, and epidemiological findings of racial disparities with an anti-racist lens. Interpretation of racial disparities must not reinforce the notion of race as biology, but rather use nuance in attributing evidenced contributors to disease. In genetics in particular, race should not serve as a proxy for genotype, social determinants of health, or systemic healthcare barriers. Medical and graduate education should be explicit in defining race as a socioeconomic variable, taking care to warn against confounding with ancestry, and discussing how ancestral variations in genetic risk do not reflect a biological definition of race. Here, we have developed a framework for inquiry-based restructuring of scientific discussion of race in genetics research and medical education. We systematically probed medical educational materials discussing estimated glomerular filtration rate (eGFR) and HIV-associated nephropathy (HIVAN), two topics where race is routinely used and taught in clinical decision-making. We found improper attributions of disease risk to race, rather than known genetic polymorphisms, and inappropriate omissions of the historical and present systemic factors, such as systemic racism, that contribute to observed racial disparities. Through these two examples, we demonstrate that race is often attributed as a causal factor for disease, and even when genetic causes are known, imprecise approximations of ancestry and confounding of this with race are often used to justify clinical decision making and mnemonics. We suggest that critical appraisal using our inquiry-based process, applied to medical and graduate education, will allow for collaborative reframing of the way race is described in research, and used in clinical decision-making.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2153. Community data-driven approach for generating cross-ethnic population carrier screening panel

Authors:

M. Einhorn, T. Ben-Simhon, Y. Einhorn; Genoox, tel aviv, Israel

Abstract Body:

Background/Objectives: The ACMG has recently published tier-based recommendations for carrier screening. While many pan-ethnic conditions are well established, some genes carry pathogenic founder variants (PFVs) prevalent only in specific ethnic groups, and may therefore not be assigned into the correct tier or not represented at all. Using an interconnected, data-driven repository including a cohort of cases from the Israeli population we created a workflow for assistance in generating a cross-ethnic carrier screening panel. **Methods:** The dataset included de-identified WES data from a cohort of 2792 Israeli patients. Frequencies of pathogenic/likely pathogenic variants were calculated for each subpopulation and compared them with existing carrier screening panels used in Israel. Manual curation was done for suspected variants using evidence from Franklin community members. **Results:** We detected a number of relevant PFVs not included in existing carrier screening panels. These included two variants that should be included in Tier 2 (carrier frequency >1:100) and nine variants for Tier 3 (1:100-1:200). We detected two PFVs of autosomal recessive traits, a PFV in PTPN23, associated with a severe neurodevelopmental disorder (carrier frequency 1:53) as well as a PFV in WFS1 associated with Wolfram syndrome (carrier frequency 1:68). **Conclusion:**Community data-driven and sharing approaches may detect PFVs missing from currently available panels. Moreover, with the increasing use of next-generation sequencing for carrier screening, it is expected to result in a deluge of variants falsely classified as P/LP. Such data-sharing approaches will facilitate proper classification of frequent variants that have no clinical implications, thereby reducing unnecessary workload and patient anxiety

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2154. Community Engagement in genetic and genomic research in a low literate setting: Challenges and perspectives

Authors:

K. Diallo¹, A. Maiga¹, A. Yalcouyé¹, C. Guinto¹, S. Diop¹, I. Ballo², G. Landoure¹; ¹Univ. of Sci., Techniques and Technologies of Bamako, Bamako, Mali, ²Univ. of Letters and Human Sci. of Bamako (ULSHB), Bamako, Mali

Abstract Body:

Introduction: Community engagement (CE) is a process of working collaboratively with a group of persons on a common interest, and is recognized as an important component of any healthcare project. Such collaboration is particularly beneficial in genetics and genomics research (GGR) for the development of genuine partnerships with communities to build trust and incorporate local perspective into the design and conduct of research. Achieving this in a low literacy population setting is particularly challenging. **Objectives:** To address issues related to implementing community engagement strategies in genetics and genomics research in a low literacy setting **Methods:** A multidisciplinary team including geneticists, epidemiologists, anthropologists, psychologists, and CE specialists was set up and held information and education workshops and sessions throughout the countries. Interviews, focus group discussions and exchanges during the consent process were used to engage patients, their relatives and community. Both modern (radio and TV broadcasts) and traditional communication tools were used to vehiculate the purpose, the approaches, and the implications of GGR in the Malian context. The place of consanguineous and intra ethnic marriages in the occurrence of recessive diseases were highlighted **Result:** Our preliminary results showed that participants and field staff seeking consent were overall satisfied with their understanding of the study project even though most of the participants were not familiar with the concepts of GGR. Participants were less aware of the methodologies used during GGR and their implications, such as sharing the participants clinical and genomic data and the use of these data for potential secondary research. In addition, it was more difficult than anticipated to make a clear distinction between therapy and research especially those coming to healthcare centers for cure. One major issue was the language barrier. In fact, as many genetic words do not exist in our local languages, translating them was challenging. **Conclusion:** Until recently, GGR were less regarded in developing countries, overshadowed by communicable diseases. Hence, community engagement is critical in enrolling participants and in disseminating research findings for the global benefits of the community. In addition, a collaboration with local languages specialists is necessary to develop a dictionary of genetic words. Capacity building for local ethics committees to enable appropriate consideration of ethical issues raised by genomic research is likely to be necessary in many countries with limited experience of this type of research.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2155. Considerations for assessing and optimizing remote participant recruitment for a pragmatic trial of polygenic risk score testing

Authors:

C. Brunette¹, A. Antwi¹, E. Harris^{1,2}, K. MacIsaac^{1,2}, T. Yi¹, M. Danowski¹, J. Vassy^{1,2,3,4}; ¹VA Boston Hlth.care System, Boston, MA, ²Harvard Med. Sch., Boston, MA, ³Brigham and Women's Hosp., Boston, MA, ⁴Precision Population Hlth., Ariadne Labs, Boston, MA

Abstract Body:

Background: The translation of genomic medicine innovations from concept to care relies on valid and generalizable evidence, the generation of which hinges on the efficient recruitment of research participants. Here we use data from the Genomic Medicine at VA (GenoVA) Study, a pragmatic trial of polygenic risk score testing at the Veterans Affairs (VA) Boston Healthcare System, to characterize recruitment outcomes among 4499 potential research participants.

Methods: Data were recorded in real-time using a VA-developed clinical trial informatics system. Data associated with recruitment letter distribution and phone outreach between 6/1/20 and 5/16/22 involving 4531 letters and 9968 phone calls to Veterans aged 50-70 were assessed. We used Markov chains to model participant transition probabilities across 10 phone outreach attempts and 4 decision states: undecided, eligible, opt out, and accept.

Results: Overall, 723 (16.1%) participants completed an eligibility screen and accepted a study invitation, 1046 (23.2%) opted out of further contact either before or after screening, and 2730 (60.7%) were not consistently reached or remained undecided. On average, participants were called 2.2 times (range 1-20), study invitations were accepted within 4.3 calls (range 2-20) and opt out decisions were made after 1.9 calls (range 1-14). We observed higher probabilities of invitation acceptance among women (9.1% v. 7.8%), non-white/Hispanic (13.3% v. 12.8%), and participants 60 or older (21.0% v. 19.0%) than expected. End of week calls (Wednesday-Friday) yielded better invitation acceptance rates (7.4-8.5%) compared to early week calls (Monday-Tuesday; 6.0%). August, December, and January yielded the worst acceptance to opt out ratios (0.3-0.5) compared to other months (\bar{x} = 0.8). Participants were most likely to opt out (0.15, 95% CI 0.14-0.16; 0.05, 95% CI 0.04-0.06) or complete the eligibility screen (0.18, 95% CI 0.17-0.20; 0.10, 95% CI 0.09-0.11) within the first two call attempts. Opt outs occurring post-screening were low, but consistent up to the eighth call attempt (range 0.02-0.04). Invitation acceptance most often occurred between the second and seventh calls (range 0.07-0.19). Only 1.7% (95% CI 0.00-0.09) of undecided or difficult to reach participants transitioned to any other state after the fourth call attempt.

Discussion: Assessing recruitment outcomes across demographic, seasonal, and longitudinal factors is important to the operational efficiency of large genomic medicine research initiatives. Structuring outreach that focuses on optimal timing and direction of labor can help maximize participant engagement and recruitment efforts.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2156. Considerations for embedding genomic discovery research in clinical settings: Perspectives from medical geneticists

Authors:

K. Muenzen¹, H. A. Burkhardt¹, N. Reid¹, P. Tarczy-Hornoch¹, D. W. Tsuang², A. T. Chen¹, G. P. Jarvik³, D. R. Crosslin⁴; ¹Univ. of Washington, Seattle, WA, ²VAPSHCS, Seattle, WA, ³Univ. of Washington Med. Ctr., Seattle, WA, ⁴Tulane Univ. Sch. of Med., New Orleans, LA

Abstract Body:

Background: Accelerating the discovery and clinical uptake of new genomic knowledge is important for advancing genomic medicine. Adoption of new knowledge into clinical practice is hindered in part by separate norms and regulations dictating standard practices in research and clinical care. Here we explore considerations for conducting genomic discovery research within clinical environments from the perspectives of medical geneticists. **Methods:** In this IRB-approved study, we conducted 1-hour, semi-structured, remote interviews with 20 board-certified medical geneticists from the Electronic Medical Records and Genomics (eMERGE), Clinical Sequencing Evidence-Generating Research (CSER), and University of Washington (UW) networks. Interviews investigated the feasibility, utility, and challenges of clinical genomic discovery. Using Grounded Theory methods, we analyzed interview transcripts by conducting line-by-line, axial, and theoretical coding to synthesize themes that emerged from interviews. **Results:** The sample was largely white and had equal male and female representation. Participants expressed high hopes for the utility of integrating genomic research into clinical settings, but warned that feasibility was limited by the challenges of using clinical data for research, questions of consent, funding, concerns about clinical misuse of genetic testing results, healthcare system fragmentation, and conflicting goals for patients, research participants, clinicians, and researchers. Suggested requirements for implementing clinical genomic research included rigorous actionability and validity assessments of new discoveries, strong community engagement, evolved consent models, enhanced sharing and cleaning of clinical and genomic data, novel funding schemas, defined roles and responsibilities for clinicians and patients, and standards for returning research results to patients. However, participant views on how to address each of these requirements often conflicted with one another, highlighting the complexity of the issue and the challenge of developing consensus recommendations. **Conclusions:** While the potential benefits of merging research and clinical genomics are clear, the practical, ethical, social, and political considerations are extraordinarily complex. This work provides a baseline reference for the potential benefits and challenges that should be considered when integrating genomics research into clinical environments.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2157. Cultivating a data-driven computational culture within biomedical institutions by empowering graduate students with code-based data science skills.

Authors:

C. Ronkowski¹, K. Peng¹, D. Yu¹, R. Ayyala¹, S. Knyazev², M. Wu¹, I. Haworth³, J. Zaro³, S. Mangul¹; ¹Univ. of Southern California, Sch. of Pharmacy, Titus Family Dept. of Clinical Pharmacy, Los Angeles, CA, ²Univ. of California, Los Angeles, David Geffen Sch. of Med., Dept. of Pathology and Lab. Med., Los Angeles, CA, ³Univ. of Southern California, Sch. of Pharmacy, Dept. of Pharmacology and Pharmaceutical Sci., Los Angeles, CA

Abstract Body:

To address an emerging disparity between biomedical researchers' lack of computational prowess and the rising importance of massive omics datasets in translational and clinical research, we created a 2-unit graduate-level course to teach reproducible data science to Pharmaceutical and Translational Sciences (PHTS) graduate students at the University of Southern California. In our course, we emphasized data analysis, data visualization, and open data science practices using code-based assignments written within a free online Jupyter notebook environment. This framework enabled us to combine the benefits of literate programming and cloud computing to promote scientific transparency in a manner accessible to anyone with an internet connection. In addition, we implemented simplified explanations of programming fundamentals to gradually ease students into using Python libraries, empowering them with the foundational knowledge necessary to troubleshoot one's own code. We did so as a significant deterrent to adopting code-based data science solutions is the frustration of not being able to resolve errors, leading many biomedical researchers to resort to using poorly reproducible, Graphical User Interface (GUI)-based software. Moreover, we measured the effectiveness of this strategy using an anonymous Institutional Review Board (IRB)-approved survey conducted during the first and final lectures. Our findings show greater self-reported confidence, increased comprehension of the course material, and an agreement that our course was highly impactful. Based on these results, we conclude that our course framework and teaching approach represents an effective strategy for introducing the fundamentals of modern biomedical data science to students with minimal prior computer science background.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2159*. Delivery of an unfavorable genetic diagnosis: Developing a methodology framework

Authors:

O. Chorin^{1,2}, L. Sheelo Salzer^{3,2}, A. Koifman^{4,5}, B. Pode-Shakked^{1,6}, A. Raas-Rothschild^{1,6}, M. Tsur^{7,8,9}, I. Maya^{3,6}, ¹The Inst. of Rare Diseases, Edmond and Lily Safra Children's Hosp., Sheba Med. Ctr., Ramat Gan, Israel, ²Sackler Faculty of Med., Tel Aviv Univ., Tel Aviv, Israel, ³Raphael Recanati Genetics Inst., Rabin Med. Ctr., Petah Tikva, Israel, ⁴Med. Genetics Service, Assuta Univ. Med. Ctr., Ashdod, Israel, ⁵Faculty of Hlth.Sci., Ben Gurion Univ. of the Negev, Beer Sheva, Israel, ⁶Sackler Faculty of Med., Tel-Aviv Univ., Tel Aviv, Israel, ⁷Mitchell- Hamline Sch. of Law, Saint Paul, MN, ⁸Cardozo Sch. of Law, Yeshiva Univ., New York City, NY, ⁹Faculty of Law, Hebrew Univ., Jerusalem, Israel

Abstract Body:

Background/Objectives: Genetic counseling consists of the crucial aspect of conveying and discussing unfavorable genetic results. This event is experienced as difficult for the consulting physician, the patients, and the families. There is no structured methodology for delivering an unfavorable genetic result. **Methods:** We collaborated with "Shakla & Taria", an Israeli institute for negotiation and difficult conversations, to develop a structured methodology for delivering unfavorable genetic results in a genetic counseling setting. The principles and the different stages of the method were introduced to a team of medical geneticists and processed to form an agreed generic methodology taught to clinical geneticists in designated workshops. **Results:** The methodology consists of a preliminary questionnaire filled out by the geneticists and aims to improve the level of emotional preparation for the meeting. Additionally, preplanned details are addressed, including the timing, setting, desired participants, and atmosphere of the genetic medical encounter. The methodology for the conversation itself is structured, with a stepwise gradual delivery of the unfavorable genetic result, while affirming the patient's desire to receive the data at every stage and allowing the patient to lead the conversation. These stages are conducted sensitively and professionally, focusing on the advantages of the diagnosis and only later delving into additional medical details and offering future planning possibilities, including options for prenatal diagnosis. The last, most crucial stage consists of affirming the understanding of the data by the patient and offering the opportunity for a second meeting within a few weeks to ensure comprehension and allow for follow-up questions and care. The methodology requires delivering information while maintaining patients' autonomy, feelings, and identity inside the room and during the day after. The stages include a declaration of intent, preparation for the result, delivering the outcome itself and the medical diagnosis, discussion about the diagnosis advantages, more informative medical information, and future implications. The conversation ends after verification of the patient's understanding. **Conclusion:** We constructed a structured methodology for conveying unfavorable genetic results. We plan to incorporate these single-day workshops into the residency program for medical geneticists, genetic counselors, and other professionals in the field. Further studies are required to assess the overall satisfaction and well-being of both the patients and the geneticists from the medical genetic encounter.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2160. Designing a Comprehensive Education Program for Participants, Providers, and Research Teams: The eMERGE Consortium Experience

Authors:

J. Connolly¹, E. Berner², M. Smith³, A. Allworth⁴, P. S. Applebaum⁵, H. Aungst⁶, H. Bangash⁷, S. T. Bland⁸, R. L. Chisholm⁹, K. Dufendach⁶, M. Hamed⁷, M. Harr¹, I. Holm¹⁰, D. Karavite¹¹, A. Khan⁵, N. Limdi¹², C. Liu⁵, A. Miller⁷, K. F. Mittendorf¹³, C. Nelson⁶, S. Nigbur⁷, T. Schmidlen¹⁴, A. Sherafati⁷, E. Soper¹⁵, S. Suckiel^{16,17}, S. Terek¹, W-Q. Wei¹⁸, H. Hakonarson^{1,19,20,21}, M. Sabatello⁵; ¹Ctr. for Applied Genomics, Childrens Hosp. of Philadelphia, Philadelphia, PA, ²Dept. of Hlth.Services Admin., Univ. of Alabama, Birmingham, AL, ³Ctr. for Genetic Med., Dept. of Med., Northwestern Univ., Chicago, IL, ⁴Univ. of Washington, Seattle, WA, ⁵Columbia Univ., New York, NY, ⁶Cincinnati Childrens Hosp. Med. Ctr., Cincinnati, OH, ⁷Cardiovascular Diseases, Mayo Clinic, Rochester, MN, ⁸Ctr. for Precision Med., Dept. of BioMed. Informatics, Vanderbilt Univ. Med. Ctr., Lebanon, TN, ⁹Northwestern Univ, Chicago, IL, ¹⁰Div. of Genetics and Genomics, Boston Childrens Hosp., Boston, MA, ¹¹Dept. of BioMed. and Hlth.Informatics, Childrens Hosp. of Philadelphia, Philadelphia, PA, ¹²Dept. of Neurology, Heersink Sch. of Med., Univ. of Alabama at Birmingham, Birmingham, AL, ¹³Vanderbilt-Ingram Cancer Ctr., Nashville, TN, ¹⁴Med. Affairs, Invitae, San Francisco, CA, ¹⁵Icahn Sch. of Med. at Mount Sinai, New York, NY, ¹⁶Inst. for Genomic Hlth., Mount Sinai, New York City, NY, ¹⁷Div. of Genomic Med., Dept. of Med., Mount Sinai, New York, NY, ¹⁸Ctr. for Precision Med., Dept. of BioMed. Informatics, Vanderbilt Univ., Nashville, TN, ¹⁹Dept. of Pediatrics, The Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ²⁰Div. of Human Genetics, Childrens Hosp. of Philadelphia, Philadelphia, PA, ²¹Div. of Pulmonary Med., Childrens Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Polygenic risk scores (PRS) have the potential to improve healthcare by identifying individuals at risk for common, complex conditions. Use of PRS in clinical practice, however, requires careful assessment of the needs of patients, providers, and capabilities of healthcare systems. The electronic Medical Records and Genomics (eMERGE-IV) network is conducting a collaborative study, which will return PRS to 25,000 pediatric and adult participants. All participants will receive a risk report, potentially classifying them as high risk (top 2-10% per condition) for one or more of 10 conditions based on PRS. The study population is enriched by participants from diverse groups, including racial and ethnic minority populations, underserved populations, and populations who experience poorer medical outcomes. The size, scope, and novelty of the study necessitates developing comprehensive and diverse educational resources to address the needs of providers (particularly primary care providers) and participants from various backgrounds. All 10 eMERGE-IV clinical sites conducted focus groups, interviews, and/or surveys to understand the educational needs of eMERGE-IV's key stakeholders - participants, providers, and/or study staff. Together, these studies highlighted the need to develop tools that address the perceived benefit/value of PRS, types of education/support needed (including family support), accessibility of the PRS risk report, and PRS-related knowledge, understanding, and misperceptions. Based on these findings, the network followed processes to harmonize the educational initiatives across clinical sites, developed extensive training educational material and formal/informal resources, and tested them through stakeholder and community engagement. This poster summarizes the eMERGE-IV consortium's collective approach to 1) assessing stakeholders' educational needs and 2) developing educational approaches for its primary stakeholders. It addresses the challenges encountered, and solutions provided, by a large-scale and novel clinical trial and summarizes lessons-learned for similarly ambitious programs.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2161. Developing Recommendations for Direct-to-Consumer Genetic Testing

Authors:

H. Ayoubieh¹, ISCC-PEG DTC-GT Working Group, D. Messersmith², T. Weiler³; ¹Texas Tech Hlth.Sci. El Paso, El Paso, TX, ²NHGRI, Bethesda, MD, ³FIU - HWCAM, Miami, FL

Abstract Body:

Direct-to-consumer (DTC) genetic testing (GT) for health-related conditions has become easily accessible to the general public after the US Food and Drug Administration cleared a commercial company to return health reports directly to patients with information about specific pathogenic genetic variants and pharmacogenomic variants. To navigate these health reports and discuss concerns related to their results, DTC companies suggest that consumers consult with their healthcare professional. Primary care and non-genetics specialty clinicians describe encounters with patients who have questions about DTC-GT and are interested in enhancing their genetic knowledge. However, currently there are no practice recommendations for healthcare practitioners related to the application of DTC-GT. The DTC-GT Project Group of the Inter-Society Coordinating Committee for Practitioner Education in Genomics (ISCC-PEG) at the NHGRI presented a draft set of recommendations at the American College of Medical Genetics 2022 conference. The audience was asked to rate their level of agreement with the proposed recommendations and provide open-response feedback. Over 700 participants responded to the survey. The survey included Likert scale and open-ended questions. Participant responses highlight areas of agreement and disagreement. There is controversy regarding the DTC-GT-related practice expectations for both genetics and non-genetics healthcare professionals. For non-geneticists, detailed recommendations may be too onerous to recognize and implement during a brief clinical encounter. Education is needed regarding the technology used for testing by the different DTC-GT companies, how it differs from clinical testing, how different analytical tools and reporting criteria are being used to generate the reports, and how in-depth clinical evaluation and family history intake is required to appropriately counsel patients. Additionally, limiting recommendations to selected positive DTC-GT results could result in incomplete evaluations, and false negative DTC-GT-results could incorrectly reassure patients. There is also concern that these recommendations would create a bottleneck in genetic services for DTC related referrals vs referrals where genetic evaluations may be more critical and appropriate. Moreover, genetic professionals are not typically involved in pharmacogenomic testing and relevant medication prescribing. A significant matter of contention is where and how DTC-GT results should be included in the electronic medical records, as well as the legal, policy and insurance implications of the suggested recommendations.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2162*. Development and utility of a clinical research informatics application for participant recruitment and management for a return of results pilot trial in the Million Veteran Program

Authors:

T. Yi¹, C. Brunette¹, M. Danowski¹, M. Cardellino¹, J. Vassy^{2,1,3,4}; ¹VA Boston Hlth.care System, Boston, MA, ²Harvard Med. Sch., Boston, MA, ³Brigham and Women's Hosp., Boston, MA, ⁴Precision Population Hlth., Ariadne Labs, Boston, MA

Abstract Body:

Background: Clinical research informatics (CRI) is a burgeoning domain focused on using information technologies to optimize the design and conduct of translational research. In parallel, the development of workflow processes associated with re-engaging biobank participants has become necessary as genomic repositories increasingly consider the return of actionable research results. Here we describe the development and utility of a CRI application for participant recruitment and enrollment management for the Veterans Affairs (VA) Million Veteran Program Return Of Actionable Results (MVP-ROAR) Study, a randomized controlled trial returning genetic results associated with familial hypercholesterolemia. **Methods:** We leveraged the robust VA Informatics Computing Infrastructure (VINCI) to develop an end-to-end tracking system for re-contacted MVP biobank participants. The application is housed within a secure virtual server, is password protected, and accessible only to study staff. The multi-user application integrates electronic data capture, nationwide clinical and administrative data, and dashboard analytics via a combination of commercial and open-source software. Study data are queried and stored within the VA Corporate Data Warehouse, easing access and eventual trial data analysis. We use a task-based model to transition participants through the system as research staff complete study tasks in real-time. The application is developed in Python and uses the Flask microframework, providing study staff access via a standard web browser. **Results:** The application was put into production in November 2021 and includes modules for chart review, medication reconciliation, participant contact and biospecimen logging, survey recording, randomization, and documentation of genetic counseling and return of results. As of 5/18/22, a total of 2 primary users, a genetic counselor and research coordinator, and 254 Veteran participants have been integrated into the system. The application has successfully handled 1,668 task requests involving greater than 50,000 structured data points. Specifically, application users have recorded 201 chart reviews, 408 recruitment telephone calls, 84 telephone-based surveys, and 36 return of results genetic counseling sessions, among other available study tasks. **Discussion:** The development of usable, customizable, and secure CRI tools will become increasingly important as large genomic repositories begin to return research results at scale. Our work provides a proof-of-concept for using such tools to aid in managing the return of results process within a national biobank.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2163. Do Automated Family Histories Significantly Improve Risk Prediction in an EHR?

Authors:

X. Huang¹, D. Page², S. Hebring³; ¹Yale Univ., New Haven, CT, ²Duke Univ., Durham, NC, ³Marshfield Clinic, Marshfield, WI

Abstract Body:

We recently demonstrated that electronically constructed family pedigrees (e-pedigrees) have great value in epidemiologic research using electronic health record (EHR) data. Prior to this work, it has been well accepted that family health history is a major predictor for a wide spectrum of diseases, reflecting shared effects of genetics, environment, and lifestyle. With the widespread digitalization of patient data via EHRs, there is an unprecedented opportunity to use machine learning algorithms to better predict disease risk. Although predictive models have previously been constructed for a few important diseases, we currently know very little about how accurately the risk for most diseases can be predicted. It is further unknown if the incorporation of e-pedigrees in machine learning can improve the value of these models. In this study, we devised a family pedigree-driven high-throughput machine learning pipeline to simultaneously predict risks for thousands of diagnosis codes using thousands of input features. Models were built to predict future disease risk for three time windows. At baseline models Logistic Regression and XGBoost without e-pedigree data, we achieve average areas under the receiver operating characteristic curve (AUCs) of 0.799, 0.747 and 0.710 for 1, 6, and 24 months, and AUCs of 0.820, 0.774 and 0.713 for 1, 6, and 24 months, respectively. When adding e-pedigrees features to the machine learning pipelines, AUCs of Logistic Regression increased to 0.808, 0.761 and 0.730, and AUCs of XGBoost increased to 0.826, 0.788 and 0.736 for the same three time periods, respectively. This result emphasizes the potential value of incorporating family health history into machine learning with no further human time or input for disease risk prediction.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2164. Do digital tools support a diversity of patient values? A qualitative study using the Genetics Adviser.

Authors:

S. Krishnapillai¹, M. Clausen², D. Hirjikaka², R. Kodida², C. Mighton³, S. Shickh⁴, E. Reble⁵, J. Sam⁵, E. Adi-Wauran^{6,1}, N. Baxter⁷, J. Carroll^{8,9}, A. Eisen¹⁰, A. Eisen¹¹, C. Elser¹², G. Feldman², E. Glogowski¹³, R. Kim¹⁴, J. Lerner-Ellis⁸, A. Scheer², S. Shastri-Estrada², C. Shuman¹⁵, E. Seto¹, K. Schrader¹⁶, Q. Pham^{1,12}, H. Faghfoury¹⁷, Y. Bombard^{1,2,18}; ¹Inst. of Hlth.Policy, Management and Evaluation, Univ. of Toronto, Toronto, ON, Canada, ²Li Ka Shing Knowledge Inst., St. Michael's Hosp., Toronto, ON, Canada, ³St. Michael's Hosp. & Univ. of Toronto, Toronto, ON, Canada, ⁴Univ. of Toronto & St. Michael's Hosp., Toronto, ON, Canada, ⁵St. Michael's Hosp., Toronto, ON, Canada, ⁶St. Michael's Hosp., Toronto, ON, Canada, ⁷Melbourne Sch. of Population and Global Hlth., Univ. of Melbourne, Melbourne, Australia, ⁸Mount Sinai Hosp., Toronto, ON, Canada, ⁹Univ. of Toronto, Toronto, ON, Canada, ¹⁰Potomac, MD, ¹¹Odette Cancer Ctr., Sunnybrook Hosp., Toronto, ON, Canada, ¹²Univ. Hlth.Network, Toronto, ON, Canada, ¹³Mem Sloan Kettering Cancer Ctr, New York, NY, ¹⁴Univ. Hlth.Network/Mt Sinai Hosp., Toronto, ON, Canada, ¹⁵Hosp for Sick Children, Toronto, ON, Canada, ¹⁶Univ British Columbia, Vancouver, NY, Canada, ¹⁷Mount Sinai Hosp./UHN, Toronto, ON, Canada, ¹⁸Ontario Inst. of Cancer Res., Toronto, ON, Canada

Abstract Body:

Background: The shared decision-making model involves patients and clinicians. Patient facing digital health can serve as a valuable adjunct to this process. Despite the well-established role of patient values in shared decision-making, there have been limited efforts to explore the range of values held by populations and how those values inform decision-making for genomic testing. Choosing to receive incidental findings (IFs) from exome sequencing (ES) is inherently complex, in part given the value-laden nature of the decision. The Genetics Adviser is a digital health application that supports patients undergoing genomic testing. **Aim:** To explore the range of values raised when deciding upon IFs to receive from ES and how these values are supported by the Genetics Adviser. **Methods:** We conducted semi-structured qualitative interviews with participants undergoing ES and choosing which IFs to receive as part of the Genetics Adviser RCT. Interviews were analyzed using interpretive description methods. **Results:** Sixteen participants were interviewed, primarily women (11/16), mean age 59 (range: 39-76 years). Participants unanimously expressed that their decision to receive all IFs available to them from ES preceded use of the Genetics Adviser. Decision-making was not described as a deliberate process for participants. Instead, their decision was predetermined and occurred within a 'black box' before using the Genetics Adviser. This 'black box' describes a complex system of motivational factors involved in decision-making. Three overarching values emerged from unpacking this 'black box', including: a sense of family stewardship, an imperative to accumulate information, and the notion of restoring and maintaining personal agency. While the Genetics Adviser did not play an active role in eliciting these values, participants appreciated having the tool to confirm their predetermined decision by clarifying, contextualizing, and instilling confidence in their decision. **Significance:** Participants entered decision-making for ES with a predetermined decision guided by three core values that motivated their decision to learn IFs: family stewardship, an information imperative, and preserving agency. The Genetics Adviser supported patients' decision-making by clarifying, contextualizing and reinforcing their predetermined decision. Further research is needed with more diverse populations including those that are frequently underserved in clinical genetics and have been historically excluded in genomics research to explore how digital tools can support a diversity of values.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2165. Early Mentoring for Medical Students in Research of Medical Genetics

Authors:

O. Samassekou¹, G. Landoure², C. Guinto¹, M. traore¹, L. CISSE³, T. Coulibaly¹; ¹USTTB, Bamako, Mali, ²CHU du Point G, Bamako, Mali, ³Teaching Hosp. of Point G, Bamako, Mali

Abstract Body:

Introduction: Research in medical genetics is at its infancy in developing countries such as Mali where a lack of human resources is the main cause of its slow. In investing in training to encourage youngsters on the path of research in medical genetics, we can reduce the gap between developed and developing countries for doing research in this field. Our research team has benefited from the H3Africa initiative, financed by the NIH, to study the genetic aspects of hereditary neurological disorders in Mali. Using some of these research funds, we have initiated a mentoring program by hypothesizing that medical genetics mentoring program for medical students would be a stepping stone to introducing them to research and opening up career opportunities for them. **Methodology:** From 2014 to 2021, we recruited graduate medical students doing thesis projects during their last two years of general medical residency at the University of Sciences, Techniques et Technologies of Bamako, Mali. We selected students based on their academic record and their motivation. We used as evaluation tools of the program, the number of scientific communications, the number of published articles, the number of awards obtained, the acquired skills, and the student satisfaction for the program. **Results:** We mentored 27 medical residents, 3 master's students, and 2 doctoral students. On average, the mentorship lasted 3 years. The students did more than 100 communications in scientific meetings, published 31 scientific papers, and won 34 awards. Among the skills acquired, in addition to theoretical and practical knowledge in medical genetics research, they have improved skills in mastering the English language, technique of scientific communications, and leadership. Over 90% of the students are pursuing a career in medical genetics or a related discipline. The students have been very satisfied with the program and want its continuation. **Conclusion:** We have implemented a mentoring program to introduce medical students to research in medical genetics. This program has had notable results that need to be encouraged and supported.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2166. Electronic health record-based recruitment for genomic research studies

Authors:

A. Miller, S. Nigbur, J. Petrzela, H. Bangash, S. Hussain, P. Caraballo, R. Freimuth, I. Kullo; Mayo Clinic, Rochester, MN

Abstract Body:

Introduction: The use of the electronic health record (EHR) as a tool for recruitment has come into focus since the onset of the COVID-19 pandemic. We established a set of EHR-based workflows to enroll patients for the Electronic Medical Records and Genomics (eMERGE) Network Phase IV project that is implementing polygenic risk scores in the clinical setting. **Methods:** A customized query of the EHR (Epic) is run weekly to identify individuals who meet study inclusion criteria (aged 3 to 74, resident of Olmsted County with a primary care provider at Mayo Clinic, and an upcoming clinical appointment). Demographic information is used to prioritize enrollment of minority participants. Study invitation messages are sent to eligible participants through the patient portal and interested patients are emailed the study consent form to review. Once consented, participants receive an email notification to complete additional study surveys in REDCap. Participants can self-schedule biospecimen collection for genetic testing and return of result appointments. Clinical variables needed for generating the study report are collected using electronic phenotyping algorithms and sent to the coordinating center via an Application Programming Interface (API). A study specific order is placed in the EHR for subsequent report upload. Discrete study data are transferred as a comma separated values (.csv) file from the Network REDCap instance to the site EHR using custom clients and institutional APIs to load the data into Epic SmartData Elements and Epic Registry to track recruitment metrics. **Results:** Recruitment workflow design began in October 2020 and was implemented in May 2022. Study investigators interacted with 28 teams and received 7 committee/group approvals for institutional support of the workflows. These teams, committees, and groups provided feedback to the study team and helped finalize the workflow outlined above. Participant recruitment began in May 2022 with the goal of enrolling 1900 adult and 100 pediatric participants over 18 months. The EHR query identified, on average, ~2000 eligible individuals in a week. A subset is sent invitations, depending on study needs. As of 6/9/2022, 1039 adults have been invited and included 62.8% female, 88.3% non-Hispanic White, with ages ranging from 18 to 75; 139 completed the prescreen survey and 39 consented to the study thus far. **Conclusion:** We describe an EHR-based recruitment workflow that includes informed consent and can be adapted for diverse research studies while maintaining privacy and confidentiality. A limitation is potential bias towards enrolling more educated, more technologically literate participants.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2167. Engagement and Feedback on Genetic Ancestry and Trait Results from >100,000 All of Us Participants

Authors:

H. Hoban¹, D. Brazel¹, C. Neben¹, C. Mbumba¹, A. Myers¹, A. Ondov¹, B. Ellis², L. Westendorf², A. Wise², S. Topper¹, A. Zhou¹, All of Us Research Program; ¹Color Hlth., Burlingame, CA, ²NIH, Bethesda, MD

Abstract Body:

The All of Us Research Program is committed to 1) creating a cohort that reflects the diversity of the US, with individuals who have been underrepresented in biomedical research (UBR), and 2) providing participants with access to their data. In November 2020, the program launched the Genetics Engagement Module (GEM), a digital platform where participants can choose to receive and view their genetic ancestry and four non-clinical trait results.

To date, 220,410 participants consented to receive their genetic results (73% UBR). Of those, 136,730 participants were eligible to receive GEM results (71% UBR). Of those, 67% elected to receive their results (67% UBR). 95% of those participants viewed their genetic ancestry results; 88% viewed at least one trait result, and 74% viewed all four trait results.

Participants provided feedback about their results through Likert-scaled questions and free text. 67% answered at least one question, and overall feedback was positive. Participants agreed with their trait results on average 4.2/5, but those levels varied, with earwax type the highest (4.7/5) and cilantro preference the lowest (3.4/5). They found their genetic ancestry and trait results very informative (3.3/4) and stated that they would like more information, specifically more precise locations for their ancestors. This matched with the 39% of participants who expressed wanting to learn more about their genetic ancestry. For trait results, participants stated their lactose intolerance results matched their lived experience best and their cilantro preference result the least, which is not surprising as cilantro preference is more heavily influenced by other genetic and non-genetic factors. Participants stated their bitter taste trait results were the most confusing because they lacked context on how other individuals' perceive taste, but they stated they "learned something" about their earwax type results, suggesting this trait was new to most participants. Finally, participants were asked if they had any additional feedback. Most participants stated that they did not, and many thanked the program for giving them access to their information. Taken together, these data show that All of Us participants are highly engaged and readily provide feedback about their genetic results in GEM. Participant interaction and satisfaction across all UBR groups drives recruitment and retention, increasing the value of the program's research database. To truly develop individualized health care based on that data, it is imperative to continue engaging with and collecting feedback from participants in this diverse cohort.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2168*. Engaging a local community & participant advisory board in a national precision medicine study: Experiences from the New York City Consortium *All of Us* Research Program.

Authors:

L. Bier¹, C. E. W. Freeland¹, J. Brown², L. Jones¹, G. M. Reyes³, L. Turner¹, A. Fuentes³, A. G. Gharavi¹, D. B. Goldstein¹, G. Hripacsak¹, R. Kaushal³, M. E. Ross³, **R. K. Trousdale**², E. G. Cohn⁴; ¹Columbia Univ. Irving Med. Ctr., New York, NY, ²NYC Hlth.+ Hosp./Harlem, New York, NY, ³Weill Cornell Med., New York, NY, ⁴Hunter-Bellevue Sch. of Nursing, City Univ. of New York, New York, NY

Abstract Body:

The All of Us Research Program (AOURP) is a nationwide NIH precision medicine initiative, which aims to enroll 1 million or more people across the United States to accelerate health research. AOURP comprises over 100 awardees and subawardees, and has a robust national governance structure which incorporates over 30 participant partners across the country. Even while participant representation is woven throughout the national program infrastructure, local community and participant engagement remains valuable, particularly as precision medicine research touches on topics around genetic research which may have complex local sociocultural implications. The New York City Consortium (NYCC), part of the AOURP, established a local Community and Participant Advisory Board (CPAB) to provide guidance and partnership around all aspects of implementing the AOURP in New York City. Here, we describe our experience of establishing a local CPAB within the context of a complex national governance structure. Initial CPAB invitations were extended in 2018 to community members with pre-existing relationships with study investigators, as well as AOURP participants nominated by local enrollment staff, with the aim of establishing a board that reflects a diversity of perspectives including social, ethnic, economic, faith-based and sexual-gender minority perspectives. The CPAB consists of approximately 15 members (membership changes over time), and meets every quarter with meetings still ongoing. Four strategies have been employed to foster engagement between the CPAB and NYCC research team, and maximize the impact of the CPAB within the AOURP: 1) Work with the CPAB to jointly set scope and goals, and develop the Board Charter; 2) Establish streamlined channels for bidirectional communication between the local CPAB and the national AOURP (both through NYCC researchers and National AOURP Participant Ambassadors); 3) Promote a reciprocal relationship between the NYCC and CPAB members, encouraging creative ways for CPAB members to contribute to NYCC activities (e.g., CPAB-led staff trainings, and CPAB feedback on questions, concerns and interests around genetic research), and how NYCC can contribute to community work of CPAB members (e.g., raising awareness of community services and events); 4) Proactively identify opportunities for the CPAB to provide feedback and guidance to national AOURP representatives. Using these strategies, the NYCC CPAB serves as a model for how a local CPAB can effectively work within a complex national program.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2169. Engaging Families and Clinicians to Advance Care Coordination and Improve Health Outcomes in Genetic Disease: The Long-Term Follow-Up Cares and Check Initiative

Authors:

K. Chan, Y. Unnikumaran, Z. Talebizadeh, R. Wiebenga, L. Barnes, C. Lumpkins, J. Taylor, A. Brower; American Coll. of Med. Genetics and Genomics, Bethesda, MD

Abstract Body:

The use of technologies to screen, diagnose, and treat newborns has accelerated a model of population-based genetic screening for diseases that benefit from early identification, are treatable but not curable, and require lifelong care and management. Each year in the United States the neonatal screening about 4 million babies leads to the identification of over 13,000 infants with a genetic condition. A key component of the NBS system is care coordination across the lifespan. In addition, some, but not all, of the genetic diseases that are identified through NBS require urgent referral for diagnosis and clinical evaluation. This requirement for rapid diagnosis coupled with life-long medical care requires coordination between a diverse group of stakeholders including birthing hospitals, state NBS programs, clinicians, and families. Although longitudinal follow-up that supports screening, counseling, and service delivery to individuals diagnosed with a condition through NBS has been a goal in the United States for many years, NBS screening is state-based with individual policies and practices that impact LTFU. For diagnosed infants, families and clinicians work together to manage care over the lifespan and the input of these two stakeholders is essential to ensuring LTFU. To develop a model of LTFU, the Health Resources and Services Administration funded the American College of Medical Genetics and Genomics to conduct a two-year effort focused on LTFU of newborns diagnosed with spinal muscular atrophy (SMA) through NBS. The Long-Term Follow-Up Cares and Check Initiative (LTFU Cares & Check) prioritizes the involvement of families and clinicians to inform which data points inform health outcomes and well-being over time. The project's overall goal is to expand the ability of state public health agencies to collaborate with families and clinicians to assure the best possible outcome for individuals with SMA. This model system consists of three main tools: (1) LTFU-Care Algorithms (LTFU-Cares) a web-based database for LTFU data collection for families, patients, clinicians, and newborn screening programs; (2) LTFU-Checklist (LTFU-Check) a checklist for LTFU monitoring, assessment, and reporting for states; and (3) LTFU-Cares Dashboard a collection of data visualization for families, patients, clinicians, and states on LTFU. Families, individuals with SMA, and clinicians who specialize in SMA will guide tool development and provide health information, while state NBS programs will receive summaries of health outcomes. The LTFU Cares & Check Initiative focuses on NBS and SMA but is a useful model for other timepoints and genetic diseases.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2170. Ensuring participant understanding of informed consent through an automated consent chatbot for a national genome sequencing study.

Authors:

E. Andrew¹, S. Savage², S. I. Berger¹, J. LoTempio¹, V. A. Fusaro², G. Mas², E. C. Delot³, E. Vilain¹; ¹Children's Natl. Res. and Innovation Campus, Washington, DC, ²Invitae, San Francisco, CA, ³Children's Natl. Res. and Innovation Campus, Washington, DC

Abstract Body:

Background: Informed consent is the cornerstone of ethical participation in biomedical research. For genomic studies, particularly with vulnerable groups and those with limited access to genetic providers, ensuring understanding of complex concepts is key. These concepts include sharing of genomic data in Federal repositories, risk of unexpected findings, and Federal legislation such as the Genetic Information Nondiscrimination Act. Traditional consenting is repetitive and time-intensive, can take weeks to schedule, and requires staff to review dense consent forms and attest that enrollees understood. Few studies audit participants' understanding of what was discussed during consent. We propose an alternative: a conversational chatbot scripted on informed consent.

Methods: We created an automated consent process using the virtual Genetic Information Assistant (Gia (R)) chatbot for the Pediatric Mendelian Genomics Research Center, an NHGRI-funded GREGoR site. Gia's script is based on the IRB-approved consent forms reviewed by study staff, rewritten and streamlined to fit text message format. Gia is informative and conversational and permits users to select the level of detail in the conversation while covering all required elements of informed consent. Gia works on a mobile device or computer and is more accessible to participants. Gia requires the completion of a short knowledge quiz prior to permitting consent. All individual responses along with actual consent are recorded. Results: Many participants opted to consent with Gia rather than traditional consenting. Participants who obtained a 100% score on a 10-question knowledge quiz within two attempts were automatically enrolled. Those who did not obtain 100% were directed to a traditional consent with study team member. Gia takes about 35 minutes to complete, compared to 60 min for the traditional consent.

Discussion: Gia chatbot-based consent for genomic sequencing studies is feasible, allows mobile access to consent, and reduces barriers. Gia ensures each participant has a uniform experience and study staff can audit understanding. Traditional consenting is challenging for large studies because the process is highly manual; Gia allows large scale population studies without increasing support staff.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2171. Ethical considerations for genomic newborn hearing screening from the SEQaBOO (SEQuencing a Baby for an Optimal Outcome) study.

Authors:

C. Morton¹, C. O. Mitchell¹, G. Rivera-Cruz², M. H. K. Chau³, E. Dong³, R. K. W. Choy³, K. T. Booth⁴, J. Shen⁵, S. Amr⁶, A. B. S. Giersch¹; ¹Brigham and Women's Hosp, Boston, MA, ²Boston Children's Hosp., Boston, MA, ³Chinese Univ. of Hong Kong, Hong Kong, Hong Kong, ⁴Harvard Med. Sch., Boston, MA, ⁵GeneDx, Gaithersburg, MD, ⁶Brigham and Women's Hosp., Boston, MA

Abstract Body:

Recent advances in genomic sequencing technologies have expanded our ability to provide timely, efficient, and accurate genetic diagnoses. However, with increasing genomic information clinicians are presented with ethical decisions about how and when genomic information should be disclosed to families. Currently, there is limited knowledge surrounding the benefits and burdens of implementing genome sequencing into newborn screening.

SEQaBOO is a research project offering genome sequencing for newborns who refer to diagnostic audiometry following failure of newborn physiologic hearing screening. The child's deafness and hard of hearing (DHH) is used as a paradigm for implementing genomic newborn hearing screening. SEQaBOO provides comprehensive genome sequencing and variant interpretation of DHH-associated genes, as well as optional (for parents only) ACMG secondary findings (SF) v 3.0. The analytic platform can detect chromosomal aneuploidy, copy number variants (CNVs), absence of heterozygosity, and chromosomal structural rearrangements including balanced translocations.

A comprehensive study of this nature has potential to identify genomic information that may or may not be associated with DHH phenotype. To date, 99 families have consented to receive genomic sequencing results (n=272 individuals), and of those, 81 families have been analyzed utilizing the analytic platform. This comprehensive method has identified balanced chromosomal translocations (n=4), sex chromosome aneuploidy (n=1), and pathogenic CNVs (n=11) in genes associated with DHH. While 9 had direct implications for the DHH phenotype, 7 were considered incidental findings, with no clear implications for intervention. Findings of incidental results led to obtaining further IRB approval for disclosure.

Many ethical issues need to be considered when presenting genomic data in the newborn period, including the effect of disclosing incidental findings in newborns that may be unrelated to the initial indication for genetic testing. Thoughtful deliberations of the burden and benefit of disclosing incidental findings in newborns for late-onset conditions is essential. Overall, there are vast benefits for implementing genomic newborn hearing screening but considerations for public trust in genomic science are requisite to facilitate large-scale implementation.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2172. Evaluation of out-of-province and self-pay for genetic testing in two publicly funded provincial programs in British Columbia, Canada

Authors:

A. Elliott¹, P. Grant¹, S. Langlois¹, J. Nuk², S. Mung², Q. Zhang³, GenCOUNSEL Study, L. Lynd¹, J. Austin¹; ¹Univ British Columbia, Vancouver, BC, Canada, ²BC Cancer Agency, Vancouver, BC, Canada, ³BC Children's Hosp. Res. Inst., Vancouver, BC, Canada

Abstract Body:

Background: In Canada's public healthcare system, each province determines which tests to fund by evaluating evidence of clinical utility, medical necessity and cost-effectiveness. When tests are not funded publicly, healthcare providers may discuss private pay options (e.g. self-pay) with patients. These tests are typically sent out-of-province (OOP). The Hereditary Cancer Program (HCP) and Provincial Medical Genetics Program (PMGP) provide genetic services to residents of British Columbia and the Yukon Territory and collectively see >12,000 patients/year. In 2018, there was a change in the funding authority for OOP genetic testing for PMGP.

Methods: We evaluated trends in total OOP genetic testing and self-pay for HCP and PMGP from 2015-2019. Linear and regression models were applied to HCP data to explore outcomes including the relationships between self-pay with patient and test characteristics. To examine the effect of a change in the funding authority for OOP genetic testing in 2018 for PMGP, interrupted time series linear and logistic regression models were used on PMGP data.

Results: Out-of-province genetic tests completed through HCP increased 320% from 2015 to 2019 (1027 to 3289) while self-pay increased 730%. For self-pay patients, the most frequent indications were hereditary breast and ovarian cancer (74.8%), familial pancreatic cancer (8.0%) and Lynch syndrome (7.7%). For each year studied, the mean individual income of self-pay patients was \geq \$3500 higher than in the group with funded testing ($p < 0.0001$). The total number of OOP tests completed through PMGP increased 260% from 2015-2019. After the change in funding authority in 2018, the total number of tests increased by 2.35 per-month (95% CI 1.03, 3.66), which differed from the trend before change ($p = 0.0089$). The likelihood of self-pay increased over months before the funding authority change (OR per month: 1.07; 95% CI: 1.04, 1.10); while this likelihood had an immediate 87% drop when the change occurred (OR: 0.13; 95% CI: 0.06, 0.32).

Implications: Advances in sequencing technologies with associated reductions in costs have resulted in increased uptake of genetic testing. Patients with higher incomes are more likely to self-pay. Financial barriers create disparities in clinical outcomes for patients based on their economic resources. Funding decisions can have a significant impact on uptake and rate of self-pay. There are currently no Canadian guidelines for genetic healthcare providers regarding self-pay and testing resulting in dramatic heterogeneity of access. A pan-Canadian study evaluating issues including funding approval and clinical practice is urgently needed.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2173. Evidence-Based Assertions of Clinical Actionability of Secondary Findings: Update from the ClinGen Actionability Working Group

Authors:

C. Pak¹, J. E. Bulkley¹, P. Chakraborty², M. J. Gilmore¹, A. M. Foreman³, I. Futchi³, C. L. Jenkins¹, A. Katz⁴, K. Lee³, A. Milosavljevic⁵, J. O'Daniel³, J. E. Posey⁵, N. Shah⁵, R. D. Steiner⁶, A. B. Stergachis⁷, S. Subramanian⁸, T. Trotter⁹, M. S. Williams¹⁰, A. H. Buchanan¹⁰, K. A. B. Goddard¹¹, K. Manickam¹², B. C. Powell³, J. E. Hunter¹³, ClinGen Resource; ¹Kaiser Permanente Ctr. for Hlth.Res., Portland, OR, ²Children's Hosp. of Eastern Ontario and Univ. of Ottawa, Ottawa, ON, Canada, ³Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁴Horizon Therapeutics, Rockville, MD, ⁵Baylor Coll. of Med., Houston, TX, ⁶Univ. of Wisconsin Sch. of Med., Marshfield, WI, ⁷Univ. of Washington, Seattle, WA, ⁸TwinStrand BioSci.s, Seattle, WA, ⁹John Muir Hlth., San Ramon, CA, ¹⁰Geisinger, Danville, PA, ¹¹Natl. Cancer Inst., Bethesda, MD, ¹²Nationwide Children s Hosp., Columbus, OH, ¹³RTI Intl., Durham, NC

Abstract Body:

The ClinGen Actionability Working Group (AWG) uses an evidence-based framework to generate summary reports and consensus scores of the clinical actionability of secondary findings from genome-scale sequencing in adult and pediatric contexts. Recently, the AWG expanded its framework to generate assertions of actionability of gene-condition topics to inform decision-making for the return of secondary findings. The assertion process begins with a preliminary assertion of Limited, Moderate, or Strong Actionability based on consensus scores. The AWG then discusses the default actionability score-based preliminary assertion and assigns a final assertion which may differ based on the curated gene-condition evidence. The AWG may also choose an assertion of Definitive Actionability if the topic meets predetermined criteria. The AWG documents the rationale for the final assertion. To date, the AWG has finalized assertions for 108 pediatric and 135 adult gene-condition topics (www.clinicalgenome.org). The AWG changed the preliminary assertion to a higher category of actionability in 10% of pediatric and 11% of adult topics and to a lower category in 3% of pediatric and 7% of adult topics. While most assertions were unchanged, differences between preliminary and final assertions highlight nuanced factors not directly reflected in the scoring process that are nonetheless important considerations for actionability. For example, inclusion criteria used in studies could affect generalizability. In addition, when evidence is limited to descriptive or case studies, it is important to consider whether evidence from larger studies and/or controlled trials are likely forthcoming due to factors such as rarity of the condition. Importantly, our assertion process illuminates topics where gaps in available evidence makes assessment of actionability challenging. For example, *G6PD* deficiency has little published evidence for actionability, which may be related to it being common in populations traditionally underrepresented in research. This led to a preliminary assertion of Limited Actionability despite long-standing management practices, availability of clinical assessments of severity, and the likelihood of severe complications. The discussion of *G6PD* deficiency highlighted the risk of perpetuating disparities by relying on an evidence base with insufficient representation of population groups and the need to account for this representation in the assertion process. Overall, the expansion of the AWG framework to provide actionability assertions aims to facilitate decision-making around return of secondary findings from genome-scale sequencing.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2174. Expanding ethics education for trainees: creating and evaluating a course on genetics, ethics, and society.

Authors:

R. Patel, R. A. Ungar, M. Chakraborty, D. Cotter, P. Goddard, E. Greenwald, T. Mackenzie, D. O. Martschenko, A. Pyke; Stanford, Stanford, CA

Abstract Body:

Human genetics research has an ethically fraught past, with a legacy of discrimination featuring racism, classism, ableism, and exploitation. At the same time, the last decade has seen the rapid and unprecedented development of technologies capable of mapping, modeling, and manipulating genomes. Despite this troubled history and uncertain future, most training programs do not provide formal education on how to situate human genetics research within its historical and societal context. Instead, current ethics requirements largely focus on researchers' responsibilities to individuals. This narrow focus leaves most trainees ill-equipped to evaluate the broader impacts of research on society and neglects necessary dialogue about scientists' responsibilities to mitigate potential social harms.

Here, we describe the curriculum for a 12 hour, discussion-based course entitled "Genetics, Ethics, and Society." Developed by a team of genetic counselors, doctoral students, and postdoctoral scholars from bioethics and genetics, the course sought to achieve the following learning goals: (1) connect the historical context of human genetics research to modern-day practices; (2) evaluate the societal and ethical implications of human genetics research; and (3) analyze how societal norms and structures, along with personal identities, biases, and responsibilities, impact the conduct of scientific research. We offered this course in May 2022, and it was attended by 16 graduate students, three undergraduate students, and one postdoctoral scholar, all of whom work in biomedical research.

Using pre- and post-course surveys of enrolled trainees, we evaluate the effectiveness of this course in meeting its learning goals. We quantitatively describe the impact of the course on students' understanding of key topics, as well as their perceived importance of these topics. In addition, we report the primary motivations of trainees enrolling in this course and their main takeaways upon completion. We anticipate that this course will provide a model for how to educate trainees on the ethical and societal implications of scientific research. Our results underscore the importance of expanding ethics education to consider social risks and responsibilities.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2175. Expanding genomic testing to underserved pediatric populations

Authors:

A. S. A. Cohen^{1,2,3}, S. Davis⁴, N. J. Kane⁴, T. Zion¹, R. Moore¹, B. D. Zuccarelli⁵, C. Berrios¹, T. Pastinen^{1,3,4}, M. A. Hoffman^{3,4,6}; ¹Genomic Med. Ctr., Children's Mercy Hosp., Kansas City, MO, ²Dept. of Pathology and Lab. Med., Children's Mercy Hosp., Kansas City, MO, ³Univ. of Missouri–Kansas City Sch. of Med., Kansas City, MO, ⁴Children's Mercy Res. Inst., Kansas City, MO, ⁵Salina Pediatric Care Ctr., Salina, KS, ⁶Res. Informatics, Children's Mercy Hosp., Kansas City, MO

Abstract Body:

The “Genomic Answers for Kids” (GA4K) program at the Children’s Mercy Research Institute (CMRI) is committed to increasing access to genetic testing and finding answers for more pediatric rare genetic conditions. Patients enrolled in GA4K undergo short-read exome/genome sequencing (srES/srGS) with integrated machine-assisted strategies for rapid screening of positive cases to accelerate return-of-results. Following negative srES/srGS, an enhanced long-read GS workflow using PacBio Sequel IIe (HiFi-GS) is used to investigate variants in difficult-to-map regions and other novel variation. Despite employing diverse strategies for recruitment, we have identified **significant inequities in access to our study** that limit the reach of our pioneering methods. Patient identifiers from our cohort were used to extract data on race, ethnicity, home address, and payor type from the electronic medical record. Analysis of these variables revealed that GA4K enrollment under-represented racial and ethnic minorities and socioeconomically disadvantaged pediatric patients both within and beyond the Children’s Mercy Hospital (CMH) service area. For example, the GA4K cohort included only 5% Black and 7% Latinx children, compared to 15% and 11% respectively in the CMH patient population. These numbers didn’t include known prior inequities in access to care at CMH. **Under-representation of these groups goes beyond our study and leads to inflation of “rare” variants and reduced ability for accurate interpretation.** Geolocalization also demonstrated unequal enrollment across regions, with <25% of study participants having a home address in a rural zip code. These disparities may disproportionately impact patients from economically disadvantaged and rural communities because they rely on their local primary care providers (PCPs) for all healthcare needs. GA4K is **piloting a “direct-to-primary care provider” (DTP) strategy** to reach more rare disease patients by bringing genetic testing to their medical homes. Our strategy focuses on empowering PCPs to recognize their patients’ necessity for genetic testing and providing support for consenting, sample collection and return-of-results through the GA4K pipeline. A detailed description of our pilot strategy and preliminary results will be presented.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2176. Exploration of a Novel Cancer Genetics Learning Format: A Pilot Self-Directed Course in Genetic Cancer Risk Assessment.

Authors:

D. Tanna, P. Leung, J. M. Luna, G. Nuñez, R. Manookian, S. Dreike, S. Shehayeb, E. Sobotka, K. R. Blazer; City of Hope Comprehensive Cancer Ctr., Duarte, CA

Abstract Body:

Background: Despite efforts to integrate genetic cancer risk assessment (GCRA) services into mainstream medicine, the dearth of healthcare providers adequately trained in clinical cancer genetics remains a significant barrier to care. City of Hope delivers an annual interprofessional GCRA training program for community-based clinicians. As designed, the program cannot meet ongoing demands for GCRA training. Results of a pilot study to inform the design of a self-directed version of the training program are presented here.

Methods: A self-directed subset of the program curriculum covering Hereditary Breast and Ovarian (HBOC), Gastrointestinal (GI), and Genitourinary (GU) cancers was offered to a selected group of GCRA-practicing clinicians (n=67) who are program alumni and indicated interest in updated GCRA education. Participants could register for up to Three Course Bundles (HBOC, GI, GU), each containing video modules, quizzes, assignments, and evaluations to be completed within four months. Evaluations included Likert scale surveys to assess learning objectives and expectations, content quality, and the digital learning platform. Quantitative analysis of survey responses and qualitative assessment of open-ended feedback yielded recommendations for content and design strengths and improvements.

Results: Forty-four (17 MDs, 20 APNs, 1 PA, 6 GCs) of 67 invited participants enrolled in one or more course bundles; 56% completed all HBOC modules, 46% completed all GI modules, and 52% completed all GU modules. Participants reported that each module met stated learning objectives with HBOC scoring 4.56, GI scoring 4.51, and GU scoring 4.43 out of 5 quality points on average. Participants rated the quality and ease of use of the digital learning platform and quizzes/assignments an average of 4.30 and 4.32 out of 5 quality points, respectively.

Discussion: Achievement of learning objectives scored high, showing the self-directed curriculum to be a valuable resource for cancer genetics education. Open-ended responses suggested the need for more simplified content, additional clinical case examples, and continuous evidence-based and guideline updates. The digital learning platform was rated as user-friendly and the automated feedback on quizzes and assignments was greatly valued. Adding subtitles, improving audio/visual quality, and a mechanism to interact with faculty were recommended. Completion rates indicated the need for extended access for busy clinicians. Pilot outcomes will guide the development of a self-directed GCRA training program to scale up access to flexible GCRA training for clinicians across the spectrum of cancer care.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2177. Exploring stigma and discrimination among adults after receipt of neurodevelopmental/psychiatric genetic results in a population-based genomic screening program.

Authors:

O. Matshabane¹, K. Wain², A. Heidlebaugh³, M. Good³, K. Holdren³, B. Kikani¹, L. Walsh³, L. M. Koehly¹; ¹NIH, Bethesda, MD, ²GeneDx, Otsego, MN, ³Geisinger, Lewisburg, PA

Abstract Body:

Advances in neurodevelopmental/psychiatric disorder (NPD) genetics hold great promise for improved prevention, diagnosis and treatment. Models that increase access to NPD-related genetic information are needed. The MyCode population-based genomic screening program has returned NPD-related genetic results to 279 adults since 2017. Individuals with NPD can experience discrimination and the effects of stigma in daily life which may impact communication, coping and access to treatment. How knowledge of a genetic etiology for NPD impacts perceived stigma and discrimination experiences remains underexplored. This study utilized the Everyday Discrimination Scale and a novel adaptation of the Stigma by Association Scale to explore these concepts among adults who received a clinically relevant NPD-related genetic result from the MyCode program. In fall 2021, this study began exploring probands' communication decisions and experiences of stigma and discrimination related to a NPD genetic result amongst social networks at least 6 months after receiving results. 31 adult probands were invited to participate and 9 have participated to date (29% participation rate). Social network interviews were completed telephonically, resulting in 73 individuals, primarily family members, embedded in 9 ego-centered networks. The Everyday Discrimination Scale (5 items) assesses perceived experiences of receiving disrespectful treatment, poor service, being treated as "not smart", people acting afraid of them, and being threatened and harassed. Overall, 48% of probands report having these experiences one to four times a month. The most endorsed item (by 20%) was being "treated with less courtesy or respect than other people". The adapted Stigma by Association Scale (13 items) measures feelings, beliefs, and reactions to perceived personal, familial and associative stigma specifically related to a NPD genetic result. A total of 16.48% agreed to these stigma items. The most endorsed item was "I do not talk to others about it", by 6.35%. Network data revealed that although probands enumerated 8.11 living network members (Range: 3 - 16) through network assessments, genetic results were shared with an average of 1.78 network members (Range: 0 - 3). Results to date suggest that these participants may currently be refraining from communicating their genetic result with their social networks. Further research with larger samples exploring whether and how experiences of discrimination may be affecting the decision of communicating a NPD-related genetic result are needed to inform clinical and population-based models geared at returning these results.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2178. Exploring the ethical implications of genomic data sharing in cloud-based ecosystems: Views of developers and users.

Authors:

S. Nelson, J. Dahlquist, S. M. Fullerton; Univ. of Washington, Seattle, WA

Abstract Body:

Background: New cloud-based data storage and analysis ecosystems are being implemented to facilitate collaboration and maximize the scientific utility of costly-to-generate genomic and linked clinical data. These ecosystems differ from prior genomic data sharing efforts in bringing researchers to the data (instead of data to the researchers) to enable frictionless data search, aggregation, analysis, and dissemination. They also entail a range of user authentication and authorization processes, data access review procedures, and related oversight mechanisms that may blur responsibilities and complicate data stewardship. To gain a better understanding of these platforms and examine the ethical implications of this emerging genomic data sharing landscape, we conducted qualitative interviews with select platform developers and users, with the aim of informing researcher education and policy development. **Methods:** We conducted semi-structured confidential interviews with 20 developers and policymakers affiliated with five NIH-supported cloud platforms: the NCI Genomic Data Commons, the NHLBI BioData Catalyst, the NHGRI AnVIL (Analysis, Visualization, and Informatics Lab-space), the Kids First Data Resource Portal, and the All of Us Research Hub, as well as the NCBI database of Genotypes and Phenotypes (dbGaP) whose data access request and review systems are used by many cloud platforms. We also conducted interviews with researchers who have either shared, or made use of, data stored in these ecosystems. All interviews were recorded, transcribed, and subjected to a directed content analysis designed to identify key implications of cloud-based genomic data sharing. **Results:** Developers and policymakers described a complex array of technical and policy-based governance procedures designed to ensure data security and facilitate data sharing consistent with research participants' informed consent. Researchers making use of these new data sharing ecosystems described many advantages, but also challenges, associated with moving from local data storage and analysis to cloud-based approaches. **Conclusions:** Close analysis of the views of developers and users illustrates the complexity of these new data sharing mechanisms and suggests the need for more dialogue between users, developers, researchers, and policymakers to ensure that the promises of cloud-based ecosystems are realized.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2179. Factors contributing to preimplantation genetic testing for polygenic disorders (PGT-P) screening and recommendations by reproductive medical professionals.

Authors:

S. Shah¹, M. Quinn², J. Shamshoni¹, R. LeShay¹, R. Cantor³, C. Palmer⁴; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²Univ. of Southern California, Los Angeles, CA, ³UCLA Sch Med, Los Angeles, CA, ⁴UCLA Semel Inst, Los Angeles, CA

Abstract Body:

Background: Preimplantation genetic testing for polygenic disorders (PGT-P) is a new area of research and clinical application of which is available in the United Kingdom, Scandinavia, and the United States (U.S.). No research has been published reporting the awareness, uptake, and acceptability of PGT-P among reproductive medical professionals. **Methods:** The prevalence of awareness of PGT-P among U.S. reproductive genetic counselors (GCs) was assessed by administering an online survey, designed and validated at UCLA. GCs who are currently not practicing in the U.S. or who have not counseled patients regarding preimplantation genetic testing in the last 5 years were excluded from the study. Statistical analyses were conducted to assess the impact of subject age and years of practice on provider beliefs regarding patients' access to PGT-P. Adjusted logistic regressions were performed to assess which factors would motivate GCs to consider PGT-P as an option for their patients. **Results:** Approximately 72% respondents among our sample of 83 were aware of PGT-P prior to this survey. Only 13% of the participants reported they may recommend PGT-P to their patients, and most participants felt uncertain about patients having the option of PGT-P for polygenic conditions. Currently, nearly half of all participants (48%) did not want patients to have the option of PGT-P. GCs' opinions about patients having the option of PGT-P were not significantly associated with age strata or years of practice. Factors such as use of donor egg/sperm/embryo (aOR = 3, 95% CI = 1.2, 7.5), family history of a polygenic condition (aOR = 4.4, 95% CI = 1.7, 10.9), and prior pregnancy history (aOR = 4.31, 95% CI = 1.47, 12.63) were independently and significantly associated with GCs' uncertainty regarding patients having the option of PGT-P compared to GCs' perspectives on patients not having the option. **Conclusions:** Currently, most GCs (45%) in the U.S. are unsure if patients should have the option of PGT-P compared to other types of preimplantation genetic tests. Most (72%) of them are aware of PGT-P, but 87% would not consider recommending PGT-P to their patients. In contrast, most GCs support patient access to PGT-aneuploidy (PGT-A; 100%), PGT-monogenic (PGT-M) disorders of childhood onset (99%), PGT-M for adult-onset conditions (87%), and PGT-structural rearrangements (PGT-SR; 99%). Further research is warranted to understand whether those who are unsure about their stance would consider access to PGT-P for patients in the future.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2180. Family communication choices about neurodevelopmental/psychiatric genetic results: A social-network assessment of adult participants in a population-based genomic screening program.

Authors:

K. Wain¹, O. P. Matshabane², A. Heidlebaugh¹, M. Good¹, K. Holdren¹, B. Kikani², L. Walsh¹, L. Koehly²; ¹Geisinger, Danville, PA, ²NHGRI, Bethesda, MD

Abstract Body:

Clinical genetic testing is standard of care for neurodevelopmental/psychiatric disorders (NPD) in children; yet adults are rarely offered this service. Since 2017, the MyCode population-based genomic screening program has returned NPD-related genetic results to nearly 300 adults to explore how this model can improve access. Initial responses to receiving NPD-related results have been positive and indicate an intent to share results with others. However, the extent to which participants share is not well understood. This study aimed to assess communication patterns amongst NPD-result recipients' social networks and identify factors that influence sharing. Starting in fall 2021, nine participants were recruited after 6-months from receipt of NPD-results through the MyCode program (29% participation rate). Through phone surveys, we elicited ego-centric social networks, exploring kinship/social ties, contact frequency, physical distance between individuals, and if results were shared. Descriptive statistics, using $p=0.1$ to identify potential trends are provided. On average, probands enumerated 8.11 living network members (Range: 3, 16); results were shared with an average of 1.78 network members (Range: 0, 3). A total of 16 individuals across 8 families (22% of 73 living network members) were told about results. Spouses/partners were most likely to be told (5/7, 71%), followed by siblings (4/12, 33.3%), biological parents (3/11, 27.3%) and biological children (3/14, 21.4%). Of the 57 individuals who were not told, 11 were children (Age Range: 5-36 years) and 8 were siblings (Age Range: 19-61 years). Results were more likely to be shared with immediate family than other network members (OR=8.72, $p=0.09$), but there were no differences based on physical distance (OR=1.44, $p=0.69$), frequency of contact (OR=2.47, $p=0.37$) or if the family member also had an NPD diagnosis (OR=1.84, $p=0.59$). Network members who provide emotional support (OR=4.31, $p=0.12$) or other help (OR=2.77, $p=0.21$) were not more likely to be told. However, results were shared more often with those involved in health decisions (OR=56.03, $p<0.01$) and health-related discussions (OR=7.19, $p<0.01$). Thus, despite initially reported intentions to share NPD-related results with family, participants only shared with key immediate family members who help with healthcare. Individuals with NPD may have lower health literacy or efficacy with regards to communicating genetic information, which could be addressed by engaging these key family members. Future research is needed, with larger cohorts, to more fully describe participant communication decisions and intervention opportunities.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2181. *FLCN* variant of unknown significance complicating a kidney donation process in a family with Birt-Hogg-Dubé.

Authors:

W. Geurts-Giele¹, P. J. Gundlach¹, A. C. Houweling², N. Engels³, M. A. van den Dorpel³, R. van Minkelen¹, M. F. van Dooren¹; ¹Erasmus Med. Ctr., Rotterdam, Netherlands, ²Amsterdam UMC, VU Univ. Med. Ctr., Amsterdam, Netherlands, ³Maasstad Hosp., Rotterdam, Netherlands

Abstract Body:

Background Birt-Hogg-Dubé syndrome (BHD) is a rare autosomal dominant condition caused by germline mutations in *FLCN*. It can present with a variety of clinical manifestations, including kidney tumors, pulmonary cysts, spontaneous pneumothorax and fibrofolliculomas. **Case presentation** A 49-year-old woman without pre-existing pulmonary disease was admitted to the hospital due to a spontaneous unilateral pneumothorax. Chest CT revealed bilateral basilar pulmonary cysts. BHD was suspected and genetic testing revealed heterozygosity for the variant NM_144997.5(*FLCN*):c.324C>A, p.(Ser108Arg), classified as a variant of unknown significance (VUS) based on ACMG criteria PM2, PP3 and PP4. Concurrently, several first-degree relatives of the patient applied for living kidney donor evaluation because they wished to donate a kidney to a sister (SIB1) of the index patient with end-stage kidney disease due to granulomatosis with polyangiitis. During this screening process a CT scan performed in another sibling (SIB2) of the index patient revealed a kidney tumor. SIB2 turned out to be a carrier of the same *FLCN* variant, with loss of the wildtype allele in the tumor. Fibrofolliculomas were histologically confirmed and pulmonary cysts were found. Although the *FLCN* variant was a VUS according to the ACMG guidelines, the variant allele was considered likely pathogenic for this family because the phenotype of the sisters was consistent with BHD. In agreement with the BHD expert center preferred relatives for donor nephrectomy should not carry the VUS, and no phenotypic features of BHD should be present (e.g. kidney tumors, fibrofolliculomas and pulmonary cysts). Segregation analysis for this variant was complicated due to ambiguous classification of the variant, logistic problems and different testing strategies. Finally, four siblings (including SIB1) did not carry the variant. Their 85-year-old father was a carrier of the variant, with histologically confirmed fibrofolliculomas. Based on the segregation analysis ACMG PP1_strong was applied and the variant was re-classified as a likely pathogenic variant. Non-carriers were released from follow-up screening. **Discussion** The ambiguous classification of the *FLCN* VUS complicated the living kidney donation process in this family with BHD due to the absence of clear guidelines for this situation. Time was lost due to extra screening, the necessity of this screening is open to debate. Additionally, this scenario was confusing for both the family and treatment teams. A multidisciplinary family clinic might resolve some of these issues. Finally, a brother of SIB1 was found eligible for living kidney donation.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2182. Fragile X syndrome: Experiences with receiving a genetic diagnosis and communicating genetic risk in a Cameroonian family.

Authors:

K. Kengne Kamga¹, N. Munung¹, N. Séraphin², A. Wonkam¹, J. de Vries¹; ¹Univ. of Cape Town, Cape Town, South Africa, ²Univ. of Yaounde 1, Yaounde, Cameroon

Abstract Body:

Fragile X Syndrome (FXS) is the leading genetic cause of intellectual disability (ID) and Autism spectrum disorder. Receiving a genetic diagnosis of FXS can have an emotional impact on the patient, parents, and the rest of the family. This paper reports on the experiences of receiving a genetic diagnosis of FXS in an extended Cameroonian family. Using an ethnographic approach, we conducted 13 in-depth interviews and one focus group discussion with members of a large family segregating for FXS. We interrogated their experiences of receiving a genetic diagnosis. Two main themes emerged from the interviews relating to psychological adaptation and communication of FXS genetic risk. First, we describe the happiness and relief that are associated with diagnostic closure. Next, we describe genetic guilt, survivor guilt, and frustrations associated with a family history of FXS and taking care of developmentally delayed children. Finally, we highlight the communication styles used by this family to convey genetic risk to extended relatives and promote resilience. Understanding the complex interaction which occurs during the return of genetic findings will help improve the quality of pre-and post-genetic counselling practices and hence minimize the risk of conflicts. Future studies should focus on designing innovative tools to enhance the process of connecting family members together once they receive a genetic diagnosis.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2183. Fragmented systems of care: An overview of Canadian health system care models for hereditary cancer syndromes

Authors:

J. Sam¹, M. Clausen¹, C. Mighton², S. Rajeziesfahani³, R. Gopalakrishnan¹, M. Aronson⁴, D. Bishop⁵, L. Dawson⁶, A. Eisen⁷, T. Graham⁸, J. Green⁹, M. Krahn¹⁰, J. Pauling¹¹, C. Pavao¹², C. Remocker¹², S. Savas³, N. Stjepanovic¹, S. Sun¹³, T. Tiano¹¹, A. Tilley⁵, K. Thorpe², K. Schrader¹³, H. Etchegary⁶, Y. Bombard²; ¹St. Michael's Hosp., Toronto, ON, Canada, ²St. Michael's Hosp. & Univ. of Toronto, Toronto, ON, Canada, ³Mem. Univ., St. John's, NL, Canada, ⁴Zane Cohen Ctr., Sinai Hlth.System, Toronto, ON, Canada, ⁵Patient Partner, St. John's, NL, Canada, ⁶Mem. Univ., St. John's, NL, Canada, ⁷Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, ⁸Univ. of Toronto & Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, ⁹Mem. Univ., Middle Cove, NL, Canada, ¹⁰Univ. of Toronto & Univ. Hlth.Network, Toronto, ON, Canada, ¹¹Patient Partner, Toronto, ON, Canada, ¹²Patient Partner, Vancouver, BC, Canada, ¹³BC Cancer, Vancouver, BC, Canada

Abstract Body:

Background: Hereditary cancer syndromes (HCS) are one of the most prevalent inherited diseases, accounting for 5-10% of all cancers. Patients and their family members with a confirmed genetic diagnosis of a HCS require lifelong screening and follow up since they have an increased risk for multiple malignancies in several organ systems. However, there is limited data on the accessibility and coordination of HCS care across different health jurisdictions in Canada. **Aim:** The purpose of this study is to compare the systems of HCS care in 3 provinces in Canada to support a team grant that aims to explore the indirect socioeconomic and psychosocial impacts of HCS. **Methods:** Expert leads of provincial HCS care programs in British Columbia (BC), Ontario (ON), and Newfoundland & Labrador (NL) were consulted to provide detailed information about the care systems in their respective provinces. The care system dimensions examined included an overview of populations and regions served, structure of genetic service delivery, genetic testing eligibility and panel sizes, coordination of follow up care, and gaps in care. **Results:** Access to and coordination of HCS care across all 3 provinces is fragmented. First, there are inconsistencies in genetic testing referral criteria and management recommendations for carriers. This means that some family members across the country have access to HCS medical services (e.g., genetic testing, screening, prophylactic surgery) while others do not. Accessing genetic testing is further impaired by the varied and strict eligibility criteria for testing across the provinces. Secondly, all provinces face a genetics workforce shortage, and long wait times for families to access publicly funded testing. Thirdly, provincially-organized screening and surveillance programs only exist for some organs in BC, ON, and NL. The lack of organized screening programs leaves at risk individuals and their family physicians to navigate the system of care alone. This results in inequities in medical outcomes, further exacerbated for underserved populations and rural communities. Lastly, the genetic testing offerings across the 3 provinces vary. Differences in gene panels means that at risk individuals living in different jurisdictions will have inconsistent genetic diagnoses. **Conclusions:** A fragmented and uncoordinated HCS care system may result in socioeconomic, psychosocial and health implications for HCS families. Further investigation is needed to better examine these impacts to inform evidence-based practice for a more coordinated and patient-centered health care system for families affected by HCS.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2184. Gender medicine, Gender-Omics and Sex-Omics: What we are dealing with?

Authors:

A. Karra^{1,2}, W. Smaoui¹, S. Kammoun¹, N. Halouani¹, N. Abdelmoula¹; ¹Med. Univ. of Sfax, Sfax, Tunisia, ²Genomics of Signalopathies at the service of Med., Sfax, Tunisia

Abstract Body:

Background: Gender medicine is viewing patients through a sex and gender lens and studying of how diseases differ between men and women. It is a novel and promising approach in terms of prevention, clinical signs, therapeutic approach, prognosis, psychological and social impact. It represents the crucial assumption for achieving the precision medicine required in this third millennium. Currently, while it is approved that biological, genetic, epigenetic, psycho-social and environmental determinants interact conjointly in defining sex/gender differences and in setting up sex/gender disparities, two new terms evolved in literature in the field of Omics Sciences: Gender-Omics and Sex-Omics. Generally, the objective of omics sciences is to identify, characterize and quantify all biological molecules that are involved in the structure, function and dynamics of a cell, tissue or organism. Here, we aim to review the objective of Sex-Omics and Gender-Omics and to sign on the role of genet/epigenetics in these entities. **Methods:** We comprehensively review the scientific literature using Pubmed and others databases to state the meaning of Gender-Omics and Sex-Omics and to delineate the role of genetics and epigenetics in these new terms. **Results:** Our bibliographic research revealed a poor literature with only two papers responding to our inquiry in PubMed and 59 papers in Google scholar. Interrogation of the 61 papers demonstrated that the term Sex-Omics is more frequently cited. The authors deal with the role of both sex and gender in physiological and pathophysiological processes of diseases using a holistic omics approach. They reported studies that consider Sex and Gender Omics biomarkers involved in various diseases. They reported also studies that consider sex-specific variations in signaling pathways and gene expression patterns in response to various health processes, pathological situations, therapeutic approaches, etc. There are also the notion of the sexome, defined as the summation of sex-biased effects on gene networks or the reporting sex differences in gene systems. For gender influence in an omics approach, there are less reports. However, Genderome may be defined as the summation of gender-biased effects on epigenetic networks or the reporting gender-associated differences in the epigenetic systems (the epigenome with methylome and the imprintome). **Conclusion:** Sex-Omics and Gender-Omics are two new omics that have been proposed in investigating sex/gender specific aspects in biomedical sciences. These omics have complex and multifaceted nature that makes it important to be approached in next years to improve precision Gendered medicine.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2185. Genetic counselors and legal recognition: a made-for-Canada approach

Authors:

M. Zawati¹, D. Patrinos¹, D. M. Lambert², B. Knoppers¹; ¹McGill Univ., Montreal, QC, Canada, ²Univ. Coll. Dublin, Dublin, Ireland

Abstract Body:

Background: Genetic counseling is a fast-growing profession in Canada. However, it lacks legal recognition in the majority of Canadian provinces and territories. Legal recognition ensures safety in the provision of healthcare services by regulating professions that pose a risk of harm to the public. Without legal recognition, any individual uses the title without the proper qualifications, and professional quality standards remain unregulated. This may potentially expose the public to risk of harm, including inaccurate interpretation of genetic testing results, incomplete risk assessment and psychological harm. We assessed the feasibility of legal models to achieve the legal recognition of genetic counselors within Canadian law. Methods: We conducted a comparative analysis of professional legislation across the three provinces with the largest workforce of genetic counsellors (British Columbia, Ontario and Quebec) to identify models of legal recognition and the level of public protection offered by each. We surveyed representatives of a Canadian genetic counseling regulation special interest group to estimate the provincial numbers of genetic counselors. Results: 3 models of legal recognition were identified: 1) constitution of a professional order, 2) inclusion in an existing professional order, and 3) delegation of specific acts from another profession. There are an estimated 484 genetic counselors in Canada, with 89% found in 4 of the 13 provinces and territories. Discussion: The first two models provide a higher level of public protection, yet are more difficult to achieve. Delegation relies on another professional body to delegate a limited number of professional acts, so is more limited in scope, but is easier to achieve, yet does not provide the same degree of public protection. The significant human and financial resources necessary to obtain legal recognition may mean that creation of a new profession of genetic counselling is beyond the reach of provinces and territories with fewer genetic counsellors. Given that genetic counseling is a relatively small profession, a balance needs to be achieved between protecting the public and leveraging the resources required to seek legal recognition. Conclusion: Though legal recognition occurs at the provincial and territorial level, we advocate for a pan-Canadian approach to develop strategies to further pursuits of legal recognition. A series of expert discussion groups of Canadian genetic counselors and other stakeholders are being formed to discuss the most effective and feasible manner of seeking legal recognition.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday
PB2186. Genetic Counselors' Attitudes Towards Religiously-Based Policies and Their Impact on Professional and Ethical Obligations in the Prenatal Setting

Authors:

L. Freedkin¹, M. Bocian¹, R. LeShay², K. Osann¹; ¹Univ. of California Irvine Med. Ctr., Orange, CA, ²Univ. of California Los Angeles, Los Angeles, CA

Abstract Body:

The National Society of Genetic Counselors (NSGC) Code of Ethics outlines the professional and ethical obligations genetic counselors have to themselves, their patients, their colleagues, and society. These guidelines stress the importance of patients' autonomy and informed consent. One interpretation of the NSGC Code of Ethics is that prenatal genetic counselors should also discuss the option of abortion with patients in order to cover all necessary facts and options. However, institutions with religiously-based policies (RBPs) that prevent genetic counselors from freely discussing abortion with patients seem to contradict the NSGC Code of Ethics. There is very little information in the current literature about genetic counselors' opinions on this potential conflict and its impacts. This study aimed to explore genetic counselors' attitudes towards RBPs, what characteristics of genetic counselors are associated with attitudes towards RBPs, and what characteristics are associated with genetic counselors' willingness to work in settings with RBPs. A survey was developed to measure genetic counselors' opinions on RBPs and interest levels in working in fictitious prenatal settings with various policies regarding abortion discussions. Overall, genetic counselors did not support the implementation of RBPs that limit abortion discussions (n=152, 86%) and believed that such policies do not align with the NSGC Code of Ethics (n=157, 90%). Most of the participants viewed abortion as moral (n=144, 77%), were comfortable counseling about abortion (n=124, 71%), and described themselves as liberal (n=134, 78%) or not very religious (n=131, 76%). These participants were more likely to express low support for RBPs ($p \leq 0.002$) and were less interested in working at institutions with policies limiting abortion discussions ($p < 0.001$). The findings of this study raise concerns for the quality and equity of genetic counseling services in settings with certain RBPs. As the political landscape shifts and abortion rights continue to be challenged, the findings of this study may encourage genetic counselors to advocate for additional NSGC guidance and ethical policies in the prenatal setting at the institutional and governmental level.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2187. Genetics education in neurodevelopmental disorders for non-genetic health professionals: A scoping review.

Authors:

C. Somerville¹, I. Peltekova^{2,3}; ¹The Hosp. for Sick Children, Toronto, ON, Canada, ²Holland Bloorview Kids Rehabilitation Hosp., Toronto, ON, Canada, ³Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Background: Clinical genetic testing is considered the standard of care for individuals with neurodevelopmental disorders (NDDs), like autism spectrum disorder, intellectual disability, and global developmental delay. Genetic tests are regularly ordered by non-genetic healthcare providers who lack formal training in medical genetics. Research has shown that these providers experience challenges in their knowledge, understanding, and management of genetic concepts and information. There is a need for structured genetics education geared towards non-genetic health care providers caring for individuals with NDDs. The aim of this scoping review was to synthesize the literature on formal genetics education interventions on NDDs for non-genetics health professionals. **Methods:** MBASE, MEDLINE, CINHAL, ERIC and PsychInfo were searched for peer-reviewed publications through November 2021. Two reviewers independently screened and extracted data from each study. Discrepancies in data analysis were resolved following joint re-evaluation. **Results:** In total, 9885 studies were screened and 72 met inclusion criteria. Study participants were largely primary care physicians, residents, medical students, and nurses. There were very few genetics education interventions specifically targeting NDDs. The topic of NDDs was mostly approached as part of broader genetics education initiatives, which largely consisted of didactic/in-person seminars or web-based learning modules (e.g., clinical case scenarios). Educational content was comprised of basic and applied genetics concepts, family history evaluation, genetic testing, and genetic services. Most studies implemented a pretest-posttest design. Outcomes included participant-reported satisfaction with the intervention and measures of genetics knowledge and confidence. However, few studies evaluated the impact on clinical care or utilization of genetic services. **Conclusion:** Although several genetics education interventions exist for non-genetic health professionals that incorporate NDDs as a topic, the development of a structured program is needed. Future research should focus on developing genetics education tools that are tailored to clinical and health system contexts, and evaluating their impact on clinical care, to help address current barriers to genetic service delivery for patients with NDDs.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2188*. Genome sequencing and preferences for secondary findings in ostensibly healthy COVID-19 positive individuals: GENCOV study Canada.

Authors:

S. Casalino^{1,2}, C. Mighton^{1,3,4,2}, M. Clausen³, E. Frangione^{1,2}, M. Chung^{1,2}, S. Chowdhary^{1,2}, G. MacDonald^{1,2}, H. Faghfoury^{1,5}, Y. Bombard^{3,4}, J. Taher^{1,4}, J. Lerner-Ellis^{1,2,4}, GENCOV Study Workgroup; ¹Sinai Hlth., Toronto, ON, Canada, ²Lunenfeld-Tanenbaum Res. Inst., Toronto, ON, Canada, ³Unity Hlth. Toronto, Toronto, ON, Canada, ⁴Univ. of Toronto, Toronto, ON, Canada, ⁵Univ. Hlth. Network, Toronto, ON, Canada

Abstract Body:

Background: Genome sequencing (GS) can identify genetic factors that influence variability in COVID-19 symptoms and outcomes. GS can also be used to screen for secondary findings (SF) about inherited predispositions to other diseases. Attitudes toward the use of GS as a screening tool in the general Canadian population is unknown, warranting exploration of individuals' preferences, knowledge, and attitudes towards GS/SF. **Aims:** Determine 1) preferences for SF from GS and 2) the impact of genetic counseling (GC) and educational tools on SF preferences and knowledge. **Methods:** Online surveys were administered to ostensibly healthy individuals previously diagnosed with COVID-19 before pre-test GC (T0), and immediately after completion of a digital genomics platform (www.geneticsadviser.com) along with pre-test GC (T1). Surveys at T0 assessed: 1) sociodemographic characteristics; 2) SF preferences; 3) knowledge of GS; 4) attitudes toward genetics/healthcare; and, 5) health literacy. Surveys administered at T1 re-assessed knowledge of GS and SF preferences. Responses were analyzed with descriptive statistics. **Results:** 605/696 responses were received for at least part of both the T0 and T1 surveys. The majority were female (56%), ≥40 years of age (54%), and had a Bachelor's degree or higher (66%). 59% indicated they were white/European. Most participants at T0 (462/599; 77%) and T1 (504/604; 83%) wished to learn all SF from GS. SF preferences were similar at T0 and T1, with the greatest interest in learning clinically actionable SF (97% at both time points), followed by polygenic risk for common conditions, drug reactions, carrier status, and other rare Mendelian conditions. Less than 1% did not wish to learn any SF. Mean scores for knowledge of sequencing limitations and benefits (n=520) improved from low at T0 (3.6 and 4.3 out of 10, respectively; Standard deviation (SD) 2.5-2.7) to moderate at T1 (5.3/10; SD 2.6-2.9). 80% (485/605) had positive/ambivalent attitudes toward genetics and 20% (120/605) had negative/ambivalent/mixed attitudes. 87% (524/605) think it is important/essential to receive the most advanced medical care. Health literacy was adequate based on a mean scale score of 18.4/20 (SD 2.4, n=596). Characteristics and predictors related to initial and post-GC SF preferences will be presented at the conference. **Conclusion:** Preliminary findings suggest that most individuals are interested in receiving SF from opportunistic GS. GC in conjunction with the digital genomics platform improves knowledge and may be an appropriate means of educating the general population about GS and SF.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2189*. Genomic knowledge, orientation, and empowerment in a demographically diverse population: Preliminary findings from the Texome project.

Authors:

C. Murali¹, R. German², T. Nguyen Dolphyn³, C. Bacino², S. Lalani⁴, H. Bellen², S. Yamamoto⁵, z. liu², M. Wangler⁶, M. Wangler⁴, P. Lupo²; ¹Texas Children's Hosp., Houston, TX, ²Baylor Coll. of Med., Houston, TX, ³Stanford Clinical Genomics Program, Stanford, CA, ⁴Baylor Coll. Med., Houston, TX, ⁵Baylor Coll. of Med., HOUSTON, TX, ⁶BCM

Abstract Body:

The Texome project provides exome sequencing for underserved Texans with undiagnosed disease. Participants lack access to clinical genomic sequencing, due to gaps in insurance coverage or cost of testing. As a result of this criterion and the racial and ethnic diversity of Texas, the participant population is more diverse than those typically recruited for studies of genomic attitudes. In this study, we measured genomic knowledge using the University of North Carolina Genomic Knowledge Scale (UNC-GKS), optimism and pessimism using the Genomic Orientation Scale (GOS), and empowerment using the Genomic Empowerment Scale (GEmS). We used unpaired T-tests to compare genomic knowledge, orientation, and empowerment scores by income level, insurance status, and self-reported ethnicity. A total of 26 survey responses were analyzed, including caregivers of children (n=10) and adults (n=8) and eight adult probands. Twelve respondents were bilingual; all endorsed speaking English comfortably. Two were primary Spanish-speakers. Two reported having a Bachelor's degree or higher and 23 reported a high school education or less. The most common self-reported race or ethnicity was Hispanic (n=16); followed by White (n=8); Black (n=4); Native American, American Indian, or Alaska Native (n=4); and Asian (n=1). Six reported having no health insurance, 16 reported a government plan such as Medicaid, and five reported private health insurance. We compared genomic knowledge, empowerment, optimism, and pessimism by income (\geq vs $<$ \$40,000), insurance status (insured vs not insured), and ethnicity (Hispanic vs non-Hispanic). Individuals earning $<$ \$40,000 per year (n=18) had higher scores on the Meaning of Diagnosis GEmS subscale compared to those earning more (n=8) (p=0.03), indicating that they felt more likely to receive a genetic diagnosis and that the diagnosis would improve the proband's life. While the difference was not statistically significant, higher earners had higher genomic knowledge (p=0.06). No significant differences were found between insured (n=20) vs uninsured individuals (n=6) or between Hispanic (n=16) vs non-Hispanic individuals (n=10). Higher genomic knowledge was moderately negatively correlated with higher genomic pessimism (r=-0.45). These preliminary findings suggest that individuals with lower income may have greater expectations and hopes for the meaning of a genomic diagnosis. While many comparisons did not yield significant differences, we were limited by small sample sizes. As the Texome project continues to accrue participants, we anticipate valuable insight into the genomic perceptions of an underrepresented population.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2190*. Healthcare provider confidence in understanding unsolicited genomic results: Insights from the eMERGE III experience.

Authors:

G. Wiesner¹, S. Ziniel², J. Madden³, S. Aufox⁴, K. B. Brothers⁵, W. K. Chung⁶, E. Clayton¹, J. Connolly⁷, R. Green⁸, K. Leppig⁹, M. Myers¹⁰, J. Ralston⁹, R. Sharp¹¹, J. Williams¹², J. Wynn¹³, C. B. Zawatsky⁸, I. A. Holm³; ¹Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Univ. of Colorado Sch. of Med., Aurora, CO, ³Boston Children's Hosp., Boston, MA, ⁴Northwesterns Univ., Chicago, IL, ⁵Univ. of Louisville, Louisville, KY, ⁶Columbia Univ., New York, NY, ⁷Children's Hosp. of Philadelphia, Philadelphia, PA, ⁸Brigham and Women's Hosp./Genomes2People, Boston, MA, ⁹Kaiser Permanente of Washington, Seattle, WA, ¹⁰Univ Cincinnati, Cincinnati, OH, ¹¹Mayo Clinic, Rochester, MN, ¹²Geisinger, Danville, PA, ¹³New York Presbyterian Columbia, New York, NY

Abstract Body:

Introduction: As genomic medicine expands into all areas of medicine, health care providers (HCPs) will be confronted with interpreting unsolicited genomic test results (UGR) that the HCP did not order and that may or may not be related to the patient's health history. Research has shown that HCPs feel unprepared to understand and apply results of genomic tests. We sought to assess HCP confidence after receiving UGRs from the Electronic Medical Records and Genomics Phase III Network (eMERGE), where 21,915 participants at 10 clinical sites underwent genetic testing and return of 67 medically-actionable genes. **Methods:** 484 HCPs who received a pathogenic or likely pathogenic genomic result on an eMERGE participant within the previous 6 months were invited to complete an online survey. A composite score (low, medium, or high) was developed from the responses to 4 questions where HCPs rated their confidence in (1) knowledge of the returned variant, (2) ability to explain and (3) answer patient's questions, and (4) manage the care based on the result. Design-based Pearson's Chi Square tests were performed to account for the clustering of HCPs within eMERGE sites. The primary outcome was the confidence of HCPs in addressing the UGR with their patient. **Results:** 151 HCPs completed the survey (31% response rate). 83% were physicians, 54% female, 80% non-Hispanic White, and 63% had been in practice >10 years. Among the 95 (63%) HCPs who recalled receiving the test result, most did not feel confident in explaining it to their patients (57%), answering questions (64%), or managing patients' care based on the UGR (62%). The percentage of HCPs with a low, medium, or high composite confidence score (Cronbach's alpha: 0.96) were 25%, 40% and 35% respectively. Higher confidence was associated with HCPs who had previously ordered clinical genetic tests (p=0.014). HCPs with high confidence scores were also more likely to have a positive reaction (p=0.044) and less likely to be ambivalent (p=0.006) about the UGR. **Conclusions:** This novel study examines HCPs perceptions after receiving actual UGRs on clinic patients. Over half of our respondents self reported medium to low confidence, suggesting that HCPs do not feel prepared for UGR. We also found that a positive HCP reaction to receiving the UGR was strongly associated with higher confidence in their ability to understand and discuss the UGR with their patients. Our findings provide unique insights into the "real world" implementation of genomic medicine into healthcare that can be used in future efforts to elucidate best practices in returning UGR to HCPs. This study was supported by NHGRI R01HG010004.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2191. Healthcare providers' experiences counseling patients with results from consumer genomic testing.

Authors:

M. Trottier¹, D. Green¹, H. Ovidia¹, A. Catchings¹, J. Gruberg¹, V. Groner¹, S. Dandiker¹, A. Izhar¹, K. Blazer², J. Hamilton¹, K. Offit¹; ¹Mem. Sloan Kettering Cancer Ctr., New York, NY, ²City Hope Natl. Med. Ctr., Duarte, CA

Abstract Body:

Introduction: Consumer genomic testing (CGT) includes direct-to-consumer testing and consumer-initiated testing where individuals order a test from their physician or a laboratory-affiliated provider. CGT can include analysis of a biospecimen or raw genomic data. CGT increases genetic testing access; yet, it has limited regulatory oversight, resulting in a higher potential for error in analysis or interpretation and subsequent adverse medical and psychological risks. To assess the current state of CGT, we surveyed healthcare providers' experiences with CGT in oncogenetics and other clinical settings. **Methods:** We emailed a retrospective anonymized survey about experiences counseling individuals who obtained CGT to providers in the National Society of Genetic Counselors, the Clinical Cancer Genomics Community of Practice, and regional genetic-focused professional societies in Feb-Mar 2022. **Results:** Of 139 respondents, 123 (89%) identified as a genetic counselor (GC) and the rest as an MD/DO or other provider. Most (62%) had oncology as a clinical focus, and 88% counseled individuals after CGT within the past 3 years. When asked whether potential benefits of CGT outweigh potential harms, 21% agreed, 41% disagreed, and 38% were undecided. Those who reported seeing more CGT patients were more likely to indicate benefits outweigh harms ($p=.003$). Of respondents, 69% reported experiencing cases of raw data reports not being confirmed, 76% documented testing leading to an adverse psychosocial event, and 68% reported incorrect variant interpretation; the latter 2 events were reported more often by those who felt CGT harms outweigh benefits ($p\leq.01$). Oncology providers reported higher rates of being unable to confirm a CGT result upon repeat testing in a reference laboratory ($p=.03$). Open-ended responses identified additional CGT challenges, including patient understanding of test results, false negatives, concerns about patient communication with other providers, inappropriate medical care, testing in minors, financial concerns, and incorrect test ordered; 7 providers noted positive patient CGT outcomes. Finally, 34% of respondents consulted on circulating tumor DNA results, suggesting a new mechanism for CGT findings to be brought to clinic. **Conclusion:** GCs and other providers counseling patients for a previously obtained CGT result encounter challenges including unconfirmed results, incorrect interpretation, and adverse psychosocial sequelae. Benefits of CGT were also reported. Medical risks of CGT may be mitigated by augmented regulatory oversight of commercial laboratories providing CGT, and improved consumer and provider awareness.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2192. Homozygous *APC* p.I1307K Mutations in an Ashkenazi Jewish Female: Implications for Genetic Counseling and Clinical Management

Authors:

T. DeMarco, K. Matthijssen, M. Turner, D. Alaeddin, K. Schwarting; Inova Schar Cancer Inst., Fairfax, VA

Abstract Body:

Approximately 10% of colorectal cancer diagnoses occur in the context of an inherited cancer predisposition syndrome. One example of a condition associated with an inherited predisposition to colorectal cancer is classic or attenuated Familial Adenomatous Polyposis (FAP/AFAP), caused by mutations in the *APC* gene. It is estimated that about 10% of individuals of Ashkenazi (Eastern European) Jewish ancestry carry a founder mutation in the *APC* gene known as p.I1307K. The p.I1307K mutation does not cause FAP/AFAP, but is associated with a moderately increased lifetime risk for colorectal cancer. We present the case of an 82-year old Ashkenazi Jewish female with a personal history of multiple primary cancers, including breast cancer at the age of 59, lung cancer at the age of 71 (in the context of former smoking history), and an endometrial neuroendocrine tumor at the age of 82. In addition, the patient underwent a hemicolectomy at the age of 81 due to the presence of a significant tubulovillous adenoma. She reported a family history of breast and prostate cancer in her siblings and a gastrointestinal malignancy in her mother. In light of her personal and family history of malignancy, the patient recently underwent genetic counseling and testing utilizing a comprehensive pan-cancer panel at a commercial laboratory. The results indicated homozygous germline *APC* p.I1307K mutations. While there is a recent case series regarding four homozygous *APC* p.I1307K carriers (Rosenblum et al., Case Rep Oncology: 2021), there is no data currently available regarding the lifetime risk for cancer in these patients or appropriate medical management for these individuals. It stands to reason that this result may have been a contributor to the development of the large adenoma in this patient, but it is not clear what role, if any, it may have played in the development of her other cancers. Given the incidence of heterozygous *APC* p.I1307K mutations in the Ashkenazi Jewish population, it would be expected that the homozygous state would be reported with greater frequency than it is at present. However, there is currently a paucity of data in the published medical literature regarding *APC* p.I1307K homozygosity and the implications for cancer screening and patient-centered genetic counseling. This case contributes to the medical knowledge about the clinical variability in these patients and the need for further academic collaboration to refine cancer risks and establish mutation-specific medical management guidelines for homozygous *APC* p.I1307K mutation carriers.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2193. Identifying ELSI needs in newborn screening research.

Authors:

C. Lumpkins¹, A. Goldenberg², E. B. Goldman³, I. A. Holm⁴, A. Brower⁵; ¹ACMG, Bethesda, MD, ²Case Western Reserve Univ, Cleveland, OH, ³Univ. of Michigan, Ann Arbor, MI, ⁴Boston Children's Hosp, Boston, MA, ⁵American Coll. of Med. Genetics and Genomics, Dakota Dunes, SD

Abstract Body:

Introduction: Each year in the United States, 3.8 million newborns are screened for up to 81 genetic conditions, and over 12,000 are diagnosed and referred to clinical care. Discoveries of novel technologies to screen, diagnose, and treat genetic diseases have led to an ever-increasing list of conditions that are candidates for newborn screening (NBS). The NBS community of researchers, parents and families, advocacy groups, healthcare professionals, and state NBS programs play important roles in the research that leads to these advancements. Because NBS research often involves and impacts newborns and children, consideration of the ethical, legal, and social issues (ELSI) is crucial during the planning, execution, and reporting of NBS research. To better understand the ELSI concerns of NBS researchers and state programs, NBSTRN is conducting a survey to identify the specific challenges faced by those working in newborn screening.

Methods: The *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) Hunter Kelly Newborn Screening Research Program funds the Newborn Screening Translational Research Network (NBSTRN) operated by the American College of Medical Genetics and Genomics (ACMG). The NBSTRN Bioethics and Legal Workgroup developed a survey of NBS researchers and state programs with the goal of assessing the nature and scope of the ethical and legal challenges faced in their work and how those needs can best be met in the form of resources, tools and/or education. The survey included questions relating to a range of ELSI topics including, but not limited to, data sharing, consent, return of results, and including diverse perspectives in study design. The survey will be disseminated in summer 2022 to NBS researchers and state programs. Data obtained by the survey will be reviewed and used to create tools, white papers and recommendations for practice.

Results: This presentation will explore the preliminary results of the 2022 ELSI Survey and potential ways NBSTRN can address the ELSI needs of NBS researchers and state programs.

Conclusion: NBS researchers and state programs face challenges related to the ethical, legal, and social issues of newborn screening in their work. It is imperative that the nature and scope of these challenges is well understood and used to inform the development of relevant resources to support ethical NBS research. Although this effort is focused on programs related to newborns and children, our findings inform the use of genomics across the lifespan and provide a useful model for others looking to explore ELSI concerns in other areas of genomics research or care.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2194. Implementation and evaluation of digital patient-reported oncology questionnaire at the East Genomic Medicine Service Alliance

Authors:

P. Buczkowicz¹, E. Peacock¹, E. Davies², V. Kiesel³, R. N. Sandford², O. Buske¹; ¹PhenoTips, Toronto, ON, Canada, ²Cambridge Univ. Hosp.'s NHS Fndn. Trust, Cambridge, United Kingdom, ³Univ. Hosp. of Leicester NHS Trust, Leicester, United Kingdom

Abstract Body:

The East Genomic Medicine Service Alliance (E-GMSA) is undergoing a digital transformation in order to improve and standardize the delivery of genomic medicine across the region, serving a population of over 10 million. Direct patient engagement in the design and delivery of aspects of the digital transformation is fundamental to the success and widespread adoption of new and innovative solutions across genetics, other medical specialties and primary care within the region. Here we implement, gather feedback, and improve an oncology-focused family history and medical history questionnaire in preparation for region-wide deployment.

A total of 100 consented individuals from the E-GMSA who had previously filled out paper-based questionnaires as part of their referral to cancer genetics were provided access to a secure digital questionnaire. The questionnaire collected information pertaining to the patient's demographics, medical history with a focus on oncology and family history. Information from submitted questionnaires was recorded in PhenoTips software, including auto-drawn pedigree diagrams accessible to clinical staff for review and further refinement, raw questionnaire responses attached to the patient record in PhenoTips, and standardized Human Phenotype Ontology (HPO) terminology, where applicable. Furthermore, the patient-provided responses combined with available clinical data were used to test risk assessment models directly incorporated into the PhenoTips pedigree tool, including BOADICEA v5 (Cambridge Enterprise Ltd.), IBIS/Tyrer-Cuzick v8 (Cancer Research UK), GAIL (NIH National Cancer Institute) and PREMM5 (Dana-Farber Cancer Institute). Several metrics comparing respondent experience with digital versus paper-based questionnaires across various parameters, including ease of use, time to completion, comprehensibility, completion rate, time per page, average number of sessions, and net promoter score were collected. Feedback was assessed and incorporated into the oncology questionnaire prior to region-wide rollout.

The success of this pilot in the oncology setting has provided opportunities to expand to other specialties such as inherited cardiovascular conditions, and our work represents the first real-world implementation of digital questionnaires through PhenoTips as part of the E-GMSA project and demonstrates the utility of standardized pedigree and genomic data sharing in a clinical setting.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2195. Implementing human genetics training and research at a university hospital in Nairobi, Kenya.

Authors:

S. Ilovi, J. Mecha; Univ. of Nairobi, NAIROBI, Kenya

Abstract Body:

Medical Genetics in Kenya is still an underutilized service, largely driven by inadequate training of healthcare providers in medical genetics and a paucity of medical geneticists. Human genetic testing are largely carried out abroad because of low numbers; these being largely driven by prenatal genetic tests with some laboratories reporting less than five genetic tests annually for patients suspected to have genetic disorders. The bulk of genetic tests are limited to sequencing of pathogens such as SARS-CoV-2 and influenza. The University's curriculum prescribes teaching of medical genetics for both undergraduates and postgraduates. These training has been limited to cell biology with biochemistry. In 2022, we re-instituted clinical genetics teachings for medical students rotating in Internal Medicine in their fourth year and sixth year. The sessions had a reach of over 500 medical students. Similarly, we augmented clinical genetics training for first, second and third year residents in Internal Medicine. The University has recently established a Sanger based molecular laboratory which proved pivotal in management of the COVID-19 pandemic, HIV drug resistance analysis and also supporting molecular based PhD studies. We are currently instituting basic genetics research in sickle cell disease at our institution. It is envisioned that these endeavors will promote human genetics training and research in population genetics as well as improved diagnosis and management of patients with genetic and rare diseases in Kenya.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2196*. Implementing public health genomics: The familial hypercholesterolemia model

Authors:

L. Manace, R. Birnbaum, A. Avins, L. Taylor, A. Curran; The Permanente Med. Group, Oakland, CA

Abstract Body:

Background: Familial hypercholesterolemia (FH) is the most common autosomal dominant disorder, affecting 1:225 individuals worldwide. Diagnostic criteria, associated genes, and treatment guidelines are well established with proven morbidity and mortality reduction. Despite longstanding public health recommendations for health systems to identify and treat FH, only 5% of the population is currently appropriately diagnosed and managed. Population screening taking advantage of electronic medical records with referral to a specialized care-and-tracking program taking advantage of streamlined molecular genetic testing has the potential to be sustainable with cost savings based on improved outcomes. Objective: Develop and evaluate a pragmatic, phased implementation of FH as a model for precision medicine and genomics in a large integrated health system. Methods: Using the Make Early Diagnosis/Prevent Early Death (MEDPED) criteria, we created a registry of likely FH individuals 18 years and older within Kaiser Permanente Northern California enrollees (4.5 million members). This led to a pragmatic workflow for inviting members in partnership with their primary care providers to be evaluated in a specialty care program including cascade adult and pediatric relative assessment and longitudinal follow-up. Results: 93% of patients identified as likely having FH were confirmed to be affected with FH either with genotypic pathogenic or likely pathogenic variant confirmation (67%, 91% in LDLR gene 9% in APOB gene) or according to clinical diagnostic criteria including personal and family medical history with or without physical stigmata. 67% of offspring were confirmed to have inherited the causative gene variant and/or met clinical diagnosis of FH based on biochemical laboratory values. 48% of adults and 100% of affected children were on no lipid-lowering medication at the time of screening. Following initiation of therapy, there was a 41% median reduction in low-density lipoprotein cholesterol (LDL-C) levels. Conclusions: Familial hypercholesterolemia gives health systems the opportunity to realize significant primary and secondary prevention. Here we describe a pragmatic and sustainable approach to integrating genomics in a streamlined care pathway within pediatric and adult care in a model that can be adapted in other health organizations.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2197. Improving Community Based Genomics Research in Black Populations: A Review of Lessons Learned and Ideas for Need Based Solutions

Authors:

C. Whitted¹, T. Lee², F. Gilpin-Macfoy¹, R. Tawk², J. Luque², J. Reese-Smith³, L. Koehly¹; ¹NHGRI/NIH, Bethesda, MD, ²Florida A&M Univ., Tallahassee, FL, ³City of Houston Fire Dept., Houston, TX

Abstract Body:

Recruitment of historically marginalized populations in genomic research is an ongoing challenge, limiting generalizability of genomic discoveries and implementation of precision medicine initiatives. Fewer Black Americans have been recruited to participate in genomic research than White Americans, suggesting that translational discoveries from genomic research may not be generalizable to Black Americans, thereby increasing the potential for health inequities for this marginalized population. Many factors, such as historical mistrust of research due to past negative experiences, continue to impact research participation. This proposed paper perspective provides insights towards reaching Black Americans for genomics research based on lessons learned from a community-based genomics research project initiated in North Florida evaluating the Families Sharing Health Assessment and Risk Evaluation (SHARE) toolkit and ideas to help strengthen community-based research. Families SHARE aims to increase community members' understanding of the importance of family health history for their personal and family members' risk of common, complex conditions. Here, we present a review of the recruitment and community engagement methodology used to evaluate Families SHARE in North Florida Black communities and lessons learned from this experience that can guide future genomic research endeavors that aim to reach historically marginalized communities. Ideas and recommendations will be shared to improve community engagement in genomics research which includes technology needs to improve data collection, sharing of family health history and educational toolkits to improve genomic research within community settings and communities.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2198. Incidental findings in study participants: What is the researcher's obligation?

Authors:

D. Schaare, L. Boccuto, S. Sarasua, L. Ward; Clemson Univ., Clemson, SC

Abstract Body:

Background: With genomic testing becoming a standard diagnostic tool, incidental findings, or the discovery of unrelated results, have become commonplace. Due to the severity of diseases linked to specific pathogenic variants, the American College of Genetics and Genomics (ACMG) has established guidelines for clinicians to return findings on 50 plus genes associated with medically actionable conditions. However, in the research setting, the lack of a definitive guiding policy for investigators compounds ethical dilemmas that may arise when incidental findings that are clinically actionable occur in the conduct of research studies. The concern for potential loss of clinically valuable information could undermine the research subject's trust and commitment to participate in a process that offers more individual risk than benefit. **Purpose:** The analysis aimed to ascertain the most ethically valid approach to manage incidental research findings, offering a useful approach for investigators and participants. **Methods:** We performed an ethical evaluation using Wueste's Framework of Identify, Analyze, Justify, and Decide (AIJD). We then analyzed current research policies, as well as an alternative, to determine viability. **Results and Discussion:** The current policy of allowing the decision to return incidental findings up to the researcher fails the majority of ethical principles, including the consequential, deontological, and intellectual freedom perspectives. We find that by applying the ACMG guidance for clinicians to investigators resolves many of the ethical challenges and provides needed guidance to researchers. This presentation integrates suggestions from multiple organizations to establish a favorable policy for both the investigator and the research subject that will achieve the HHS aim, stated in the 2014 Joint Rule, of transitioning participants to "partners" in research.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2199. Individual attitudes towards identity, privacy, and health insights from personal genomics: Evidence from a full archival search on Twitter

Authors:

B. Zhao, M. Mills, E. Akimova; Univ. of Oxford, Oxford, United Kingdom

Abstract Body:

Abstract

In recent years, the massive growth of direct-to-consumer genomics testing services (DTCG) has raised public attention and posed pressing regulatory questions. However, to date, very little is known about how the newly found accessibility of personal genome information is perceived by individuals, and accompanied concerns on privacy, self-identity, and health insights. We collected all public tweets related to two major DTCG companies from 2006 to 2020 (n=1.6M, from 757,750 individual users), and analysed word, terminology, topic, and sentiment prevalence. We discovered on average most tweets are positive, even among concerning topics that are privacy related or classified as hate or offensive language. However, our content analysis identified several words clearly related to customer concerns; it is apparent that people express interest in such topics as ancestry, privacy, and risks.

Results

Monthly frequencies over the period from 2007 to 2020 shows online discussions experienced a substantial increase since 2014, in line with the steady and rapid growth of new genetic tests on the market. Percentage of organic tweets has been gradually declining over the years, negatively correlated with the annually increased DTC medical marketing spending in the US, in line with the prevalence of Western users mainly from the US and Europe.

Using most frequently used words via a simple case insensitive bag-of-words approach, we were able to identify several words that are clearly related to customer concerns such as *don't*, *fda* and *private*. In our LDA topic modelling analysis, we explored the models with topics n=8, chosen based on local maximum coherence score and visualised intertopic distances with the fewest overlapped areas.

Among the sentiment analysis on English tweets, we discovered on average most tweets are sentimentally positive, even among concerning topics that are privacy related or classified as hate speech or offensive language, possibly due to Twitter enforces relevant policies and restrictions to minimise harmful contents.

White supremacy is often behind the motivations for personal genetics testing, which may lead to racist speeches. Our preliminary analysis shows the percentage of offensive tweets remained stable until 2016, then fluctuated afterward. In comparison, hateful tweets remained low despite two significant spikes.

To understand opinions towards health risk reports, we extracted 493 terms and 378 concepts with semantic type *dsyn*. Tweets containing clinical terms has a constant increase starting 2020 with some previous spikes in 2015 and 2017. Importantly, in 2017 such tweets were negative in sentimental analysis.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2200. Integrating Predictive Genetic Testing into clinical care in the Preventive Genomics Clinic

Authors:

C. Srisukajorn^{1,2}, P. Teekakirikul¹, R. Ngathong¹, S. Kongmongkol¹, T. Viangteeravat³, J. M. Ford³; ¹Preventive Genomics Clinic and Family Check-up Services, Dept. of Med., Bumrungrad Intl. Hosp., Bangkok, Thailand, ²BIR Program, Faculty of Political Sci., Thammasat Univ., Bangkok, Thailand, ³Res. and Dev., Bumrungrad Intl. Hosp., Bangkok, Thailand

Abstract Body:

Background: Emerging insights into genomics has yield great benefits in modern medicine. Technological advancements made genetic tests readily accessible and affordable to general population. Many commercial test panels have been available to predict future disease risk in asymptomatic individuals. Yet, there are little data on how well predictive genetic testing has been integrated in the health space. **Aims:** To assess the trends of integrating genomics into healthcare at a quaternary care hospital and review how actions were taken after likely pathogenic/pathogenic variant(s) (PVs) were identified from either genetic screening test (GST) or diagnostic test (DT) panel according to the ACMG guidelines. **Methods:** The data was collected for 365 days from May 2020 using the number of index patients (PT) who received services at Preventive Genomics Clinic at Bumrungrad International Hospital. Patients were categorized in two groups: healthy and diseased. How they were referred to our clinic, types of services, management plans proposed by the board-certified geneticist, and actions taken from post-test counseling were recorded and examined. **Results:** Across the year, the clinic served 387 PT ranging from 2 to 93 years old (mean=44), 78.3% of which are healthy, the rest were diagnosed with certain disease(s) prior to the visit. We found 35.4% purchased GST panel prior to the visit while 27.9% were referred for genetic counseling from other clinics mostly by Ob-Gyn (n=21), Check-up Center (n=17), and oncology (n=16). Also, 49.6% acquired GST to see predisposition to inherited cancer and heart diseases while 26.4% received further genetic test using DT panel for various diseases including cancer. Of all population, 294 PT (76%) who obtained either GST or DT, PVs were detected in 29 PT (9.9%) from both healthy and diseased group. PVs detection rates in healthy PT were 6.9%. Post-counseling actions include follow-up (F/U) at the clinic, referring PT to other specialist(s), and cascade genetic testing (CGT) to determine heredity risks. F/U visits were appointed for 14 PT. Ten were referred to other specialist(s), 4 of which did show up for the referral visit, 1 was booked, and the rest were not indicated. CGT was recommended to all PT with PVs, only 10 of which have their family members tested, ranging from 1-14 members. Total 47 family members were presented at the clinic for F/U, 44.7% were found to have the family PVs. **Conclusion:** Predictive genetic testing has great potential for appropriate screening and preventive strategies. The results prompted appropriate surveillance, further referral to specialists, regular clinical follow-up, and follow-up family variant testing.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2201. International Best Practices for Genomic Data Sharing in the Health Technology Industries

Authors:

B. Knoppers, M. Zawati, Y. Joly; McGill Univ., Montreal, QC, Canada

Abstract Body:

Health technology industries play a central role in supporting the introduction of genomics into healthcare and must adapt to data sharing norms promoting collaboration, creative data re-use, and scientific reproducibility. Data sharing presents opportunities for industry, but also particular interests and challenges. Industry could therefore benefit from greater clarity and consistency over what constitutes responsible data sharing practices. An Industry Core Group was formed to develop International Best Practices for Genomic Data Sharing, consisting of representatives from 10 companies, working in partnership with the Centre of Genomics and Policy at McGill University. The Group's mission is to facilitate and promote the international harmonization and sharing of genomic and health data by industry. The effort is inspired by the Framework for Responsible Sharing of Genomic and Health-related Data developed by the Global Alliance for Genomics and Health, founded on the human rights of everyone to share in scientific advancement and its benefits, and to the protection of the moral and material interests resulting from any scientific production. The best practices are based on a systematic review of legislation, guidelines, and industry practices. Tracking the Framework's core principles, the Industry Core Group has identified 8 interrelated themes to be addressed in the best practices: Data control/sharing - how to reconcile commercial interests in data control with opportunities to generate commercial and societal value through data sharing? Identifiability - how to determine the applicability of diverse international data privacy laws and regulation? Consent - how to obtain appropriate individual permissions and to ensure commitments to patients and participants providing their data are respected when sharing internally or externally? Access - how to provide individuals and researchers access to data while protecting commercially sensitive information? Privacy and Security - how to provide for consistent and effective levels of protection for sensitive genomic and health data? Governance - how to realize commercial and societal value from data while respecting requirements from data protection law, research ethics and human rights? Partnerships - how to establish transparent, fair, and mutually beneficial partnerships involving industry, academia, as well as patient groups and publics? Intellectual Property - how to reconcile data sharing with securing intellectual property rights and commercial secrets? The Core Group seeks to dynamically engage the broader genetics community to ensure wide acceptance of the best practices.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2202*. Is it time for a paradigm shift? Design and formative evaluation of a randomized clinical trial of the sequence of genetic counseling and testing to optimize efficiency, patient empowerment and engagement, and medical adherence for cardiovascular genetic testing indications (*RESEQUENCE-GC*).

Authors:

C. Pendleton¹, C. Applegate², C. Tichnell¹, B. Murray¹, B. Rivers¹, R. McClellan¹, E. Brown¹, A. Barth¹, L. Yanek³, C. Overby Taylor⁴, J. Day⁵, C. A. James¹; ¹Div. of Cardiology, Dept. of Med., Johns Hopkins Univ., Baltimore, MD, ²McKusick-Nathans Dept. of Genetic Med., Johns Hopkins Univ., Baltimore, MD, ³Div. of Gen. Internal Med., Dept. of Med., Johns Hopkins Univ., Baltimore, MD, ⁴Dept. of BioMed. Engineering, Johns Hopkins Univ., Baltimore, MD, ⁵Dept. of Art as Applied to Med., Johns Hopkins Univ., Baltimore, MD

Abstract Body:

Genetic test (GT) results are increasingly critical to medical management. Traditional pre-test genetic counseling (GC) may facilitate informed decision making but is a constrained resource. Post-test GC can be tailored to GT results and may be optimal for adults seeking guideline-directed GT. RESEQUENCE-GC is an NHGRI-funded 3-arm trial to evaluate 2 complementary approaches to shifting the primary GC session post-test for adults with a cardiovascular GT indication. Here we describe RESEQUENCE-GC design and formative evaluation. RESEQUENCE-GC participants are randomized to 1) pre-test GC (“standard of care” arm) or 2) a pre-test genetic education video with an optional (“efficiency” arm) or 3) required (“flipped” arm) call with their GC before GT. Results are returned by phone (1) or during post-test GC (2,3). Primary outcomes are patient empowerment, psychosocial well-being, medical adherence, and GC efficiency. Preparatory to the trial, animated pre-test genetic education videos are being developed using principles from the Cognitive Theory of Multimedia Learning, with iterative professional and patient stakeholder review. The video for cardiomyopathy or arrhythmia panel GT was tested with adults from the general population recruited via a usability testing company (UserHappy) and members of a patient support group (Hope for ARVD). Participants were emailed a questionnaire link that included measures of pre-genetic education knowledge [Genomics Knowledge Scale; pilot 8-item Multidimensional Model of Informed Choice (MMIC)], the video, engagement questions [User Engagement Scale (UES); 4 items measuring trust in the video content and comfort making a GT decision], and post-genetic education knowledge. Participants (N=48) were 54% male, 42% female, 4% non-binary; 36% Black, 52% White, 4% Asian, 8% Multiracial; and 31% age 18-34, 36% 35-44, 25% 45-54, 8% ≥55; similar to the RESEQUENCE-GC recruitment pool. They had high levels of knowledge and engagement post-genetic education, averaging 7.4/8±0.9 correct MMIC responses and a 4.14±0.5 UES score (range 1-5). Trust in genetic education and comfort making a GT decision were highly rated (4.61±0.7 and 4.48±0.73 respectively); particularly among support group members (4.84±0.4 and 4.6±0.8; N=16)]. In resource-constrained environments, models to efficiently provide GT education are crucial. Our findings suggest that our theory-driven and iterative strategy has produced pre-GT education videos that are both trusted and result in high comfort making a GT decision. This finding was particularly true for cardiovascular disease patients, the target population for RESEQUENCE-GC.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2203. Lessons learned from population-based biobank participants' responses to individual genomic results

Authors:

A. Metspalu^{1,2}, A. Reigo¹, M. Palover³, T. Nikopensus¹, K. Läll³, K. Krebs³, S. Reisberg^{4,5}, R. Mägi³, M. Kals^{3,6}, H. Alavere³, M. Nõukas³, A. Kolk⁷, I. Normet⁸, M-L. Tammesoo³, E. Käär⁹, M. Puusepp³, K. Metsalu³, A. Allik¹⁰, K. Fischer^{3,9}, L. Milani³, N. Tõnisson^{1,11}, L. Leitsalu¹²; ¹Estonian Genome Ctr., Inst. of Genomics, Univ Tartu, Tartu, Estonia, ²Inst. of Molecular and Cell Biology, Univ Tartu, Tartu, Estonia, ³Estonian Genome Ctr., Inst. of Genomics, Univ Tartu, Tartu, Estonia, ⁴Inst. of Computer Sci., Univ Tartu, Tartu, Estonia, ⁵STACC, Tartu, Estonia, ⁶Inst. for Molecular Med. Finland, HiLIFE, Univ Helsinki, Helsinki, Finland, ⁷Inst. of Clinical Med., Univ Tartu, Tartu, Estonia, ⁸Family Med. Ctr. of Medicum, Tallinn, Estonia, ⁹Inst. of Mathematics and Statistics, Univ Tartu, Tartu, Estonia, ¹⁰Estonian Res. Council, Tartu, Estonia, ¹¹Tartu Univ. Hosp., Tartu, Estonia, ¹²Estonian Genome Ctr., Inst. of Genomics, Univ Tartu, Tartu, Estonia

Abstract Body:

The Estonian Biobank has over 200 000 individuals in the cohort, which is close to 20% of adult population of the country and one objective was, in addition to genomic research, to use the data for improving the public health. The whole biobank is genotyped, has longitudinal health data and finally we are moving towards the implementation of the genomic information in the healthcare. The return of individual genomic results (ROR) to research participants is still in its early phase, and insight into how individuals respond to ROR is scarce. Studies contributing to the evidence base for best practices are crucial before the latter can be established. Here, we describe a ROR procedure conducted at a population-based biobank, followed by surveying the responses of almost 3000 participants to a range of results, and discuss lessons learned from the process, with the aim of facilitating large-scale expansion. Overall, participants perceived the information that they received with counseling as valuable, even if the reporting of high risks initially caused worry. The face-to-face delivery of results limited the number of participants who received results. Although the participants highly valued this type of communication, additional means of communication need to be considered to improve the feasibility of large-scale ROR. The feedback collected sheds light on the value judgements of the participants and on potential responses to the receipt of genetic risk information. Biobanks in other countries are planning or conducting similar projects, and the sharing of lessons learned may provide valuable insight and aid such endeavours.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2204*. Low interest and high drop off in population genetic screening of adults: Results from a "real world" pilot study.

Authors:

N. Rao, J. Kaganovsky, S. M. Fullerton, A. T. Chen, B. H. Shirts; Univ. of Washington, Seattle, WA

Abstract Body:

Population genetic screening may identify individuals at risk for preventable hereditary cancers or hypercholesterolemia and potentially address disparities in genetic testing. Learning why people may or may not engage in population screening is important to assess the acceptability of this strategy and understand if it will lead to more equitable access. We sought to understand participation rates and decision-making about screening among a diverse cohort invited to participate in the University of Washington Medical Center's (UWMC) population genetic screening study.

This study invited adults unselected for medical history who have received care at the UWMC to participate in genetic screening for hereditary cancers and familial hypercholesterolemia between June 2020 and July 2021. Email invitations contained a link to a study FAQ describing risks and potential benefits. People had the option to participate in genetic screening and provide feedback about their reasons for participating in a series of surveys. Uninterested individuals could provide reasons for nonparticipation in a free text box prior to or after the initial survey. Participation trends were compared across demographic variables and reasons for participation or non-participation were evaluated using quantitative and qualitative methods.

A total of 40,855 people were invited to participate in the study, and 2,855 (7%) people returned signed consent forms and DNA samples. The average age of enrollees was 40 years old, and 60% were female. Participation rate was lowest among Black individuals (3.33%) and highest among Multiracial or Other Race/ethnicity identifying individuals (12.99%). Learning about disease risk and identifying risk early for prevention purposes were strong motivators for receiving screening. Compared to other participants, Asian participants considered knowing more about the test and test accuracy factors of greater importance during screening decision making. Analysis of 139 survey free text responses identified 3 themes related to nonparticipation: benefits do not outweigh risks, don't want to know, and challenges with study logistics.

Uptake of genetic screening in the UWMC study is low compared to screening studies among biobank participants. However, individuals contributing to biobanks may be more interested in research and genetic findings generally. Our study may more closely mimic implementation challenges and participation outcomes in a stand-alone genetic screening program among the general population and suggests strategies that could be pursued to enhance uptake in future screening programs.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2205. Masking for whole-exome sequencing data: An analysis of the impact of masking choices in rare coding variants association tests.

Authors:

T. Nguyen¹, R. Koesterer¹, O. Ruebenecker¹, A. Moriondo¹, Q. Hoang¹, D-K. Jang¹, J. Flannick²; ¹The Broad Inst. of MIT and Harvard, Boston, MA, ²Boston Children's Hosp., Boston, MA

Abstract Body:

Whole exome sequencing (WES) has emerged as a valuable counterpart to genome-wide association studies (GWAS) because it helps implicate causal genes via rare variants associations. However, there are many different annotation algorithms and many more ways to combine rare variants to perform filters, and there is no consensus on how the results of gene-level analyses depend on grouping strategies (i.e., masks). In this paper, we analyze the impact of masking choices on association results in WES studies by conducting SKAT and burden testing and downstream analyses for various traits. Our findings suggest that as variants get rarer, higher effect sizes are observed although they cannot overcome the lack of power; in addition, masks that are more lenient than loss of function produce more gene-level associations although they may be less biologically relevant. This implies the importance of using multiple masks to aggregate variants into genes to gain more insights into trait/gene associations and ultimately, disease biology.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2206. Medical genetics and genomics services *across the lifespan*: Patient and referral trends over 17 years at the Maritime Medical Genetics Service

Authors:

M. Mackley¹, S. Goobie^{2,3}, A. Alshammari⁴, A. Miller⁴, M. Ari^{4,5}; ¹Div. of Clinical and Metabolic Genetics, Dept. of Pediatrics, The Hosp. For Sick Children, Toronto, ON, Canada, ²Dept. of Pediatrics, Dalhousie Univ., Halifax, NS, Canada, ³Maritime Med. Genetics Service, IWK Hlth.Ctr., Halifax, NS, Canada, ⁴Dept. of Med., Dalhousie Univ., Halifax, NS, Canada, ⁵Div. of Endocrinology and Metabolism, Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Following diagnostic and therapeutic advances in genomics it is expected that patterns of access to medical genetics services have changed. Adult patients—with presentations and needs distinct from their pediatric counterparts—may be increasingly accessing genetics services. Presently, however, literature documenting patterns in access is lacking. Such information is essential to facilitate resource planning and improve care delivery to genetics patients, from prenatal through adulthood.

Maritime Medical Genetics Service (MMGS), in Halifax, Nova Scotia, Canada, is a comprehensive medical genetics service offering clinical and metabolic genetics care to patients of all ages. MMGS provides all genetics services to the three Canadian Maritime provinces, uniquely affording a complete picture of access in a discrete population of nearly 2 million. To understand how services have changed and to inform future care provision we undertook a quality improvement project where we retrospectively reviewed data collected on all patients referred to MMGS over a 17-year period (January 2002 and December 2019). To assess referral trends, patient demographic, referral source, and clinic data were extracted and subjected to descriptive analysis.

Over the 17-year period, a total of 43052 patient visits were documented, of which 27865 were new patient visits. Total clinic volumes increased steadily, with an average annual growth rate of 13.1%. The proportion of adult patients grew from 61.9% in 2002 to 73.3% in 2019, and the average age of new patients increased from 24.8 to 36.5. The clinic type that saw the most growth was cancer, making up 39.1% of new visits in the final year of data (compared to 12.5% in 2002). Of note, new visits referred from internal medicine and its subspecialties, including oncology, increased more than 3-fold over the study period—increasing from 4.9% to 16.2% of new visits. In contrast, conventional sources of referrals—family medicine, pediatrics, obstetrics and gynecology—contributed proportionally less to the referral volume over time.

This quality improvement study highlights that access to genetics services at our centre has changed, necessitating a reconceptualization in local care delivery. The increasing referral volume, as well as the trend towards increasing numbers of adult patients, highlights the need for new service delivery models, particularly for adult care. Greater consideration must be given to multidisciplinary disease-specific clinics, increased numbers of medical geneticists, expanded roles of genetic counselors and nurses, and greater empowerment of non-geneticists in the delivery of genetics services.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2207. Newborn Screening: a multidisciplinary approach based on landscape analysis, genetic screening, whole genome sequencing, and artificial intelligence to achieve an early diagnosis of rare diseases

Authors:

A. Ferlini¹, F. Fortunato¹, M. Farnè¹, R. Selvatici¹, C. Blankart², S. Gillner², R. Roettger³, J. Kirschner⁴, J. Schenk⁵, K. Zarakowska⁵, J. Zschüntzsch⁶, Y-S. Man⁷, L. Goodman⁸, M. Trad⁹, S. Sansen¹⁰, M. Zuccola¹¹, M. Martinez-Fresno¹², L. Tome¹³, L. Baillon¹⁴, E. Gross¹⁵, N. Garnier¹⁶; ¹Univ. OF FERRARA, FERRARA, Italy, ²KPM Ctr. for Publ. Management and Swiss Inst. for Translational and Entrepreneurial Med., Bern, Switzerland, ³Dept. for Mathematics and Computer Sci., Univ. of Southern Denmark, Coopenhagen, Denmark, ⁴Univ. of Freiburg, Freiburg, Germany, ⁵Takeda Pharmaceuticals Intl. AG, Neuchâtel, Switzerland, ⁶Georg-August-Univ.et Goettingen, Goettingen, Germany, ⁷Novo Nordisk Hlth.Care AG, Zurich, Switzerland, ⁸Univ. Coll. Dublin, Dublin, Ireland, ⁹Lysogene France, Paris, France, ¹⁰Sanofi France, Brussels, Belgium, ¹¹F. Hoffmann-La Roche, Basel, Switzerland, ¹²illumina, London, United Kingdom, ¹³ProQR Therapeutics, Leiden, Netherlands, ¹⁴PTC Therapeutics, South Plainfield, NJ, ¹⁵Eurordis, Paris, France, ¹⁶Pfizer Ltd, Tadworth, United Kingdom

Abstract Body:

Rare diseases (RDs) affect over 30 million people in the EU and result quality-of-life limiting and life-threatening. The RDs diagnosis is challenging, due to a multitude of different symptoms and syndromes, low prevalence, and lack of general awareness. Altogether, this leads to a lengthy and burdensome path to diagnosis, severely delayed by a “diagnostic odyssey”. Lack of timely diagnosis affects disease management and identification of potential beneficial treatments and/or clinical trials. Indeed, less than 10% of RD patients have a treatment and only 1% take advantage from approved therapy in Europe. Since 72% of RDs are genetic and 70% have an exclusively paediatric onset, strategies for genetic newborn screening (NBS) are of pivotal importance, possibly representing a particular diagnostic “check point”. The EU-IMI Research Project Screen4Care (S4C) is an international public-private collaboration of 35 partners. The 5-year project focuses on accelerating diagnosis for RDs patients through two central pillars: genetic newborn screening (NBS) and artificial intelligence (AI)-based tools. The project will drive genetic NBS in about 20.000 infants in Europe, and will design innovative gene panel including treatable (TREAT-map gene panel) and actionable (ACT-map gene panel). Whole genome sequencing (WGS) is also planned in early symptomatic infants, tested negatively during NBS. The project includes a large landscape analysis on existing knowledge about NSB worldwide and will design new Artificial Intelligence (AI)-based algorithms to identify patients using Electronic Health Records (EHR) and to build a repository “*symptom checkers*” which will allow to design an “app” user friendly for patients and healthcare providers, to help self-diagnosis and referral pathways to physicians for diagnostic workup. S4C goals will be accelerating the time to RD diagnosis and treatment, and the establishing of a digital infrastructure for families and healthcare workers. *This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 101034427. The JU receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA (www.screen4care.eu)*

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2208. Non-inferiority of letter versus telephone genetic counseling for negative exome-based cancer genetic test results disclosure in a historically underserved population.

Authors:

M. J. Gilmore¹, M. C. Leo¹, L. Amendola², K. A. B. Goddard¹, J. E. Hunter³, G. Joseph⁴, T. L. Kauffman¹, B. Rolf⁵, E. Shuster¹, J. Zepp¹, B. Biesecker⁶; ¹Kaiser Permanente Ctr. for Hlth.Res., Portland, OR, ²Univ. of Washington, Seattle, WA, ³RTI Intl., Research Triangle Park, NC, ⁴Univ California San Francisco, San Francisco, CA, ⁵Univ. of Washintgon, Seattle, WA, ⁶RTI Intl., Washington, DC

Abstract Body:

Scalable alternative delivery models for genetic counseling are needed to realize the promise of precision medicine and ensure broad and equitable access to these services. The Cancer Health Assessments Reaching Many (CHARM) study, as part of the Clinical Sequencing Evidence-Generating Research (CSER) consortium, prioritized the evaluation of genomic medicine implementation in historically underserved populations. We assessed whether a mailed letter (n=65) was non-inferior to phone genetic counseling (n=73) in informing participants, without a personal or family history of cancer, of their negative exome-based panel results. Participants recruited earlier in the study received results by phone, and participants recruited later in the study received results by letter. Using the CHARM definition for ‘historically underserved’ (non-White, Hispanic, low-income, low-education, Spanish-speaking, resident of medically underserved area, uninsured, or LGBTQ+ identity), 76% (105/138) of participants met this criteria. Data was collected via participant surveys - follow-up 1 (FU1) at 2-30 days after result disclosure, and follow-up 2 (FU2) at 5-7 months after result disclosure. Primary outcomes and non-inferiority margins were subjective understanding of results (75% ratio of means) and test-related distress using the FACToR (Feelings About genomiC Testing Results) measure (75% ratio of means). Secondary outcomes and non-inferiority margins were related to satisfaction (80% ratio of means). The satisfaction measures, assessed at FU1 only, included questions about mode of results delivery and if the “right amount of information” was provided at results disclosure. For subjective understanding of results, letter was inferior to phone at FU1 but was non-inferior at FU2. Letter was non-inferior to phone at FU1 and FU2 using FACToR (overall score). Letter was non-inferior to phone for satisfaction with mode of results delivery. Notably, letter was inferior to phone for having the “right amount of information” provided at results disclosure. The transient inferiority of a letter for subjective understanding may be acceptable for *a priori* low-risk individuals receiving negative test results, where there are not recommended health behavior changes. Limitations include a small sample size. Future applications of this alternative delivery model could explore increasing the amount of information provided in the letter. Our work provides evidence that a letter is acceptable in conveying genetic test results to *a priori* low-risk individuals in a scalable manner to allow for more equitable access to genomic medicine.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2209. Parents' attitudes towards return of results from genome sequencing research for their healthy children: A qualitative study

Authors:

C. Di Carlo¹, C. Mighton^{2,3}, M. Clausen², E. Joshi², S. Casalino², T. Kim¹, C. Kowal¹, C. Birken^{3,4}, J. Maguire¹, Y. Bombard^{1,3}; ¹Li Ka Shing Knowledge Inst., St. Michael's Hosp., Toronto, ON, Canada, ²Li Ka Shing Knowledge Inst., St. Michael's Hosp., Toronto, ON, Canada, ³Univ. of Toronto, Toronto, ON, Canada, ⁴The Hosp. for Sick Children, Toronto, ON, Canada

Abstract Body:

INTRO: With widespread genomic sequencing (GS) research efforts, there is increasing impetus to return individual results to research participants. Operationalizing return of results in a healthy paediatric population is complex, as parents face decision-making conflicts regarding the manner and to what capacity to receive their child's GS results. AIM: We explored parental attitudes and preferences towards their healthy children's participation in GS research, expectations for return of results, and the extent of involving their child in the process. METHODS: Parents with children enrolled in TARGet Kids! were recruited for participation. TARGet Kids! is a Canadian primary care research network where healthy children are followed longitudinally to obtain and return genomic data. Parents viewed a slide show on GS prior to semi-structured individual interviews exploring expectations for return and recontact of results, attitudes towards ownership of genetic information, and perceived utility of participation. Transcripts were coded by multiple members of the research team and analysed using thematic analysis and constant comparison. RESULTS: There were 26 parents interviewed; 20 were aged 30-39, 22 were female, 19 self-reported as White/European, and all were college or university educated. Parental perspectives of GS centred on three themes: (1) *Downstream impacts of testing*; Parents expressed concerns of genetic discrimination and unintentional differential treatment of their child, while acknowledging clinical utility of GS with provider support. (2) *Reciprocity*; Parents preferred to receive medically actionable, childhood onset results at minimum, with the expectation to be recontacted overtime with reclassifications. (3) *Power and empowerment*; Some parents felt empowered to learn and take preventative action for their child and relatives if presented with the opportunity. Other parents felt hesitant with the decision-making power on behalf of their child, expressing concerns regarding child autonomy. CONCLUSION: Parents focus on downstream impacts, reciprocity and power dynamics when considering their healthy child's participation in a GS research study. Incorporating these themes in patient-facing engagement materials can optimize research and promote participation centred on the needs and expectations of parents and caretakers.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2210*. Partnering with patients to explore the psychosocial and socioeconomic impacts of hereditary cancer syndromes

Authors:

H. Etchegary¹, J. Sam², M. Clausen², D. Bishop³, J. Pauling⁴, C. Pavao⁵, C. Remocker⁵, T. Tiano³, A. Tilley³, C. Mighton⁶, S. Rajeziessfahani¹, R. Gopalakrishnan², M. Aronson⁷, L. Dawson¹, A. Eisen⁸, T. Graham⁹, J. Green¹⁰, M. Krahn¹¹, S. Savas¹², N. Stjepanovic², S. Sun¹³, K. Thorpe⁶, K. Schrader¹³, Y. Bombard⁶; ¹Mem. Univ., St. John's, NL, Canada, ²St. Michael's Hosp., Toronto, ON, Canada, ³Patient Partner, St. John's, NL, Canada, ⁴Patient Partner, Toronto, ON, Canada, ⁵Patient Partner, Vancouver, BC, Canada, ⁶St. Michael's Hosp. & Univ. of Toronto, Toronto, ON, Canada, ⁷Zane Cohen Ctr., Sinai Hlth.System, Toronto, ON, Canada, ⁸Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, ⁹Univ. of Toronto & Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, ¹⁰Mem. Univ., St. John's, NL, Canada, ¹¹Univ. of Toronto & Univ. Hlth.Network, Toronto, ON, Canada, ¹²Mem. Univ., St. John's, ON, Canada, ¹³BC Cancer, Vancouver, BC, Canada

Abstract Body:

Background: Patient oriented research aims to improve patient outcomes and the quality of research by focusing on patient-identified priorities and engaging patients as partners. To date, methods for meaningful patient engagement and the role of patient partners in genomics health research are not well described. **Aim:** To describe the engagement of patient partners in a large team grant in Canada exploring the economic and psychosocial impacts of hereditary cancer syndromes (HCS). **Methods:** Six patient partners from 3 provinces with a hereditary breast and ovarian cancer syndrome or Lynch syndrome were recruited from team members' networks. Regular meetings among study PIs, staff and patient partners provide consistent communication and co-development opportunities. **Results:** Patient partners were invited to engage across all phases of the study: reviewing the grant application, co-developing the study design and materials, and ultimately assisting with recruitment, data analysis, and knowledge dissemination. Offering choice in level of involvement is best practice for patient engagement and has been appreciated by patient partners as affording them flexibility around their contributions. Patient partners meet quarterly with the study team, with email communication in between, to provide feedback on study design and materials. Opportunities to review study material on their own time allows for equitable involvement when attending meetings is not always feasible. As this study involves qualitative interviews and data collection about sensitive topics (e.g., cancer journey, emotional & financial impacts), patient partners' lived experiences have informed study materials. To date, patient partners provided feedback on the content and length of interviews, probing questions, and the language to be used. Feedback on early iterations of the interview guide revealed a bias towards negative language about the impact of HCS. Partners reminded the team that there were positive impacts as well and cautioned about the use of exclusively negative language (e.g., 'burdens'). Patient partners also took part in mock interviews to finalize the interview guide and provide a training opportunity for students and study staff. **Conclusions:** To date, patient partners provided important insights on the study population and methods. Their role reflects true research co-development. Ultimately, our aim is to build a rigorous patient oriented research program in hereditary cancers, informed by patient voices, from which lessons learned are shared and the care of families affected by HCS is improved.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2211. Patient reported utility of cancer results from genomic sequencing.

Authors:

S. Shickh^{1,2}, M. Clausen², C. Mighton^{1,2}, E. Adi-Wauran^{1,2}, D. Hirjikaka², R. Kodida², S. Krishnapillai^{2,1}, E. Reble², J. Sam², N. Baxter^{3,4,5}, A. Laupacis⁶, Y. Bombard^{2,1,7}; ¹Inst. of Hlth.Policy, Management and Evaluation, Univ. of Toronto, Toronto, ON, Canada, ²Genomics Hlth.Services & Policy Res. Program, Li Ka Shing Knowledge Inst. of St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, ³Univ. of Melbourne, Melbourne, Australia, ⁴Li Ka Shing Knowledge Inst., St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, ⁵Dept. of Surgery, Univ. of Toronto, Toronto, ON, Canada, ⁶Univ. of Toronto, Toronto, ON, Canada, ⁷Ontario Inst. for Cancer Res., Toronto, ON, Canada

Abstract Body:

Background Genomic sequencing (GS) holds great promise to improve clinical utility for patients with suspected hereditary cancer syndromes (HCS), most of whom remain undiagnosed after standard panel tests. Although GS may increase HCS detection, it also generates a large volume of non diagnostic results (e.g., variants of uncertain significance; VUS), whose impact on patients is largely unknown. This limits our understanding of all possible benefits and harms of GS and the evidence available for adoption and reimbursement decisions.

Aim Explore patient-reported utility of cancer GS results, including primary and secondary VUS and low/moderate risk results.

Methods A qualitative interpretive description study, using semi-structured interviews with patients who received GS for cancer as part of the Incidental Genomics Trial. Thematic analysis employing constant comparison was used. Two coders coded transcripts, with use of a third coder to resolve conflicts.

Results Twenty-five patients participated, most of whom were female (22), >50 years (18), European or Ashkenazi Jewish (17) with breast cancer (20), and who received a variety of results including: primary & secondary VUS and pathogenic low/moderate risk variants. Patients were initially disappointed or relieved about their results but ultimately focused on clinical utility, which they described as clear follow up plans, frequent surveillance and follow up by a specialist. Their focus on clinical utility was driven by the type of GS result received. For example, when patients were provided with clear clinical actions for low/moderate risk findings, they perceived the results to be very “useful” and “reassuring” and of moderate-high utility. In contrast, when receiving low/moderate risk results or primary VUS without clear clinical follow up, patients perceived results as “concerning”, leading to potential harms such as anxiety or vigilance about cancer symptoms. Lastly, patients receiving secondary VUS interpreted their results as “negative” and not requiring any additional clinical follow up, and therefore they began to refocus on other aspects of their life and “move forward”. Overall, having a supportive family or primary care provider enhanced perceptions of utility while unsupportive families or providers diminished utility.

Conclusion From patients’ perspectives, GS for cancer may add limited utility and potentially cause harms. These findings of preliminary experiences of harms and limited utility warrant further evaluations in addition to practice interventions to support patients receiving low/moderate risk results and primary VUS from GS.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2212. Patterns and barriers to clinical trial participation for patients with primary mitochondrial disease: A survey of patients/caregivers, physicians, and pharmaceutical industry professionals.

Authors:

M. Windrem¹, P. E. Yeske², J. M. Ridgway³, A. Karaa⁴, **M. DiMatteo**¹; ¹Stealth BioTherapeutics, Needham, MA, ²United Mitochondrial Disease Fndn., Pittsburgh, PA, ³Bruno and Ridgway Res. Associates, Lawrenceville, NJ, ⁴Genetics Div., Massachusetts Gen. Hosp./Harvard Med. Sch., Boston, MA

Abstract Body:

Objective: Clinical trial recruitment is challenging, especially for rare, genetic disorders. Clinical research fails to attract the necessary level of clinical trial participation due to multiple factors. We conducted a study to better understand reduced participation in clinical trials among mitochondrial disease patients/caregivers, physicians, and industry professionals. **Methods:** Mitochondrial disease patients or caregivers, physicians treating patients with mitochondrial disease and being involved as an investigator in at least one clinical trial, and pharmaceutical industry professionals were invited via email to participate in a survey. The surveys were conducted from May 4-16, 2022. Interview topics for each group included discussion regarding clinical trial participation and recruitment as well as perceived barriers to participation in clinical trials. Results were compared between the groups. **Results:** A total of 207 surveys were completed (n=170 patients/caregivers, n=20 physicians, n=17 industry professionals). At least one clinical trial was recommended by a physician to 39% of patients surveyed, primarily being recommended by a neurologist (46%) or a geneticist (27%). Of patients surveyed, 29% participated in a clinical trial. Most physicians (85%) reported frequently recommending clinical trials for their patients. Physicians reported that most patients were either very (50%) or somewhat (35%) receptive to clinical trials. Industry professionals reported that most patients are very (23%) or somewhat receptive (59%) to participating in clinical trials. Various reasons exist for the different levels of perceived receptiveness. Travel/transportation was a common theme across all three groups as a leading barrier to patients participating in clinical trials. Other leading barriers agreed upon by more than one group included: trial site availability; not knowing how to find clinical trials; no time to participate; and not likely to qualify. **Conclusions:** Mitochondrial disease patients/caregivers, physicians, and pharmaceutical industry professionals indicate similar barriers to clinical trial participation. Increased understanding of such barriers may assist in improving clinical trial awareness and enrollment for patients with mitochondrial diseases.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2213. Personal utility in genomic medicine: what is it and can we measure it?

Authors:

E. Turbitt¹, J. Kohler², F. Angelo³, I. Miller⁴, K. Lewis⁵, K. Goddard⁶, B. Wilfond⁷, B. Biesecker⁸, M. Leo⁹; ¹Univ. of Technology Sydney, Ultimo, Australia, ²Stanford Med., Palo Alto, CA, ³Northwestern Univ., Seattle, WA, ⁴Children's Natl., Washington DC, DC, ⁵NIH, Bethesda, MD, ⁶Natl. Cancer Inst., Bethesda, MD, ⁷Seattle Children's Res. Inst., Seattle, WA, ⁸RTI Intl., Bethesda, MD, ⁹Kaiser Permanente Ctr. for Hlth.Res., Portland, OR

Abstract Body:

Rationale: It is important to understand the benefits of a medical test, as defined by those undergoing testing and receiving results. Genomic test results do not always provide information that is medically actionable. Yet, end users report benefits from receiving genomic information even if medical actions are not indicated. The ability to assess this personal utility in genomic medicine relies on a validated, patient-defined outcome measure, which is currently lacking. We aimed to construct and psychometrically evaluate a measure of personal utility - the PrU.

Methods: We used an evidence-based, operational definition of personal utility, with data from a systematic literature review and Delphi survey. This definition formed the basis from which we developed 24 items for the PrU. We piloted the PrU with 24 adults enrolled in a genome sequencing research study. The results from this pilot along with expert recommendations led to a revised PrU consisting of 17 items. We administered the PrU to adults enrolled in one of two Clinical Sequencing Evidence-Generating Research (CSER) studies after receiving genomic results ($n=841$).

Results: Most respondents were female (77.2%), educated to some post-high school or beyond (73.6%), and identified as White (37.7%), Black (28.5%), and/or Hispanic/Latino (21%). A principal-axis factor analysis suggested a three-factor solution, accounting for 66% of the variance in the items. Using the items as a guide, we labeled the three factors "self-knowledge and control" ($\alpha=.94$), "reproductive planning" ($\alpha=.89$), and "practical benefits" ($\alpha=.87$).

Conclusions: We provide strong initial evidence for the reliability and validity of three subscales of personal utility valued by adults who receive genomic test results that converge on the broader central concept of personal utility. The PrU is designed to measure the personal utility people report from receiving genomic information. Further validation of the scale in other contexts is required. The ability to measure the dimensions of utility in genomic medicine contributes to understanding patient experiences and guiding the implementation of genomics in clinical care.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2214. Population DNA screening for medically actionable conditions in young adults: The DNA Screen national pilot study

Authors:

P. Lacaze¹, J. Tiller¹, I. Winship², The DNA Screen Investigator Group; ¹Monash Univ., Melbourne, Australia, ²Royal Melbourne Hosp., Melbourne, Australia

Abstract Body:

Population DNA screening for medically-actionable conditions in early adulthood has the potential to enable timely prevention, early detection and treatment of cancer and heart disease. For medically-actionable genomic conditions such as hereditary breast and ovarian cancer (HBOC), Lynch syndrome (LS) and familial hypercholesterolemia (FH), pathogenic germline DNA variants confer high risk of developing future disease, and can be used to prompt risk management and family cascade testing. Population DNA screening for these conditions could thereby identify medically actionable genetic risk factors early, enabling timely risk-management and informed decision making, to facilitate early detection or prevention, and realise the preventative potential of genomics. Despite this opportunity, diagnostic rates for these conditions remain low, limited by restricted access to genetic testing and lack of awareness. Offering preventive DNA screening to healthy adults through a national healthcare system, with minimal or even no testing criteria using a low-cost DNA screening test, has the potential to save thousands of lives and achieve healthcare savings - if it can be delivered responsibly, cost-effectively and at scale. In Australia, a national pilot study of 10,000 adults aged 18-40 years has been funded by the Australian Government's Genomic Health Futures Mission, and begins recruitment in mid-2022. The DNA Screen national pilot study has been designed specifically to answer key implementation questions and address socio-ethical challenges associated with population DNA screening. This presentation will cover various aspects of the DNA Screen study, including: the purpose-built prospective study design; social-media driven recruitment strategy; scalable DNA screening test and assay; telehealth genetic counselling and return of results process; downstream clinical support pathways; policy, implementation and cost-effectiveness aspects; and integration into the national healthcare system. The DNA Screen pilot study is positioning Australia to take a world-first step towards offering preventive DNA screening through a national healthcare system.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2215. Predictors and Trends in Cancer Genetics Clinic Attendance Rate After the Adaptation of Telemedicine During the COVID-19 Pandemic

Authors:

A. Smullin, P. Flodman, D. Nathan, M. Smith; UC Irvine, Irvine, CA

Abstract Body:

Despite the clear benefit of cancer genetic counseling, many eligible patients never meet with a cancer genetic counselor. Several elements contribute to this, including the growing demand for genetic services, lack of genetic professionals, and patient non-attendance. Prior research in cancer genetic counseling and other medical specialties has investigated the use and outcomes of alternate service delivery models, however, little is known about the specific impact of telemedicine on patient attendance over a substantial period of time. This study analyzed demographic and clinical data from 800 adult patients seen for cancer genetic counseling before and after the adaptation of telemedicine during the global COVID-19 pandemic. The purpose of this research was to investigate telemedicine's impact on attendance rate at follow-up appointments as well as explore patient predictors of attendance status. Logistic regression analyses identified that patients were 3.54 times more likely to attend their first scheduled follow-up visit if they were in the telemedicine cohort ($p < 0.001$). Additionally, patients who had more relatives with cancer and patients of Asian descent were more likely to attend their first follow-up visit. Patients were less likely to attend their first scheduled follow-up visit if there was a greater amount of time between their initial appointment and their genetic test results report date. This research builds upon current literature on attendance status and contributes novel findings on the scope and impact of telemedicine's role in increasing attendance and access to cancer genetic counselors. Recognizing and understanding telemedicine's positive outcomes may lay the foundation for the adoption and permanence of this service delivery model in the cancer genetic counseling setting.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2216. Preferences of parents from diverse backgrounds on genomic screening in newborns.

Authors:

J. Omorodion^{1,2,3}, C. Hsu⁴, G. Rivera-Cruz^{1,5}, M. del Rosario⁵, S. Pereira⁶, S. Ziniel⁷, I. Holm^{5,8}, N. Gold^{8,9}; ¹Harvard Med. Sch. Genetics Training Program, Harvard Med. Sch., Boston, MA, ²Boston Combined Residency Program, Dept. of Pediatrics, Harvard Med. Sch., Boston, MA, ³Boston Combined Residency Program, Dept. of Pediatrics, Boston Univ. Sch. of Med., Boston, MA, ⁴Ctr. for Bioethics, Harvard Med. Sch., Boston, MA, ⁵Div. of Genetics and Genomics, Boston Children's Hosp., Boston, MA, ⁶Ctr. for Med. Ethics and Hlth. Policy, Baylor Coll. of Med., Houston, TX, ⁷Dept. of Pediatrics, Univ. of Colorado Sch. of Med., Aurora, CO, ⁸Dept. of Pediatrics, Harvard Med. Sch., Boston, MA, ⁹Div. of Med. Genetics and Metabolism, Massachusetts Gen. Hosp. for Children, Boston, MA

Abstract Body:

Newborn screening (NBS) is a long-running and successful public health initiative which has improved morbidity and mortality in infants with a range of rare disorders. The current Recommended Uniform Screening Panel specifies 35 core conditions that are suggested for inclusion in each state's NBS program. As genomic testing modalities and therapies for genetic diseases have expanded, a larger scope of disorders have been considered for NBS. Parents are key stakeholders in expanding NBS. Previous studies that have examined parental attitudes towards genomic sequencing have been largely limited to white, highly-educated populations. If genomic sequencing is to be implemented in an ethical and just manner, however, it is crucial that the perspectives of parents from a broad range of backgrounds are included. We conducted a series of semi-structured interviews and focus groups with parents from varying racial, ethnic, and socioeconomic backgrounds, which focused on several themes that may be relevant to future policymaking in NBS, including preferences regarding which genetic disorders to screen, coping with the uncertainty of genetic test results, data privacy, and barriers to pursuing follow-up care. Overall, 20 interviews (15 in English, 5 in Spanish) and 5 focus groups (4 in English, 1 in Spanish) were conducted. Emerging themes have included enthusiasm for expanded newborn screening, an appreciation for the nuances of variants of uncertainty and low penetrance disorders, and interest in shared ownership of the genetic data with the child's pediatrician. We are hopeful that by eliciting parental attitudes towards expanded newborn screening from a diverse population, and then using this data to inform its implementation, there will be maximal participation and benefit to children and families.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2217*. Priorities to promote participant engagement in cancer genomics research: A report from the Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network

Authors:

A. Schuster¹, N. Crossnohere¹, M. Bachini², C. K. Blair³, J. Carpten⁴, E. Claus⁵, G. Colditz⁶, L. Ding⁶, B. Drake⁶, R. Fields⁶, K. Janeway^{7,8}, B. M. Kwan⁹, H-J. Lenz⁴, Q. Ma¹, S. I. Mishra³, E. D. Paskett¹, T. R. Rebbeck^{10,8,11}, C. Ricker⁴, M. C. Stern⁴, A. Sussman³, J. Tiner¹², J. Trent¹³, R. Verhaak¹⁴, N. Wagle^{15,8,16}, C. Willman^{17,18}, J. F. P. Bridges¹; ¹The Ohio State Univ., Columbus, OH, ²Cholangiocarcinoma Fndn., Herriman, UT, ³Univ. of New Mexico Hlth.Sci. Ctr., Albuquerque, NM, ⁴Univ. of Southern California, Los Angeles, CA, ⁵Yale Univ, New Haven, CT, ⁶Washington Univ. Sch. of Med., St. Louis, MO, ⁷Dana-Farber / Boston Children's Cancer and Blood Disorders Ctr., Boston, MA, ⁸Broad Inst. of MIT and Harvard, Boston, MA, ⁹Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ¹⁰Harvard TH Chan Sch. of Publ. Hlth., Boston, MA, ¹¹Dana-Farber Cancer Inst., Boston, MA, ¹²Natl. Cancer Inst., Bethesda, MD, ¹³TGen, Phoenix, AZ, ¹⁴The Jackson Lab., Farmington, CT, ¹⁵Harvard Med. Sch., Dana-Farber / Harvard Cancer Ctr., Dana-Farber Cancer Inst., Boston, MA, ¹⁶Count Me In, Boston, MA, ¹⁷Mayo Clinic Comprehensive Cancer Ctr., Rochester, MN, ¹⁸Univ. of New Mexico Sch. of Med. and Comprehensive Cancer Ctr., Albuquerque, NM

Abstract Body:

Background: Engaging diverse populations in cancer genomics research is of critical importance and is a fundamental goal of the National Cancer Institute (NCI) Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network. The Network sought to identify research priorities in participant engagement for the Network and strategies to address these priorities, and where possible bring medical benefit. **Methods:** The Network conducted a four-phase process with affiliated stakeholders to identify research priorities and associated strategies. The process comprised: (1) an online survey to elicit potential priorities; (2) a two-day virtual meeting, including participants and advocates from each Research Center to discuss and vote on priorities; (3) recommendations from PE-CGS External Advisory Panel to refine priorities; and (4) a one-hour virtual meeting to rank priorities and identify strategies. **Results:** Nearly 150 Network-affiliated stakeholders were engaged in the process, including clinicians, scientists, genetic counselors, funders, and patient/community representatives and cancer advocates. Six priorities in order of importance were: (i) tailor education and communication throughout the research process; (ii) identify measures of engagement; (iii) identify optimal engagement strategies; (iv) understand cancer disparities in the context of cancer genomics research; (v) personalize the return of genomics findings; and (vi) consider ethical, legal, and social implications of genomics research and participant engagement. Six associated strategies to support meaningful coordination and cooperation were: leverage Network infrastructure; develop shared models of knowledge; establish subcommittees; hold Network-wide “All Hands” meetings; securely share data and materials; and standardize and harmonize data. **Discussion:** The Network is actively pursuing these research priorities by implementing the associated strategies that support meaningful coordination and cooperation to bring direct benefit to our patients. These efforts harness the collective strengths of all Network stakeholders and position the Network to generate discoveries that advance the science of engagement in cancer genomics research.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2218*. Prioritizing Research in Newborn Screening: Tools from the Newborn Screening Translational Research Network

Authors:

J. Taylor, K. Chan, Y. Unnikumaran, G. Tona, C. Lorkovic, M. Lietsch, A. Brower; American Coll. of Med. Genetics and Genomics, Bethesda, MD

Abstract Body:

As advances in science and technology provide treatment options for genetic conditions, the number of diseases that could be included in newborn screening (NBS) continues to grow. A variety of stakeholders have led efforts to develop the evidence to expand NBS. A federal advisory committee reviews conditions that are nominated and evaluate the evidence of feasibility to screen and a net benefit of early identification and intervention. The increasing use of genomics and gene-targeted therapies has led to a large pipeline of conditions that are candidates for newborn screening. There are several efforts such as the Clinical Genome Resource (ClinGen) that curate the clinical relevance of genes and variants that are clinically actionable through the use of data standards and evidence frameworks. Using these efforts as a model, we developed a process for evaluating conditions that are candidates for newborn screening. We describe the process for identifying conditions, the evidence gathered to evaluate conditions and the process for prioritizing conditions that have a high potential to be nominated to the recommended uniform screening panel (RUSP). An initial list of candidate conditions was developed with a questionnaire sent to subject matter experts that included the name of the condition, analytical method, second-tier test, and available treatment(s). In 2021 NBSTRN updated the approach in consultation with the Pilot Research & Implementation Workgroup to include additional sources for identifying candidate conditions. The information about each condition was expanded to include information about research and pilot studies, incidence rates, and advocacy groups. The list was then prioritized by including criteria such as understanding of the condition (severity/urgency) test efficacy, and treatment efficacy. The new method has identified 71 conditions that are candidates for newborn screening, 21 conditions are currently a part of NBS state panels, 19 conditions may be detected in the differential diagnosis of current RUSP conditions, and 7 conditions have been nominated for addition to the RUSP. NBSTRN will continue to evaluate the evidence for NBS conditions and propose strategies to evaluate changes and improvements to NBS expansion. In conclusion, NBSTRN created a process to help researchers identify diseases for the discovery of screening, diagnostic, and therapeutic advancements. Information about these conditions can be found in the Newborn Screening Conditions Resource (NBS-CR), an interactive tool that is available online at <https://nbstrn.org/tools/>.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2219. Psychosocial impacts in patients/parents with secondary findings reporting from exome sequencing: 1st French multicenter qualitative and quantitative study (FIND Study).

Authors:

E. Viora-Dupont^{1,2}, F. Robert^{3,4}, C. Binquet^{1,5}, A. Chassagne^{1,6}, S. Staraci^{4,7}, M. Rossi^{3,8}, D. Sanlaville^{3,8}, P. Ederly^{3,8}, G. Lesca³, A. PUTOUX³, N. Chatron³, L. Pons³, A. Cadenes³, E. Gautier², S. Moutton², A. Sorlin², C. Thauvin-Robinet^{1,2}, C. Philippe^{1,2}, A. Baurand², C. Sawka², C. Mignot⁹, P. Charles⁹, A. Afenjar⁹, I. Marey⁹, D. Héron⁹, B. Keren⁹, M. Spetchian⁹, M. Gargiulo^{4,10}, Consortium FIND, L. Faivre^{1,2}; ¹FHU TRANSLAD, GAD INSERM UMR 1231, Univ. of Burgundy Franche-Comté, Dijon, France, ²Genetics Dept., Reference Ctr. for Dev.al Disorders, Univ. Hosp., Dijon, France, ³Genetics Dept., Reference Ctr. for Dev.al Disorders, HCL, Bron, France, ⁴Clinical Psychology Lab., Psychopathology, Psychoanalysis (EA4056, ED 261), Univ. of Paris, Sorbonne Paris City, Paris, France, ⁵CIC 1432 Module Épidémiologie Clinique, Université de Bourgogne, Dijon, France, ⁶Lab. of Sociology and Anthropology (LaSA, EA3189), Univ. of Burgundy Franche-Comté, Besançon, France, ⁷Genetics Dept., Reference Ctr. for Hereditary Cardiac Disorders, GH APHP, Paris, France, ⁸INSERM U1028, CNRS UMR5292, CRNL, GENDEV Team, Univ. of Claude Bernard Lyon 1, Bron, France, ⁹Genetics Dept., Reference Ctr. for Dev.al Disorders, GH APHP, Paris, France, ¹⁰Inst. of myology, GH APHP, Paris, France

Abstract Body:

The development of next generation sequencing improves the primary diagnosis but can reveal additional information (incidental findings or secondary findings (SF)). The search for secondary findings raises debates among geneticists. Faced with the contradictions in the recommendations of American and European learned societies, a national study has been set up to contribute to the evolution of French recommendations.

Our study was deployed during 18 months in 3 French reference centers for developmental diseases. Patients/parents who have undergone exome sequencing have been asked their choice to access to SF on themselves or their children for 3 categories (group 1 late onset actionable diseases, group 2 genetic counselling, group 3 pharmacogenetics)

Using a mixed, qualitative and quantitative, method, we included 344 patients or parents of children with undiagnosed developmental disorders. We studied the initial perception, the repercussion of the results at the time of the return, at 6 months and 12 months after the return via questionnaires designed for our study, and standardized scores of anxiety and depression completed. Furthermore, semi-directed interviews allowed us to deepen the impact of the results.

80% of the parents accepted the search for SF in their child, and we showed an influence of the medical discourse on the parents' choices. 52 SF were returned to 47 families, 44 families completed the questionnaires at the time of return and 38 and 39 at 6 months and 12 months. 26 families were interviewed. We show that the return of the SF has no impact on the parameters studied but the semi-directed interviews showed that the individual situation could be very different, and distress can exist. The follow-up interviews up to one year did not show regrets, but it was not without any psychological impacts in group 1.

The degree of anxiety and anguish is correlated with the nature of the result and the life context of the subjects concerned. This study highlights the importance of an interdisciplinary team to accompany this type of opportunistic presymptomatic diagnosis, the effects of which are difficult for parents to anticipate, especially since their main motivation remains the search for an etiological diagnosis of their child's disorders and potential actionability.

ASHG 2022 Annual Meeting Poster Abstracts

Our study confirms the limited impact of secondary findings reporting on patients' anxiety and quality of life.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2220. Q-A sequences in prenatal genetic counseling in Japan; conversation analytic study of NIPT and NT consultations

Authors:

M. Kawashima¹, H. Maeda², K. Horike³, A. Hayashi³, A. Tachibana³, M. Morine³, K. Hinokio³, T. Iwai⁴, K. Maeda⁴, N. Okamoto⁵, A. Kondo⁴; ¹Kyoto Sangyo Univ., Kyoto, Japan, ²Rikkyo Univ., Tokyo, Japan, ³Prenatal Med. Ctr., Shikoku Med. Ctr. for Children and Adults, NHO, Zentsuji, Japan, ⁴Med. Genetics Ctr., Shikoku Med. Ctr. for Children and Adults, NHO, Zentsuji, Japan, ⁵Osaka Women s and Children s Hosp., Izumi, Japan

Abstract Body:

Genetic counseling should be precisely tailored based on the clients' perspectives and understanding, especially when clients face serious and important decision-making. Genetic counselors are tasked with providing ample information and support for clients especially in cases like nuchal translucency (NT). This study focuses on question-answer (Q-A) sequences in prenatal genetic counseling by comparing noninvasive prenatal genetic testing (NIPT) and NT cases. Previous studies have examined linguistic strategies for establishing the supportive and non-coercive stance of genetic counselors, which enhance clients' autonomy in counseling (Sarangi & Clark 2002; Sarangi 2000). This study resides in the similar stream using conversation analysis to elucidate details of interactions based on 12 video-recordings. The recordings took place in a tertiary hospital in Japan and all genetic counseling was done by registered clinical geneticists (CGs) at this hospital. General medical encounters are quite structured with an opening/problem, presentation/history, taking/physical exam/diagnosis/treatment, and recommendation/closing (Heritage & Maynard 2006). However, we found that prenatal counseling is structured through Q-A sequences to fit with the client's knowledge and perspective. There are three major findings. First, both NIPT and NT cases involved Q-A sequences before information giving to gather clients' previous knowledge and stance toward screening tests. In NIPT cases, the CG starts with inquiries on intention for testing and previous understandings about NIPT. In NT cases, the CG starts with unpacking results from ultrasounds and asks questions on clients' perspectives and understandings. Second, especially in NT cases, both information giving and decision-making activities are sequentially initiated by clients' questions. During the CG's explanations, clients frequently ask questions on certain parts. The CG uses that occasion to include additional information, which are tailored to the client's stance and knowledge level. Third, the clients' questions are initially concerned on the given information. Clients then move to focus on decision-making autonomously in the later part. The CG can advance the decision-making process by including recommendations within the slot of answers regarding the screening test as well as based on the attitude and stance that clients can take to make a decision in the future. Through the Q-A sequences, the genetic counseling providers and clients are engaged in establishing relationships trustworthy enough to reach a decision about screening tests in prenatal genetic counseling.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2221. Raising Genetic Disease Awareness: NBSTRN Efforts to Leverage Social Media to Realize the Promise of Genomics

Authors:

Y. Unnikumar¹, K. Chan², R. Wiebenga³, G. Tona⁴, E. Wilson¹, H. Weiss¹, M. Hartnett¹, A. Brower⁵; ¹American Coll. of Genetics and Genomics, Silver Spring, MD, ²American Coll. of Med. Genetics and Genomics, Bethesda, MD, ³Newborn Screening Translational Res. Network, DELL RAPIDS, SD, ⁴NBSTRN, Bethesda, MD, ⁵American Coll. of Med. Genetics and Genomics, Dakota Dunes, SD

Abstract Body:

Social media has transformed genetics and genomics research by facilitating the global exchange of ideas among individuals impacted by disease with the researchers. This rapid transmission of information has the potential to dramatically raise awareness of newborn screening (NBS) and empower advocates to work with researchers to realize the promise of genomics to screen, diagnose, and treat babies. To realize the promise of genomics and advance NBS research, a diverse network of researchers, health professionals, families and advocacy groups, and state NBS programs need to have a mechanism to connect, collaborate and communicate the most relevant and cutting-edge research and knowledge. The Newborn Screening Translational Research Network (NBSTRN) has facilitated this new partnership, collaboration, and capacity building through the use of social media. The NBSTRN created a four-stage process of disseminating research efforts and increasing genetic disease awareness: 1) outreach, 2) engagement, 3) membership, and 4) awareness. In the Stage 1 (Outreach), NBSTRN uses website, social media, podcast, and publications; in stage 2 (Engagement), NBSTRN uses experts in their workgroup and the co-creation of data tools and resources, in Stage 3 (Membership), NBSTRN uses a forum platform to engage with stakeholders, and in Stage 4 (Awareness), NBSTRN provided specialized consultation services for grant applications, publication and study design. As of June 2022, the NBSTRN represents over 194 institutions from around the world, including 51 academic institutions, 45 for-profit institutions, 6 governmental agencies, 34 international institutions, 38 state programs, and 29 health care systems. In less than a year since conception, the Newborn Screening SPOTlight podcast has reached over 38 countries with over 1,100 episodes downloads. We will highlight the impact of NBSTRN social media awareness campaigns have made on severe combined immunodeficiency (SCID) and spinal muscular atrophy (SMA). This poster will share strategic approach using the four-stage process to engage with different stakeholders, lessons learned and unlearned in engagement activities on the different social media channels, and impact made on raising awareness of rare disease conditions. These efforts serve as an important model of how research initiatives can leverage social media to advance genomics and improve the health of all individuals with genetic disease.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2222. Real-world implementation of an efficient germline Point-of-Care Testing model in a multidisciplinary cancer center.

Authors:

K. Schaffer^{1,2}, E. Watkins³, P. Hurley^{1,2}, K. Moses⁴, M. Donahue⁴, K. E. Beckermann^{1,2}, N. Davis^{1,2}, K. K. Ancell^{1,2}, B. Park^{1,2}, T. Pal^{1,2}, K. F. Mittendorf¹, K. Lang^{1,2}, G. L. Wiesner^{1,2}; ¹Vanderbilt-Ingram Cancer Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Vanderbilt Univ. Sch. of Med., Nashville, TN, ⁴Dept. of Urology, Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Background: National guidelines recommend germline genetic testing for many cancers given the potential for precision therapy. Traditionally, patients see a genetic counselor (GC) for counseling and testing. NSGC has reported that a full-time cancer GC provides pre-test counseling for 10 new patients per week. Thus, limited GC availability and increasing volume of patients can delay testing and precision therapies. To address this barrier, the Vanderbilt-Ingram Cancer Center developed a novel genetic counseling and point-of-care testing (POCT) model to streamline the process. We describe a POCT pilot for patients with genitourinary (GU) cancers.

Methods: In conjunction with the Vanderbilt Hereditary Cancer Clinic (HCC), POCT protocols were developed outlining germline testing processes in medical oncology (MO) and urologic oncology (UO) clinics including patient identification, education, orders, and lab processing. Samples were sent to a national reference lab for hereditary cancer panel testing. Oversight of the program was done by a HCC GC and oncology clinicians from MO and UO. These champions were responsible for implementation, ensuring appropriate follow-up, and developing a POCT database to track results. Patients were provided a previously validated online interactive educational video focused on inherited cancer predisposition and following receipt of test results, were offered HCC referral to discuss results.

Results: POCT germline testing was completed on 356 patients with GU malignancies from April 2020 to April 2022. The majority (336; 94%) had prostate cancer, 14 (4%) had urothelial carcinoma or renal cell carcinoma, and 6 (2%) had >1 GU malignancy. 319 (98%) were male, mean age was 66 (37 - 90), 305 (86%) were White, 43 (12%) Black, 8 (2%) Other; 337 (95%) Non-Hispanic. MO completed testing for 161 (45%), UO for 115 (32%), and a multidisciplinary team for 80 (22%). Pathogenic or Likely Pathogenic variants were identified in 39 patients (11%); of these 22 (56%) had therapeutic implications and 17 (44%) accepted post-test counseling. Compared to the NSCG report, the completed tests are equivalent to 74% of a GC annual workload.

Conclusions: This pilot demonstrates the feasibility of a streamlined, multidisciplinary germline POCT model. The essential elements include designated clinician champions, coordination of POCT with other visits, and online patient education tools. The POCT model dramatically improved access to germline testing for patients undergoing cancer care, allowed patients with eligible variants to receive precision therapies, and identified those needing additional screening or cascade testing for family members.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2223. Rethinking general consent for stem cell-based embryo model research

Authors:

B. Mittleman; Stanford Univ., Stanford, CA

Abstract Body:

Advances in stem cell research now allow us to create embryo models outside the womb from stem cells originally derived from human skin. The models provide novel opportunities to understand human development, fertility, and genetics, yet bring potential ethical and legal challenges. Current ethical and legal work centers on the moral status of embryo models but no one has evaluated the complexity the new research adds to the relationship between researchers and biological material donors or between donors and the research. I argue the use of previously collected biological material for stem-cell based embryo models open novel ethical and legal concerns. Broad consent models take autonomy away from research participants in a morally controversial area of research. Current common law property doctrine would not apply easily to a dispute over stem-cell based embryo models and current statutes do not protect aggrieved research participants. This work outlines an inform consent structure for stem-cell based embryo model research that considers both research participant autonomy and possibilities for future innovation. I advocate for researchers to apply dynamic consent in their research by collecting new biological material and investing in continued relationships with donors.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2224. Return of secondary genomic findings: Experiences of sickle cell disease research participants.

Authors:

J. Floyd¹, A. Buscetta¹, G-A. Fasaye², V. Bonham¹; ¹Natl. Human Genome Res. Inst., Bethesda, MD, ²Natl. Cancer Inst., Silver Spring, MD

Abstract Body:

While the body of research investigating research participants' views on the return of actionable secondary genomic findings results (RoR) grows, there has been limited study of individuals who are living with a chronic genetic condition, such as sickle cell disease (SCD). People with SCD have historically faced discrimination and difficulty accessing treatments for their condition, which can potentially be compounded when seeking additional healthcare downstream due to RoR. As RoR is increasingly incorporated into studies, it is imperative that the views of diverse research participants be investigated and rooted in the context of health equity. This qualitative study of adult SCD research participants (n=30) was conducted in 2021-2022 with the aim of exploring views and attitudes regarding RoR. Each participant completed a semi-structured interview to ascertain perceived benefits, harms, and other implications associated with RoR. Results suggest that adults living with SCD support the return of secondary findings (SF) from genomic sequencing studies. Study findings adopt the view that already living with a chronic genetic condition may influence potential psychological burden of receipt of SF. Themes from the data include the expectation that researchers should provide referrals to medical and mental health care specialists, downstream financial support for low-resource study participants, and support with communicating results with family. Understanding the views of research participants based upon race, ethnicity, genetic condition, insurance coverage and other sociodemographic factors is important in the developing body of research on return of secondary genomic findings.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2225. Secondary findings in a large Pakistani cohort tested with whole genome sequencing

Authors:

A. Skrahin¹, H. A. Cheema², M. Hussain³, N. N. Rana⁴, K. U. Rehman⁵, R. Kumar⁶, G. Oprea¹, N. Ameziane¹, A. Rolfs¹, V. Skrahina¹; ¹Arcensus GmbH, Rostock, Germany, ²The Children Hosp., Lahore, Pakistan, ³Pakistan Inst. of Med. Sci., Islamabad, Pakistan, ⁴Children Hosp. and ICH, Multan, Pakistan, ⁵Town Women and Children Hosp., Peshawar, Pakistan, ⁶Liaquat Natl. Hosp., Karachi, Pakistan

Abstract Body:

Background. Early detection of genomic secondary findings (SFs) leads to the timely initiation of therapeutic or preventive measures. The studies in the field of SFs are diverse regarding participants' characteristics (country, ethnicity, presence of symptoms, and consanguinity); sequencing methods: whole genome sequencing (WGS) or whole exome sequencing (WES); versions of the American College of Medical Genetics and Genomics (ACMG) SFs list (v1.0, 2013; v2.0, 2016; v3.0, 2021); inclusion medically actionable SFs not yet recommended for reporting by the ACMG.

Aim and methods. Taking advantage of WGS over WES and version 3.0 ACMG SF (SFs v3.0) list over previous versions we studied SFs in individuals clinically suspected of genetic diseases at five different hospitals in Pakistan. In addition to the SFs v3.0, we have defined a list of gene/disease pairs that are not included in the SFs v3.0 list but have a clear medically actionable value if detected and reported (non-ACMG SFs), that includes e.g., Glucose-6-phosphate dehydrogenase deficient hemolytic anemia (*G6PD* gene) and Factor V Leiden thrombophilia (*F5* gene).

Results. We included 744 individuals; 91% of the participants were clinically symptomatic (n=674); 90% younger than 18 years (n = 666); 81% of the participants were reported as product of consanguinity (n = 606). The SFs v3.0 were detected in 21 (2.8%) individuals. Most of the SFs v3.0 belonged to cardiovascular diseases (CVD), 81.8% and cancer predisposition conditions, 9.1%. Non-ACMG SFs were detected in 55 (7.4%) participants. Most of the non-ACMG SFs belonged to hemolytic anemias (35.1%), followed by CVD (17.5%) and bleeding disorders (8.8%); cancer predisposition conditions accounted for 1.8% of non-ACMG SFs.

Conclusions. WGS is a reliable and easy test format to identify SFs v3.0 and non-ACMG SFs. The overall rate of SFs in the Pakistani individuals tested for genetic diseases is unexpectedly high (10.2%) due to a very high proportion of non-AGMG SFs (7.4%). The most frequent SFs were in genes associated with CVD - 81.8% (SFs v3.0), and hemolytic anemias - 35.1% (non-ACMG SFs). Our findings serve as a resource to inform decision-making in individuals undergoing genomic testing, aid the development of practice standards in genomic medicine, and drive future research efforts.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2226. Shared in Rare: Engaging stakeholders to develop a shared ELSI research agenda across rare diseases.

Authors:

C. Berrios¹, J. Garrett¹, L. Jones², N. Petersen², M. Strenk¹, Rare Voices Advisory Group; ¹Children's Mercy Kansas City, Kansas City, MO, ²Rare Voices Advisory Group, Kansas City, MO

Abstract Body:

Research on individual rare diseases presents challenges such as small sample sizes and limited funding sources, while research efforts spanning across rare diseases present difficulty in applying outcome measures to conditions with variable manifestations. Common experiences in rare disease (extended diagnostic odyssey, chronic conditions, need for advocacy, and challenges finding and managing multi-specialty care) create shared ethical, legal, and social issues (ELSI). Shared in Rare is a stakeholder engagement project to leverage the shared experiences of patients and families impacted by rare disease to develop an ELSI research agenda that is both applicable across rare diseases and guided by patient and family priorities.

A stakeholder group, Rare Voices (RV), is collaborating with ELSI and community engaged researchers to complete the project. RV includes 4 teens and 12 parents of children impacted by rare disease and 6 clinicians/researchers. This project team is working together to conduct listening sessions with diverse parents, caregivers, and teens in the rare disease community that will inform development of the ELSI research agenda.

Early engagement with RV members revealed anticipated challenges including identifying research topics across all rare diseases, communication across diverse stakeholders, moving from qualitative listening session data to research questions, and addressing stakeholder feelings about traumatic experiences surfaced by the project work. This input informed modifications to the project plan including participatory RV training in qualitative research utilizing interviews with RV parent members and collaborative development of research questions emerging from those interviews. Frequent small group work has been incorporated to facilitate communication and engagement. A trauma workgroup was formed and completed training to integrate trauma sensitive practices into all project activities.

This incorporation of stakeholder input has built a highly collaborative and engaged project team working equitably to reach the project aims. 17/22 (77%) of RV members are directly involved in the listening sessions as part of teams developing recruitment materials, drafting the discussion guide, and moderating listening sessions. The team's actions to address anticipated challenges inform future stakeholder engagement and the collaboration in every aspect of this project has poised the listening sessions to capture the voice of the rare disease community, facilitating identification of research needs that will inform a patient and family-centered ELSI research agenda to improve care for rare diseases.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2227. The future of diversity, equity, and inclusion: Achieving the vision of genetic counseling for the All of Us Research Program

Authors:

K. Onyeneho; All of Us Res. Program, NIH, Bethesda, MD

Abstract Body:

Objectives: Genetic counseling is germane to helping individuals learn about genetic contributions for developing chronic disease, including diagnosis, treatment, and prevention. The *All of Us* Research Program (“program”) seeks to achieve its vision of participant return of value and evaluation of its impact in genetic counseling. The program is poised to address practical challenges to enable genetic counseling for those who want to receive health-related DNA results, with an emphasis on diversity, equity, and inclusion (DEI) among populations known to be historically underrepresented in biomedical research (UBR). UBR groups are based on age, racial and ethnic identity, sexual and gender identity, sexual orientation, disability status, educational attainment, geography, and income. The program’s Genetic Counseling Resource will support the unique needs of UBR groups who want to receive genetic counseling for their health-related DNA results. The future of DEI in genetic counseling is pivotal for UBR groups and for the program overall in achieving its vision of providing value to its participants and evaluating its impact in research. A literature review was conducted to investigate evidence-based approaches toward increasing the efficacy of genetic counseling by understanding what types of information collected from genetic counseling sessions may be useful for returning value to research participants and for evaluation of impact. **Methods:** The literature review used PubMed and ScienceDirect 2016-2021 publications involving value and impact of genetic counseling among adults and 48 publications met inclusion criteria modeled after the Preferred Reporting Items for Systematic Reviews and Meta-Analysis based on their relevance to the review. **Results:** Efficient methods for genetic counseling that provide value to individuals and evaluation of impact include phone-based counseling accompanied by patient portals and educational tools (e.g., chatbots, videos, web-based, and printed materials); in-person counseling with visual aids; and post-genetic counseling routine patient engagement. **Conclusions:** Patient portals and educational tools enabling genetic counseling were shown to increase health and medical knowledge among individuals, especially UBR groups such as low-literacy populations and those who are not geographically in close proximity to genetic counseling services. This was significant for populations receiving access to and making use of genetic counseling services; including UBR populations, the Deaf community, and communities affected by orofacial clefts.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2228. The Million Veteran Program - Return Of Actionable Results (MVP-ROAR) Study: Participant Baseline Characteristics

Authors:

M. Danowski¹, M. Cardellino¹, C. Brunette¹, T. Yi¹, J. Vassy^{2,1,3,4}; ¹VA Boston Hlth.care System, Boston, MA, ²Harvard Med. Sch., Boston, MA, ³Brigham and Women's Hosp., Boston, MA, ⁴Precision Population Hlth., Ariadne Labs, Boston, MA

Abstract Body:

Introduction: The MVP-ROAR Study is a randomized controlled trial returning actionable genetic results associated with familial hypercholesterolemia (FH) to a subset of participants in MVP. FH is a monogenic disease characterized by a 20-fold increased risk of premature coronary artery disease due to severe elevations in low-density lipoprotein cholesterol (LDL-C). Participants are recontacted and reconsented to the study protocol. At baseline, enrollees complete a survey, provide a family history, and provide a biospecimen for LDL-C testing and clinical confirmation of the research finding, via a CLIA-certified laboratory. The genetic counselor (GC) returns the results by telephone and provides post-test counseling, including the provision of FH-related resources, facilitation of cascade testing, and documentation in the medical record.

Results: As of 5/9/22, a total of 385 participants with a suspected variant associated with FH have been recontacted and sent a study opt out letter. Of these, 11 opted out of further contact via postcard. Study staff sent a detailed mailing to 314 Veterans across 40 states. 177 were reached by phone at least once, and 65 have enrolled in the trial. Mean (SD) age of enrolled participants is 61.0 (13.3) years. 52 (80.0%) participants are men and 25 (38.5%) identify as non-white race. At enrollment, 13, 3, and 1 participants were taking a statin, fish oil, and ezetimibe, respectively. 44 participants have completed the baseline survey and biospecimen collection (blood = 39, saliva = 5). Initially 3/8 (37.5%) results were not clinically confirmed due to a high false positive call rate for rare variants on the MVP genotyping chip. After implementing an improved variant calling algorithm all subsequent results 23/28 (82.1%) were confirmed. Results have been disclosed by the GC to 28 participants. Survey data indicate some degree of medication aversion [Mean (SD) 21.4 (5.8) Beliefs About Medication] and reduced health-related quality of life among study participants, compared to the general population and those with FH [34.12 (11.12) Veterans Rand -12 Item Health Survey physical component score; 29.13 (14.25); mental component score]. 41 participants had a baseline LDL-C >70mg/dL, of whom 28 had a baseline LDL-C >100mg/dL.

Conclusion: Study participants are providing valuable information about the attitudes and outcomes of return of unanticipated genetic results within MVP. Additional data are needed to understand why participants declined to enroll after initial phone contact, but preliminary results suggest that enrollees, although with poorer baseline health, are engaged and generally receptive to receiving results.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2229. The Quebec Participatory cohort, a citizen science project for direct-to-consumer genetic testing participants

Authors:

S. Girard¹, M. Gagnon¹, C. Moreau¹, S. Gravel², M. Zawati³, Y. Joly³, H. Vezina¹; ¹Université du Québec à Chicoutimi, Chicoutimi, QC, Canada, ²Mc Gill Univ, Montreal, QC, Canada, ³McGill Univ., Montreal, QC, Canada

Abstract Body:

Direct-to-consumer genetic testing have become mainstream in the past decade. Through these tests, participants expect to learn more about their health and their ancestry. The collective investment of these participants could be further enhanced by academic efforts aiming to collect the raw data from these tests for research aiming at fostering knowledge and the public good. With this in mind, we have recently launched the pilot phase of the Quebec Participatory Cohort (CopaQ). Through this project, we aim to mobilize participants who acquired direct-to-consumer genetic testing through a private company. We first built a robust data management framework to ensure that our project will adhere to the appropriate ethical and legal policies in Canada. As expected from a citizen science project, data conservation and protection are key issues that need to be addressed in the right way. As such, we opted to develop our platform through a sustainable health hub geared toward data privacy. To collect raw genetic testing data, we built a secured web portal (<http://copaq.ca>) that provide users with a platform to collect informed consents, genotyping data as well as demographic and genealogical information. Through citizen participation, we will build a large cohort that will be made available to academic researchers. Amongst the conditions to access this cohort will be the need for the researchers to provide regular updates on the progress of their research project, in the form of a public-friendly video. Additionally, to attract and involve a large number of citizens, our team will provide participants with personalized genetic and genealogical reports. We believe that CopaQ can attract hundreds if not thousands of citizens who want to be involved in research projects. Our citizen science platform will also be a golden occasion to create a science-based community to drive citizen engagement even further.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2230*. The role of oncology nurses in the genetic counseling workflow: A literature scan

Authors:

E. Hasser¹, K. A. Leppig², N. B. Henrikson³; ¹Univ. of Washington, Seattle, WA, ²Kaiser Permanente of Washington, Seattle, WA, ³Kaiser Permanente Washington Hlth.Res. Inst, Seattle, WA

Abstract Body:

As the demand for genetic testing grows, a shortage of genetic counselors and bottleneck at the training stage limits the reach of the genetic counseling workforce. Nurses, especially infusion nurses spending extended time with cancer patients, are uniquely suited to relay genetic testing information. Training nurses in genetics could increase genetics competence and alleviate the workload burden on geneticists and genetic counselors. Minimal research has been published on the role of nurses in the genetic counseling workflow, but existing programs showed promising success. Collaboration with nurses to increase access to genetic counseling should be considered as a possible service delivery model and research area.

We conducted a literature scan and annotated bibliography to investigate the potential role of nurses in genetic counseling and to identify existing programs leveraging the nursing skill set, with a focus on oncology and infusion nurses. Combinations of terms in the categories of genetics (genetics, genomics, genetic testing, genetic counseling), nursing (nursing, nurses), and cancer (cancer, oncology) were searched in PubMed and Google Scholar. After an initial review, more specific terms such as “genetic counselor extender” and “service delivery model” were also implemented. The literature search window was Jan 2000 to May 2022. We included sources in the annotated bibliography if they contained relevant discussions about nurses in the genetic counseling workflow. We excluded sources that did not relate to nursing and genetics and those that addressed specialties other than oncology or a general care setting. We included 37 articles in the annotated bibliography. Sources discussing nurses’ knowledge of genetics were the most common (15 articles). Some articles identified that both practicing nurses and nursing students have limited training in genetics; others established core competencies for genetics in nursing curricula. Two dedicated genetics courses increased nurses’ confidence in genetics concepts. Several sources supported a potential for nurses to discuss genetic testing due to their high patient contact but did not propose a direct integration of nurses in the counseling workflow. Proposed innovative service delivery models can increase genetic counseling access and relieve administrative burden on genetic counselors. These alternatives to the traditional genetic counseling pathway included the use of telehealth genetic counseling; genetic counseling assistants; and nurses and nurse practitioners as genetic counselor extenders who actively participate in the counseling process by taking low complexity cases.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2231. The SeDeN-p2 study : Perceptions of French health professionals on the extension of newborn screening with or without genetics as a first-line test.

Authors:

C. Level^{1,2,3}, F. Huet^{4,3,5}, D. Salvi¹, C. Thauvin-Robinet^{1,3,2}, C. Binquet^{6,3}, E. Simon⁷, Inter-Filières de Santé Maladies Rares, France, Association Francophone de Génétique Clinique, Société Française de Médecine Prédictive et Personnalisée, C. Peyron^{8,3}, L. Faivre^{1,2,3,9}; ¹Ctr. de Génétique et Ctr. de Référence Anomalies du Développement et Syndromes Malformatifs, CHU Dijon Bourgogne, Dijon, France, ²Equipe Génétique des Anomalies du Développement, INSERM UMR 1231, Université de Bourgogne-Franche-Comté, Dijon, France, ³FHU TRANSLAD, Dijon, France, ⁴Service de Pédiatrie Multidisciplinaire, Hôpital d'Enfants, CHU Dijon Bourgogne, Dijon, France, ⁵Société Française de Dépistage Néonatal, National, France, ⁶Ctr. d'Investigation Clinique - Epidémiologie Clinique, INSERM UMR 1432, CHU Dijon Bourgogne, Dijon, France, ⁷Service de gynécologie obstétrique et médecine fœtale, CHU Dijon Bourgogne, Dijon, France, ⁸Laboratoire d'Économie de Dijon, Dijon, France, ⁹Filière de Santé Natl.e AnDDI-Rares Anomalies du Développement avec ou sans Déficience Intellectuelle de causes Rares, Dijon, France

Abstract Body:

Context: The recent modifications of the French bioethics law, the therapeutic progress and the massive development of advanced genetic techniques (such NGS) with a rapid decrease in costs imply to question the extension of newborn screening (NBS) to new actionable pathologies and the acceptable and relevant methods for its possible expansion. International studies are beginning to determine the potential place of NGS in NBS. In this perspective, the SeDeN project aims to fully assess the social acceptability of these issues by measuring the diversity and consistency of expectations of health professionals, parents and public policy makers. **Material and method:** The component of the SeDeN project presented here focuses on the opinions of professionals. This part is composed of a qualitative exploratory phase (12 interviews), a quantitative investigation phase based on a questionnaire survey and a qualitative in-depth phase. The online questionnaire, carried out by University Hospital Federation TRANSLAD in partnership with the French Society of Newborn Screening, was distributed via networks and learned societies to pediatricians, geneticists, genetic counselors, midwives, gynecologists, and molecular biologists in mainland France and the French overseas territories. Only the results of this questionnaire are presented here, with particular emphasis on the differences in responses between geneticists and other professionals. **Results and discussion:** The questionnaire was distributed from June 22 to December 20, 2021. 1199 professionals have completed all or part of the questionnaire and nearly 20% are geneticists or genetic counselors. Professionals expressed concerns about the use of genome-wide techniques in NBS and preferred the use of target panels. Concerns about the wider use of genetic techniques include ethical issues, potential anxiety for families, organizational difficulties in implementation, and lack of trained professionals. When practitioners were asked about adding different disease groups (childhood onset with no available treatment; adult onset with treatment or not; or able to provide information to relatives) to NBS, geneticists and genetic counselors were the most reluctant, as they were particularly aware of the impacts of genetic disease announcement. The same is true for the transmission of uncertain results. These data are important to better understand the opinion of French health professionals in a logic of decision support for health policies.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2232. The Texome Project: Initial enrollment in a rare and undiagnosed disease program for underserved communities in Texas

Authors:

R. German^{1,2}, C. Bacino^{1,3,4}, C. N. Murali^{3,1}, S. Lalani^{1,3}, T. Nguyen Dolphyn⁵, S. Baskin^{6,1}, E. Roeder^{6,1}, C. Schmid^{6,1}, R. Okashah Littlejohn^{6,1}, O. Juarez^{6,1}, L. Vossaert^{4,1}, N. Owen^{4,1}, C. Eng^{4,1}, P. Liu^{4,1}, Z. Liu^{1,2}, D. Mao^{1,2}, S. Pasupuleti^{1,2}, S. Kim^{1,2}, S. Yamamoto^{1,2,7,8}, M. Wangler^{1,2,8}, H. Bellen^{9,2,10,8}; ¹Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, ²Jan and Dan Duncan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, ³Texas Children's Hosp., Houston, TX, ⁴Baylor Genetics Lab., Houston, TX, ⁵Stanford Clinical Genomics Program, Stanford Hlth.Care, Stanford, CA, ⁶Dept. of Pediatrics, Baylor Coll. of Med., San Antonio, TX, ⁷Dept. of NeuroSci., Baylor Coll. of Med., Houston, TX, ⁸Program in Dev.al Biology, Baylor Coll. of Med., Houston, TX, ⁹Dept. of Molecular and Human Genetics, Baylor Coll. Med., Houston, TX, ¹⁰Howard Hughes Med. Inst., Houston, TX

Abstract Body:

The Texome Project provides crucial genetic testing, advanced bioinformatics analysis, and functional modeling to diagnose and study genetic disease in minority and underserved populations. National programs providing comprehensive genomic medicine have demonstrated the utility of genetic analysis on rare and complex undiagnosed disease. However, underserved groups and ethnic minorities have not equally benefited from large-scale genomics research and are largely underrepresented in genomic databases. To reach this population within Texas, we partnered and promoted enrollment with safety-net hospitals and community health clinics. To maximize accessibility to genetics services among our cohort, we offer virtual appointments via telemedicine, remote sample collections, and multilingual recruitment. The Texome Project also includes longitudinal follow-up which seeks to uncover the attitudes and feelings towards genomic medicine from an understudied population. At present, the Texome Project has received referrals for 55 ethnically and socioeconomically diverse families, 39 of whom we accepted. We completed enrollment for 28 families and whole exome sequencing (WES) for 14. A definitive diagnosis has been determined in 4 cases. Two of our solved cases are caused by previously reported variants in *ATPIA3* and *ASXL3* while the other two diagnostic variants in *DNMT3A* and *RPGR* are novel and not well studied. This gives our WES analysis a diagnostic yield of 29% to date. The clinical and molecular data provided by these patients contribute to new scientific discovery about the phenotypic spectrum and pathogenicity of specific variants. One significant challenge in genetic testing for our population is the availability of family members who may be separated by national borders, interpersonal issues, or premature mortality. This leads to some participants being analyzed on a proband or duo basis (n=10, 36%) instead of a more ideal trio analysis (n=18, 64%). For cases with non-definitive WES findings (n=10, 71%), further bioinformatic and functional modeling in *Drosophila* is ongoing which may implicate new gene-disease relationships. Preliminary findings of the Texome Project highlight the need and value of including underserved and underrepresented patients to provide equitable care and further scientific discovery.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2233. Transforming genetics service delivery: Genetics providers' perspective on integrating digital tools into clinical practice.

Authors:

W. Lee^{1,2}, D. Hirjikaka³, S. Grewal³, M. Clausen³, **S. Luca**¹, Y. Bombard^{3,2}, R. Hayeems^{1,2}, Genetics Navigator Study Team; ¹Program in Child Hlth.Evaluative Sci., The Hosp. for Sick Children, Toronto, ON, Canada, ²Inst. of Hlth.Policy, Management, and Evaluation, Univ. of Toronto, Toronto, ON, Canada, ³Genomics Hlth.Services Res. Program, Li Ka Shing Knowledge Inst., St. Michael's Hosp., Toronto, ON, Canada

Abstract Body:

Background: eHealth tools can provide scalable solutions to current challenges of access, efficiency, and equity in genetic service delivery. Various digital tools have been integrated into one or more stages of the genetic service pathway and have contributed to improving access to genetic services. Yet, genetics providers' perspectives on and comfort with digital tools are not well characterized. We explored providers' perspectives on the integration of digital tools into practice for both pediatric and adult patient populations. **Method:** Healthcare providers from genetics clinics across Canada were recruited via professional networks and snowball sampling. Using an interpretive description approach, key informant interviews were conducted. The interviews were transcribed and analyzed using thematic analysis. **Results:** Eleven geneticists and 21 genetic counselors across five provinces were interviewed (yrs in practice: 3 to >25). Participants had generally favorable attitudes towards digital tool integration as a strategy for alleviating pressures and inefficiencies in triaging, clinical assessment, pre-test counseling and education, thereby allowing providers to practice at the top of their scope. While many were hesitant, some providers noted that digital tools can be used for return of results in less complex cases (e.g., negative results). However, they expressed concerns that digital tools may create unintended burdens on providers' workloads associated with managing new digital applications. In addition, providers described ways in which digitization of genetics services could both enhance and compromise the patient-provider relationship. For example, some providers expected that the increased efficiency in practice would allow them more time to better address patient needs, while others worried that using digital tools would interfere with building patient/caregiver rapport. Finally, providers felt that digital tools could empower patients in their medical care, but they were concerned that this benefit may be limited to patients with adequate access to and proficiency with technology. **Conclusion:** Providers considered digital tools to be a potential solution for improving access, efficiency, and patient-centered care. However, successful integration of digital tools into care delivery requires careful consideration, evaluation and research of their potential unintended impacts related to digital burden for providers, compromised patient rapport, and widening the inequity that already exists in patient access to genetic services.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2234*. Tribal community perspectives on genomics research and data sharing: a mixed-methods study.

Authors:

K. S. Tsosie^{1,2,3}, J. Kim², J. J. Vasquez⁴, J. M. Yracheta^{2,5}, J. Li-Pook-Than⁴, M. Snyder⁴, L. G. Rosas⁴; ¹Arizona State Univ., Tempe, AZ, ²Native BioData Consortium, Eagle Butte, SD, ³Vanderbilt Univ., Nashville, TN, ⁴Stanford Univ. Sch. of Med., Palo Alto, CA, ⁵Missouri Breaks Industries Res., Inc., Eagle Butte, SD

Abstract Body:

Purpose: Precision medicine will likely fail to redress health disparities without equitable approaches for the collection and sharing of genomics and health data from American Indian and Alaska Natives. **Methods:** Convergent mixed methods characterized Tribal members' perspectives on genomics research and data privacy, sharing, and deposition. Survey data were collected from adult (median age=40.0 years) enrolled members of the Cheyenne River Sioux Tribe (N=125) recruited from Tribal district health fairs to assess positive-negative motivators contributing to trust and study participation. Respondents were purposively selected for focus group (total N=20, 3 cohorts) interviews, coded for content and thematic analyses. **Results:** Survey respondents expressed increased willingness to participate in studies led or partnered with the Indian Health Service (IHS; 72.8%) or nonprofit organizations (70.4%), compared to non-IHS federal institutions (52.8%) or pharmaceutical companies (36.8%). Respondents with higher educational attainment were more likely to be concerned about who can access their genetic data ($P=0.023$) and fate of biological material at the end of a study ($P=0.010$) compared to those without college experience. Focus group participants expressed concerns that removing individual identifiers would not adequately mitigate risks of hacking, re-identification, and reification of racial purity assumptions about Tribes. Intriguingly, Tribal participants called for increased transparency in who is accessing their data, which could be a pathway for building trust. Trust is intrinsically linked to derived utility of Tribal data back to communities. Participants stressed the importance of having data sharing protocols codified under Tribal law before Tribal data is collected. **Conclusions:** Trust is an important factor for Tribal members on whether to participate in research, as contextualized by contentious relationships between Tribal nations and federal agencies. Tribal members are wary of research motives, particularly for-profit interests, and open data sharing contributing to commercialization and privatization of Indigenous data. **Policy:** These findings inform data policies (e.g., National Institutes of Health Genomic Data Sharing Policy) and Tribal engagement. A model was developed encapsulating themes of data sharing, access, security, and stewardship with embedded histories of trust and distrust which, in turn, relate to facilitating direct community benefits and long-term Tribal data governance. Hence, fostering Tribal data agency can dismantle power imbalances and encourage Tribal research participation.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2235. Understanding views towards gene editing in Switzerland

Authors:

K. Ormond, S. Gächter, E. Vayena; ETH-Zurich, Zurich, Switzerland

Abstract Body:

Although the debate about modifying human genetics has been going on for several decades, the discussion about gene editing remains heated. Several large groups have presented views of experts and/or the general public, but through this work it is also clear that views vary from country to country, due in part to cultural and religious variation as well as existing legal frameworks. In Switzerland, in addition to having ratified the Oviedo Convention, there are a number of laws that restrict the genetic manipulation of embryos or reproductive germ cells, which would subsequently prohibit human germline gene therapy. In contrast, somatic gene therapy research can be carried out when approved. In this study we present data from semi-structured interviews with Swiss experts. The potential participants included basic research and translational scientists, physicians, lawyers and bioethicists. A list of ~150 potential expert participants were identified based on their roles at the major Swiss Universities, University Hospitals and Children's Hospitals, including professionals working specifically in genetics, genomics and the field of rare disease. We also utilized lists of board certified (FMH/FAMH) medical geneticists in the country, and individuals who participated in a Swiss Expert Committee for Biosafety (SECB) Gene Therapy Working group, the Swiss National Ethics report on gene editing, or TA-Swiss, the Foundation for Technology Assessment. Interviews are underway as of June 2022 and will be conducted until data saturation is reached. Data will be recorded, transcribed, and coded using an inductive coding method. From this, themes will be identified and presented through discussion with the research team and a feedback and deliberation process with study participants. Their views will be compared with already published literature in this area, and presented with a summary of relevant laws that exist in Switzerland. The results from this interview project will inform a country-wide survey of the general public (planned for Fall 2022) and a deliberative democracy event in early 2023. We aim to incorporate social, cultural, religious, and traditional aspects unique to Swiss society among other things that might play a role in defining bioethics, policy-making, and regulatory practices of gene editing for the therapy of hereditary diseases in Switzerland.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2236. Users' evaluation of artificial intelligence-based diagnostic program in pediatric rare diseases

Authors:

I. Choi¹, J. Kim¹, G-H. Kim², B. Lee^{2,1}; ¹Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, ²Asan Med Ctr, Seoul, Korea, Republic of

Abstract Body:

Since the early 2000s, with the advent of the next generation sequencing (NGS) technique, genome sequencing using the NGS technique, such as whole-exome sequencing (WES) or whole genome sequencing (WGS), showed higher clinical utility than the traditional tests. Furthermore the artificial intelligence (AI)-based genetic diagnostic program has been applied to genome sequencing to facilitate the diagnostic process. The Korean version of the artificial intelligence (AI)-based diagnostic program utilized big data obtained from whole-exome sequencing (WES) performed by next-generation sequencing to shorten the time required for pediatric rare disease diagnosis. This project was a Korean government-driven project to develop the AI-based medical service in South Korea, which has been named as “Dr. Answer 1.0.” It was started in 2018 to support clinical treatment and diagnosis using medical big data. It aims to provide precision medical services through early diagnosis and personalized gene treatment using genetic diagnosis. The objective of the current study was to evaluate the experience and level of satisfaction with the program for patients, guardians, and physicians who participated in the AI-based genetic diagnostic program for rare pediatric genetic diseases. The study period was from April 2020 to March 2021. A survey was designed to assess their experience and level of satisfaction. A total of 30 physicians and 243 patients and guardians (199 neurodevelopmental disorders and 44 hearing impairments) completed the survey. DNA samples of the subjects were collected through buccal swabs or blood collection: 211 subjects (86.8%) through buccal swab and 29 subjects (11.9%) through blood collection. Average turnaround time for result receipt was 57.54 ± 32.42 days. As a result of the AI-based diagnostic program, the diagnostic rate of neurodevelopmental disorder was found to be 40.1% and that of hearing impairment was 52.3%. The level of satisfaction of the two groups participating in the AI-based diagnostic program was 8.31 ± 1.71 out of 10 in the patient and guardian group and 8.42 ± 1.23 in the physician group. As the number of AI-based health care solutions will increase in the medical genetics field, genetic counseling will be considered more important. In future studies, the awareness and satisfaction survey on genetic counseling service provided to patients with rare diseases will be investigated. * It was written by referring to a article of the same author that had been accepted for publication through Medicine journal at April, 2022.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2237*. Using variant databases to estimate disease prevalence for rare recessive disease

Authors:

S. Baxter¹, M. Singer-Berk¹, J. Wood¹, N. Watts¹, M. Wilson¹, H. Rehm², A. O'Donnell-Luria³; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²Massachusetts Gen. Hosp./Broad Inst. of MIT and Harvard, Boston, MA, ³Broad Inst. of MIT and Harvard/Children's Hosp. of Boston, Cambridge, MA

Abstract Body:

Prevalence of a condition is an important factor in decision-making by researchers and pharmaceutical companies to determine goals and resource allocation toward the condition, yet the prevalence of the vast majority of rare diseases is unknown. Increased knowledge about the prevalence of a condition, both globally and in specific sub-populations (e.g. Ashkenazi Jewish, East Asian), is important to advocacy groups and organizations that aim to extend outreach, support and to build a network that includes as many affected individuals and families as possible. Over the last 2 years our team has partnered with a network of patient organizations to estimate prevalence for 15 rare autosomal recessive conditions in the RareAsOne network. We collected pathogenic (P) and likely pathogenic (LP) variants from databases like ClinVar and the Human Gene Mutation Database (HGMD), as well as all predicted loss of function (pLoF) variants in gnomAD. All LP and P ClinVar/HGMD variants that are present in gnomAD were re-curated following modern ACMG/AMP curation guidelines and pLoF were curated following our own internally-developed loss of function curation protocol. Carrier frequency and disease prevalence were estimated for all populations in gnomAD with over 2,000 alleles, using the Hardy-Weinberg equation and gnomAD allele frequencies (AFs). The results were returned directly to the patient organizations and their experts on 90 minute zoom calls, accompanied by written materials. Across the 15 genes assessed, global carrier frequencies ranged from 1/284 to 1/11,891 and prevalence from 1/322,831 to 1/565,546,084. For some genes, we were able to identify populations with increased carrier frequencies that had not been previously recognized in the literature (ex. TANGO2, PLA2G6). Two of the key factors impacting the estimates were formal classification of the variants and time since gene discovery. Including high frequency variants with weak or minimal evidence for pathogenicity would overinflate the estimates, highlighting the importance for having conservative standards for which variants are included. Conversely, genes that were more recently associated with disease typically had lower estimated prevalence due to fewer LP/P variants in ClinVar, underlining the need for including these genes on panels and functional studies to resolve variants of uncertain significance. Groups used this data for a variety of purposes including but not limited to publications, patient recruitment, and outreach to pharmaceutical companies. A future goal of this project is to release an aggregate frequency calculator utilizing gnomAD data for use by the community.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2238. Utilizing chatbots for family communication: Uptake and engagement among familial hypercholesterolemia probands from a prospective, pragmatic trial.

Authors:

N. Walters¹, Z. T. Lindsey Mills¹, A. Brangan¹, G. Campbell-Salome¹, S. Savage², T. Schmidlen², K. M. Morgan¹, E. P. Tricou^{1,3}, M. N. Betts^{1,4}, L. K. Jones¹, A. Sturm^{1,5}; ¹Geisinger, Danville, PA, ²Invitae, San Francisco, CA, ³Family Heart Fndn., Pasadena, CA, ⁴WellSpan Hlth., York, PA, ⁵23andMe, Sunnyvale, CA

Abstract Body:

Background: Familial hypercholesterolemia (FH), a CDC Tier 1 condition, significantly increases one's risk for early heart attack and stroke. FH is treatable but vastly underdiagnosed, making it imperative that probands share risk information with relatives to identify more cases. Chatbots, internet-based conversational agents, can aid in family communication. A Family Sharing Chatbot (FSC) was sent to FH probands identified through Geisinger's MyCode® Community Health Initiative. The FSC allows probands to send relatives a link to a Cascade Chatbot (CC) that explains their FH risk and recommends action steps. This study examined uptake of and engagement with the FSC. **Methods:** Probands included in a prospective, pragmatic trial were sent three messages prompting use of the FSC via email, text, and/or the patient electronic health record (EHR) portal. An opt-out methodology was used. Data on chatbot engagement from 7/1/2021 through 5/12/2022 was collected from the EHR and the chatbot's HIPAA-secure web portal. Statistical analyses were conducted using SPSS version 26. **Results:** Of the 175 included probands, 21 (12%) opted out of receiving the FSC. A logistic regression was performed to examine the effects of age on the likelihood that probands would opt out, which was statistically significant ($\chi^2(7, 154) = 10.35, p < .001$). Older probands were more likely to opt out than younger probands [$\text{Exp}(B) = .942, (.909, .977)$]. No significant differences between sex and opt-out status were found. Communication preferences varied among the 154 (88%) remaining probands; most chose to receive the FSC via the patient EHR portal (91/154, 59%), followed by email (40/154, 26%), text (21/154, 14%), and both email and text (2/154, 1.3%). Seventy-five (49%) probands clicked the FSC link, 62 (40%) started the FSC, and 36 (23%) shared a CC link with at least one relative. More females shared (24/89, 27%) than men (12/65, 18%). The average age of probands who shared ($M = 52$ years) was lower than those who did not ($M = 56$ years). Probands who read two or more invites on the patient EHR portal shared more (20/69, 29%) than those who read one or none (1/22, 4.5%). Probands who had multiple contacts with the study team (genetic counseling, study follow-up calls) shared more, with 6/42 (14%), 23/90 (26%), and 7/22 (32%) completing one, two, and three contacts respectively. **Conclusions:** The FSC can help FH probands facilitate family communication with at-risk relatives. It is essential to offer alternatives (e.g., family letters) as these may be more desirable for older probands who were more likely to opt out of the FSC. Reminder messages and additional contacts may help increase engagement with the FSC.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2239*. Validation of automated electronic health record (EHR) data capture of hereditary breast and ovarian cancer and Lynch syndrome phenotypes

Authors:

J. Savatt¹, C. J. Buoy¹, S. M. Ney¹, C. R. Kelsey¹, A. H. Buchanan¹, N. Banet², M. A. Kelly¹, R. Puttagunta¹, H. Ramey¹, W. Fairbrother³, N. T. Strande¹; ¹Geisinger, Danville, PA, ²Cleveland Clinic, Cleveland, OH, ³Brown Univ., Providence, RI

Abstract Body:

Electronic health record (EHR)-linked genomic biobanks offer a wealth of data that can be leveraged to increase understanding of genetic disease. The power of such data to expand our understanding relies on capturing accurate phenotypic information. Yet, EHR International Classification of Disease (ICD 9/10) codes alone may not accurately reflect patient phenotypes. While manual chart reviews enable more reliable phenotyping, they are low throughput and often not feasible for large scale studies. We assessed the performance of EHR data extraction for 14 cancer and precancerous phenotypes associated with hereditary breast and ovarian cancer (HBOC) and Lynch syndrome in Geisinger's MyCode Community Health Initiative. MyCode is a biobank of blood and other samples from over 312,000 unselected participants who consent to health-related research including exome sequencing linked to EHR data. To identify phenotypes of interest, diagnosis codes (ICD 9/10 and Epic Diagnostic Grouper (EDG)) and diagnoses from the Geisinger tumor registry that collects data for all individuals diagnosed or treated in the health system for the National Cancer Database were annotated for inclusion and exclusion using clinician generated phenotype definitions. Data were then extracted for 7,710 MyCode participants. For each phenotype, blinded manual EHR reviews were completed for the first 50 phenotype positive participants, if available, and at least 100 phenotype negative individuals. Discrepancies between the extracted data and manual review were re-reviewed and diagnoses were added or removed from pulls, as needed. Sensitivity and specificity were determined for diagnostic codes alone, tumor registry data alone, and combined diagnostic code and tumor registry data abstraction. Eight phenotypes had at least 50 participants flagged as positive by the data pull. Combined diagnostic code and tumor registry data abstraction had a median sensitivity of 99.4% and median specificity of 98.6%. Combined data extraction for 10 phenotypes had sensitivities and specificities that were both greater than 95%; abstraction of rare and noncancerous phenotypes had poorer performance. Diagnostic codes alone had fewer false negatives compared to tumor registry data alone, however tumor registry data had higher specificity. While our data are promising for EHR phenotype extraction, systems that lack tumor registries may need additional EHR elements to enhance specificity. These results show most HBOC and LS phenotypes can be accurately identified using a combination of diagnostic codes and tumor registry data facilitating phenotype collection for future large scale genomic studies.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2240. Voices from the online right: A qualitative study of race and genetics discourse.

Authors:

O. Sommers¹, J-H. Yu², D. Bowen¹; ¹Univ. of Washington, Seattle, WA, ²Univ. of Washington Sch. of Med., Seattle, WA

Abstract Body:

Background: Scientific racism maintains unjust racial hierarchies by portraying racial inequality as an inevitable result of innate differences between races. Proponents of modern scientific racism often cite geneticists' ability to predict racial categorization from genetic ancestry as proof that there are significant biological differences between races. These ideas are often spread through online discourse and have been used to support White Supremacist "replacement" theories and justify racial violence. Methods: To investigate the prevalence of scientific racism in "civil conversations" online, semi-structured interviews were conducted with 9 adult individuals, recruited from the social media site Reddit, who reported having conversations about race and genetics online. Results: Most participants had concordant explanations about the relationship between race and genetics. There was disagreement about the primary cause of health differences between black and white Americans. Discrimination was either ranked as an important environmental factor or dismissed completely. Among those who dismissed discrimination, socioeconomic factors within the environment were given as more important than genetic factors in determining health differences; however, those socioeconomic factors were then explained by references to innate racial differences, thus naturalizing existing racial hierarchies. Conclusion: Despite variation in participants' articulation on the causes of racial health differences, the logics of those actively involved in online race and genetics discourse ultimately endorse racial differences as innate. Further analysis will explore how discussions about scientific racism shape what is considered "common sense" (i.e., taken for granted) in online communities.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Wednesday

PB2241. Whole-exome sequencing in Puerto Rican patients: approach to diagnosis in complex traits.

Authors:

E. Albino¹, C. Buxo², C. Chapel-Crespo³, S. Carlo⁴, A. Cornier⁵, Z. Powis⁶; ¹Univ. of Puerto Rico, Med. Sci. Campus, Sch. of Hlth.Professions, San Juan, PR, ²Univ. of Puerto Rico, Med. Sci. Campus, Sch. of Dental Med., San Juan, PR, ³Univ. Pediatric Hosp. Dr. Antonio Ortiz, Sch. of Med., Univ. of Puerto Rico, San Juan, PR, ⁴Ponce Hlth.Sci. Univ., Cabo Rojo, PR, ⁵San Jorge Children s Hosp./ Ponce Hlth.Sci. Univ., San Juan, PR, ⁶Quest Diagnostics, Kalamazoo, MI

Abstract Body:

Whole-exome sequencing (WES) has become the first diagnostic option, especially in populations with difficult access to genetic tests. The use of WES in diagnosis will accelerate the management in Hispanic Puerto Rican patients undergoing a diagnostic odyssey; however, interpretation can be complex as Puerto Ricans are poorly represented in genomic databases. In addition, due to the differences in the medical system in Puerto Rico, exome testing is often ordered as a first-tier test and as proband-only testing due to the reduced cost to the family. Our objective was to evaluate the use and diagnostic yield of genetic tests analyzed in a diagnostic reference laboratory for Puerto Rican patients with complex traits.

We retrospectively reviewed consecutive results and clinical information of 50 Puerto Rican probands undergoing WES. WES included intronic variants, copy number variant analysis, and mitochondrial genome sequencing. Statistical analysis was performed using Fisher's exact test.

WES reports from 50 Puerto Rican probands between 16 months to 36 years were evaluated. WES was the test choice for these probands who have many heterogeneous neurogenetic phenotypes and a normal chromosomal microarray result. WES positive results were reported in 12 of 50 (24.0%) of probands; 15 of 50 (30.0%) probands had an uncertain result. The remainder of results of probands; 23 of 50 (46%) were negative. While the positive and uncertain rates trended towards higher numbers than probands from other regions of the United States (for the same time period), differences were not statistically significant ($P = 0.0453$). All uncertain results were due to variants of uncertain significance in clinically relevant genes. Of 27 probands, 11 (40.7%) had a variant of uncertain significance reported in addition to the primary positive or VUS finding.

Almost a quarter (24%) of patients received positive results from WES in this cohort of Puerto Rican patients with complex traits and 30% of the probands had a variant of uncertain significance. As diversification occurs in population genomic databases and variant data is shared in knowledge bases like ClinVar, the number of reported VUSs is hypothesized to decrease; the trends in this study suggest the study population may be underrepresented in these databases, though larger studies are needed. Such studies may also aid in variant classification and add to the knowledge of genetics for Puerto Rican patients with complex traits.

Genetic Counseling, ELSI, Education, and Health Services Research Posters - Thursday

PB2242. Your genes & you: patient motivations regarding secondary findings from genomic studies

Authors:

P. Brock¹, S. Hovick¹, L. Byrne¹, L. Peterson², G. Brock¹, I. Lattimer¹, J. Lin¹, R. Webber¹, L. Woodruff¹, M. Towne³, C. Milliard³, K. Sweet¹; ¹The Ohio State Univ., Columbus, OH, ²Loyola Univ. Med. Ctr., Chicago, IL, ³Ambry Genetics, Aliso Viejo, CA

Abstract Body:

Genomic sequencing provides opportunity to identify pathogenic variants unrelated to the primary testing indication. These secondary findings (SF) may be actionable and detected in the absence of any signs of disease, and individuals and families may benefit from increased preventive care and/or management based on these results. Most guidance on returning SF comes from expert recommendations. In this study, we assessed participants' interests, attitudes, and motivations regarding SF, as well as perceived disease risk, risk-related beliefs, and informational needs. Following standard genetic counseling and clinical multigene panel testing, patients at our cardiology and hereditary oncology clinics are offered enrollment in the Genomic Medicine Initiative (GMI) biobank study for further genomic investigation including research exome analysis and return of actionable SF results. A REDCap survey that assessed interests, attitudes, and motivations regarding SF, as well as perceived disease risk, risk-related beliefs, and information needs was emailed to 315 GMI participants. We received a total of 109 responses (35% response rate) from patients who underwent testing for a personal history of cancer (N=103) or heart disease (N=6). Of the respondents, 63% were male and 95% were White with an average age of 59 years (range 30-82). The most common primary motivation for enrolling in the GMI was to have the most up-to-date genetic information to provide family members (44%) followed by contributing to scientific knowledge and understanding (23%). Participants strongly agreed that they enrolled for scientific information (>90%), to provide information for their family (>90%), to change their health care (80%), and/or to gain additional explanations for their health condition (81%). Just 36% "strongly agreed" that they signed up because they were encouraged by a health care professional. Participants who enrolled primarily to provide information for family members were more likely to be older (mean=63.75, SD=10.77) compared to participants who enrolled primarily to contribute to scientific knowledge (mean=54.28, SD=14.06, p=0.016). Two separate descriptive disease scenarios addressed interest in receiving SF for either cancer or heart disease, and respondents reported equal interest in both types of SF. These results provide important insights into the motivations and interest in genomic research with the possibility of the return of SF results. Most participants report hoping to provide information for their family over information for themselves, and personal motivations may have a bigger impact than encouragement from health care providers.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2243. A calibrated automated patch clamp assay can interpret any VUS in *KCNH2*.

Authors:

C. Ng^{1,2}, R. Ullah³, J. Farr^{1,4}, A. P. Hill^{1,2}, **J. I. Vandenberg**^{1,2}, B. M. Kroncke³; ¹Victor Chang Cardiac Res. Inst., Darlinghurst, Australia, ²Sch. of Clinical Med., UNSW Sydney, Darlinghurst, Australia, ³Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁴Sch. of Computer Sci. and Engineering, UNSW Sydney, Kensington, Australia

Abstract Body:

Introduction: Pathogenic *KCNH2* variants can cause long QT syndrome (LQTS) and increase the risk of sudden cardiac arrest. More than half of the nonsynonymous variants in *KCNH2* listed in the ClinVar database are classified as variants of uncertain significance (VUSs). *In vitro* functional assays can provide evidence to strengthen the case for reclassifying VUS as either likely benign or likely pathogenic. **Aim:** To use an automated patch clamp assay, calibrated according to the recommendations published by the ClinGen SVI working group, to assess the function and provide evidence for VUS reclassification for *KCNH2* variants identified in this cohort. **Method:** A cohort of 216 nonsynonymous missense variants in *KCNH2* were collated from LQTS cohorts in France, Italy and Japan (Kozek et al., 2021). Heterozygous *KCNH2* variants were expressed in HEK293 cells and characterised using an automated patch clamp electrophysiology platform. Variants were classified as functionally normal or abnormal based on previously established current density and gating parameters established for this assay (Jiang et al., 2022). **Results:** Out of the 216 unique missense variants identified from this cohort, 188 were identified in individuals with clinically confirmed LQTS whereas 28 in individuals with suspected LQTS. 172 of the 188 variants in clinically confirmed LQTS were found to be functionally abnormal compared to 15/28 variants in suspected LQTS cases. 154 of the 216 variants have been previously classified in ClinVar database; 58 classified as (likely)pathogenic, 79 VUSs, 17 with conflicting classification. 54/58 ClinVar variants classified as (likely)pathogenic were found to be functionally abnormal. Furthermore, 58/79 VUS and 14/16 conflicting variants were also functionally abnormal. Most of the variants in the voltage sensor domain, pore domain and cNBH domain are functionally abnormal; whereas only a minority of variants located within the proximal N and distal C terminus are functionally abnormal. Variants from the EAG domain can have normal and abnormal function. **Conclusion:** Our *in vitro* functional genomics assay for *KCNH2* variants can provide the additional evidence needed for VUS reclassification. This is invaluable, as reclassifying a VUS can facilitate cascade screening of the extended family. **Reference:** 1. Kozek et al., Estimating the Posttest Probability of Long QT Syndrome Diagnosis for Rare *KCNH2* Variants. *Circ Genom Precis Med* (2021) 2. Jiang et al., A calibrated functional patch clamp assay to enhance clinical variant interpretation in *KCNH2*-related long QT syndrome. *Am J Hum Genet* (2022)

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2244. A Case of Concurrent 3p26.3-p25.3 Deletion and 9q34.2-q34.3 Duplication in Monozygotic Twin Males: A Case Report.

Authors:

T. Maini¹, J. Black¹, I. Cherrick², N. Brescia³, J. A. O'Malley², R. Lebel¹; ¹Section of Med. Genetics, SUNY Upstate Med. Univ., Syracuse, NY, ²Dept. of Pediatrics, SUNY Upstate Med. Univ., Syracuse, NY, ³Dept. of Neurology, SUNY Upstate Med. Univ., Syracuse, NY

Abstract Body:

Monochorionic diamniotic male twins were vaginally delivered at 33 weeks, to a 32-year-old G3P2 mother. The union was non-consanguineous. Intrauterine growth restriction of twin B had been noted. Both were small for gestational age and required ventilatory support. Both twins displayed micrognathia, hypotonia, bilateral ptosis, unilateral cryptorchidism and hydronephrosis, and gastric anomalies. Twin A had an atrial septal defect and a left SVC. Both failed a newborn hearing screen; twin A in both ears and twin B on the right side only. Twin A developed seizures soon after birth and twin B developed them later. Microarray revealed a 10,138 kbp deletion of 3p26.3-p25.3 and a 4,031 kbp duplication of 9q34.2-q34.3, in both boys. Karyotype confirmed 46,XY,der(3;9)(p25.3;q34.2) in both. The 3p anomaly overlaps with that of 3p deletion syndrome, with features including low birth weight, microcephaly, ptosis, micrognathia, hypertelorism, hypotonia, psychomotor and growth delay, hearing deficits and intellectual disability. Congenital heart disease, gastric and renal abnormalities are also commonly seen. The 9q34 duplication is a much rarer finding, and is associated with developmental and speech delay, poor feeding, and musculoskeletal abnormalities. Both twins remained in the NICU for an extended period. Six days after discharge from the NICU, twin A was readmitted to the pediatric ICU due to seizures. Twin B was transferred to the PICU from the NICU. Recurrent admissions followed for twin B, due to apneic episodes and feeding difficulties. He was found to have advanced pulmonary dysfunction and expired at 9 months; his autopsy showed pulmonary interstitial fibrosis, a dilated pulmonary artery, and right ventricular hypertrophy and dilation. Examination of the brain demonstrated acute on chronic diffuse hypoxic and ischemic damage and a mal-rotated right hippocampus lacking a dentate gyrus. Twin A had a similar but less fulminant course notable for apneic episodes, failure to thrive and seizures. He was transferred to comfort care, and expired at 14 months. No autopsy was performed. While 3p distal deletion syndrome is well established in the literature, 9q duplications are seldom reported. The combination of both has not been reported to date, and its occurrence in monozygotic twins is extraordinary.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2245. A combination of exome sequencing and optical genome mapping unveils a dual molecular diagnosis in a case with an unknown neurodevelopmental disorder.

Authors:

A. Acharya¹, N. Lin¹, S. Leal¹, J. Posey², I. Järvelä³, I. Schrauwen⁴; ¹Columbia Univ., New York, NY, ²Baylor Coll. of Med., Houston, TX, ³Biomedicum, Espoo, Finland, ⁴Columbia Univ., NYC, NY

Abstract Body:

An affected female with an apparent unknown syndrome that includes moderate intellectual disability, microcephaly, almond-shaped eyes, short philtrum, small chin, thin and dysmorphic fingers with a brachydactylic 4th finger and arachnodactylic 5th finger on the left, a slender habitus and mild epilepsy was investigated to determine the molecular etiology. We first performed exome sequencing, which revealed a 12bp deletion in the *GRIN2A* gene [p.(Ile151_Ala155delinsThr)] that was inherited from her father who also had mild epilepsy during childhood, that resolved similarly as previously seen in *GRIN2A* cases. The additional features of the daughter were unlikely due to the *GRIN2A* variant; therefore, we investigated this case further via optical genome mapping (OGM), a technology that utilizes long, linearized DNA to detect genomic structural variants (SVs) such as deletions, duplications, insertions, inversions and translocations. Optical genome mapping identified 6199 SVs, of which 44 were ultrarare (MAF<0.005) and 31 unique to the proband. The latter included a heterozygous 7kb insertion affecting *H3F3A*. Variants in this gene cause Bryant-Li-Bhoj neurodevelopmental syndrome 1, an autosomal dominant disorder with a variable phenotype including features seen in our case, such as intellectual disability, microcephaly, dysmorphic features including an abnormal philtrum, and distal skeletal defects. The variant is currently being further reconstructed via targeted long-read sequencing and Sanger sequencing, which will be presented at the meeting. In conclusion, via comprehensive molecular testing including exome sequencing and optical genome mapping, we identified a dual molecular diagnosis in an affected individual with an apparent unknown syndrome.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2246. A framework designed to improve the evaluation and reporting of incidental findings: Implementation and experience of a clinical laboratory in the first 15 months.

Authors:

C. M. Brown, L. M. Amendola, A. Chandrasekhar, A. J. Coffey, A. Kesari, R. J. Taft, D. L. Perry; Illumina Inc., San Diego, CA

Abstract Body:

Introduction: The Illumina Clinical Services Laboratory offers genome sequencing (GS) for rare disease and returns incidental findings (IFs), defined here as variants unrelated to testing indication not in genes on the American College of Medical Genetics and Genomics secondary findings list, that impact medical management. Lack of professional guidelines prompted us to design a framework for IF evaluation and return. Here we describe this framework and outcomes of its integration into our workflow.

Methods: Necessary properties of reportable IFs were defined and a stepwise process to evaluate potential IFs, with stopping points at which variants are deemed reportable or not, was developed. We assessed the number and types of IFs returned between January 2021 and March 2022.

Results: Clinical significance of the genotype and actionability of the associated disorder were identified as key features of reportable IFs. Single gene variants meet clinical significance criteria if they are in the correct zygotic state, classified as pathogenic (P) or likely pathogenic (LP), and have a gene-disease relationship of strong or definitive. For mitochondrial variants, heteroplasmy level is also considered. Multigene copy number variants must be classified as P or LP. Expanded repeats are not returned as IFs. Actionability criteria are met if ClinGen actionability scores for the relevant disorder are at least moderate (total score ≥ 7) with evidence of effective interventions (effectiveness score > 0). If scores are low, or unavailable, the literature is searched. IFs are returned if pre-symptomatic surveillance, interventions or circumstances to avoid are described. Variants in genes associated with disorders with palliative treatments not specific to the genotype, or management strategies in early stage clinical trials, are not returned as IFs.

Over 15 months, 726 individuals had GS. IFs were returned in 31 individuals from 30 families (4.3%) and were mostly associated with disorders of the blood and blood forming tissues (2.1%), cancer predisposition/neoplasm (1.2%), and renal or cardiac disorders (both 0.3%). Variants in the following genes were reported at least

once: *APRT, CHEK2, CLCN1, DDX41, F11, G6PD, HBB, HNF1A, HOXB13, MT-RNR1, NF1, PKD1, RAD51C, SDHA, TTN, TTR* and *WRN*. Of the 17 associated disorders, actionability was established using ClinGen data for 7 and the literature for 10.

Discussion: This framework was successfully incorporated into our laboratory workflow and led to the return of actionable findings in 4.3% of patients. It could serve as a model for other groups considering standardization in the return of IFs encountered during genomic testing.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2247. A heterozygous de novo VUS (p.M130I) in candidate gene PSMC5 in a male with autism, mild developmental delay, speech deficiency, dysmorphic features, hypospadias and chordee.

Authors:

M. Hajianpour¹, C. Kois², J. A. Lundberg¹; ¹Albany Med. Coll./ Div. of Med. Genetics and Genomics, Dept. of Pediatrics, Albany Med. Ctr., Albany, NY, ²Nemours Children's Hosp., Wilmington, DE

Abstract Body:

Proband is a 23-month-old male with autism spectrum disorder, mild developmental delay and speech deficiency, hypospadias, and chordee (status post chordee repair, meatotomy and scrotoplasty). He was born FT, BW: 6lbs 1oz, BL: 18 inches. Developmental milestones were mildly delayed: He walked at 18 months, talked in single word at 9 months and in 2-3 words sentences at 21 months of age. He is receiving PT, OT, and ST. He has now a vocabulary of ~100 words, counts to 20, and knows alphabets. **Physical exam** (at 23m): Ht: 28th%ile, Wt: 40th%ile, HC: 60th%ile for age. He was not cooperative during the exam. He had occasional hand flapping when excited. **Dysmorphic features:** Mild positional plagiocephaly (prominent on the left), dolichocephaly, prominent mid-forehead; slightly posteriorly rotated ears; down-slanted palpebral fissures, epicanthal folds, long eyelashes; low/ broad nasal bridge and root, anteverted nares; thin upper vermilion; mild microretrognathia; hypoplastic nipples; mild clinodactyly of the 5th fingers, and bilateral clinodactyly of 2nd toes laterally overlapping the 3rd toes. **CNS:** CN II-XII are grossly intact. Good muscle tone. **GENETIC WORK-UP:** [1] Chromosome microarray (CMA) analysis: A Paternally inherited 580 Kb duplication at chromosome 17q25.3: a variant of uncertain significance (VUS): Not contributing to the phenotype. [2] Fragile X studies: Normal, 31 CGG repeats. 3. Molecular studies: GeneDX expanded Autism/ID panel TRIO (2500 genes): (i) A heterozygous maternally inherited VUS in KDM1A gene: p.A197T, Autosomal dominant - not contributing to phenotype. (ii) A hemizygous maternally inherited VUS of gene **ZNF674**: p.E150K A candidate gene/ may contribute to phenotype. (iii) A heterozygous, de novo VUS in candidate gene **PSMC5**: c.390 G>p.M130I: May contribute to the phenotype (**under investigation/ research**). *No PSMC5-related disorder is listed in OMIM.*
PSMC5: Member of the AAA (ATPases Associated with diverse cellular Activities) gene family. It may be involved in the regulatory subunit of the 26S protease. *PSMC5 (Proteasome 26S Subunit, ATPase 5) is a Protein Coding gene.* Reported diseases associated with PSMC5: Mulibrey nanism and Ogden syndrome. Patient does not currently have features of these two conditions.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2248. A multiomics approach to resolving small supernumerary marker chromosomes

Authors:

C. Grochowski¹, M. Ghandi², H. Du³, M. Mehaffey², K. Park², J. Rosenfeld³, S. Darilek¹, J. Eisefeldt⁴, M. Pettersson⁴, L. Potocki⁵, J. Lupski³, C. M. B. Carvalho²; ¹Baylor Coll. Med., Houston, TX, ²Pacific Northwest Res. Inst., Seattle, WA, ³Baylor Coll. of Med., Houston, TX, ⁴Karolinska Inst.t, Solna, Sweden, ⁵Baylor Col Med/TX Child Hosp, Houston, TX

Abstract Body:

Small supernumerary marker chromosomes (sSMC) are chromosomal fragments that are inherently challenging to characterize due to their small size, abnormal genetic structure, frequent occurrence of breakpoints within centromeric sequence as well as their heterogeneous and mosaic nature. Markers are hypothesized to be formed due to a trisomy rescue event resulting in incomplete degradation of DNA contained within micronuclei structures that are generated from aberrant anaphase chromosome separation and/or through a chromoanagenesis repair process. They display sporadic Mendelian inheritance patterns and the clinical consequence when they are identified remains elusive. We sought to fully resolve the architecture and infer the mechanistic origins through a multiomics approach including custom high-resolution array comparative genomic hybridization (aCGH), short-read whole genome sequencing (WGS) as well as optical genome mapping (OGM). Our heterogeneous cohort includes four families that were referred for genetic testing due to various clinical presentations that identified a *de novo* sSMC during G-banded karyotyping studies. The probands included 1 boy and 3 girls with markers that were derived from chromosomes 2 (N=1), 17 (N=2) and 19 (N=1). Further characterization using custom aCGH revealed an apparent size distribution of 4Mb to 10Mb with a log₂ ratio confirming the suspicion of mosaicism in at least 2 probands. Visualization of copy number gains in short-read WGS and the bioinformatic tool *VizCNV* revealed complexities not previously identified. One sample revealed a complex sSMC with a triplication embedded within duplications involving genomic material from chromosome 19. Subsequent analysis using OGM revealed the triplication to be inverted within this sSMC's architecture. In all samples studied, further analysis of the breakpoint regions within the proband in comparison with both parental samples will help to elucidate the final structure, dosage sensitive genes contained on the sSMC as well points of genomic instability that may have mediated the event. As previous studies (Grochowski *et al.*, 2018 (PMID:29696747) and Bodkin *et al.*, 2019 (PMID:31279534)) have highlighted, the complete architectural mapping of an sSMC is crucial in understanding the phenotypic presentation in a patient and targeting therapies in the pursuit of precision medicine.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2249. A multiple congenital anomalies newborn with coexistence of extra der(22) chromosome marker and balanced t(11;22) maternally inherited

Authors:

M. Albalwi¹, R. Alahmadi², D. Al-Bardisi², K. Alshalan², M. Siraj², A. Alzaid², N. Aldawsari², M. Alfadhel³, M. Alshammari⁴; ¹Dept. of Pathology & laboratory Med., King Abdulaziz Med. City, Riyadh, Saudi Arabia, ²Dept. of Med. Lab., King Khalid Univ. Hosp., King Saud Univ. Med. City, Riyadh, Saudi Arabia, ³Genetics & Precision Med. Dept., King Abdullah Specialist Children's Hosp., Riyadh, Saudi Arabia, ⁴Med. Genetics Div., Pediatrics Dept., King Khalid Univ. Hosp., King Saud Univ. Med. City, Riyadh, Saudi Arabia

Abstract Body:

Supernumerary derivative 22 [der(22)] syndrome is a rare chromosomal genomic disorder that is associated with multiple congenital anomalies, congenital heart defects, renal anomalies and male genital abnormalities. Here, we describe a one month male newborn with extra supernumerary derivative 22 marker and a balanced translocation t(11;22) inherited from his mother [47,XY, t(11;22)(q23.3;q11.1)mat, +der(22)t(11;22)(q23.3;q11.1)]. Coexisting of extra der(22) and balance t(11;22) rarely reported or not seen in the literature. Proband clinical features was associated with distinctive feature of deep-set eyes, relatively large ears with preauricular tag and cleft palate. Also, he has a congenital heart defects, absent unilateral kidneys and bilateral undescended testis. His genetic analysis abnormalities were also confirmed by FISH and chromosomal microarray analysis. Proper clinical evaluation and laboratory investigations are important for a proper family planning and genetic counseling particularly in a highly consanguineous population.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2250. A national initiative working toward equitable translational genomic research for Canadian Indigenous families: Early findings indicate the need to reduce uncertainty in variant interpretation.

Authors:

A. Mohajeri¹, **K. Jacob**¹, S. McIntosh¹, J. mwenifumbo², T. Maroilley³, V. Avramovic³, F. Bernier⁴, K. Boycott⁵, L. Badalato⁵, O. Caluseriu⁶, L. Chad⁷, G. Costain⁷, I. De Bie⁸, C. Greenberg⁹, A-M. Laberge¹⁰, J-B. Riviere¹¹, V. Siu¹², L. Turner¹³, W. Wasserman¹, N. Caron¹⁴, Silent Genomes Precision Diagnosis Consortium, M. Tarailo-Graovac¹⁵, L. Arbour¹, A. Lehman¹; ¹Dept. of Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, ²BC Children's Hosp. and BC Women's Hosp., Dept. of Pathology and Lab. Med., Vancouver, BC, Canada, ³Cumming Sch. of Med., Univ. of Calgary, Calgary, AB, Canada, ⁴Alberta Children's Hosp., Calgary, AB, Canada, ⁵Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, ⁶Univ. of Alberta, Edmonton, AB, Canada, ⁷The Hosp. for Sick Children, Toronto, ON, Canada, ⁸Montreal Children's Hosp., Montreal, QC, Canada, ⁹Children's Hosp. of Winnipeg, Winnipeg, MB, Canada, ¹⁰Dept. of Pediatrics, CHU Sainte-Justine, Montreal, QC, Canada, ¹¹McGill Univ. Hlth.Ctr., Montreal, QC, Canada, ¹²Univ. of Western Ontario, London, ON, Canada, ¹³Mem. Univ. of Newfoundland, St. John's, NL, Canada, ¹⁴UNBC, Prince George, BC, Canada, ¹⁵Univ. of Calgary, Calgary, AB, Canada

Abstract Body:

Problematic barriers to health care for Indigenous populations in Canada (First Nations, Inuit and Métis People) affect Indigenous participation in clinical trials and translational precision health studies, including projects to improve genomic diagnosis. Furthermore, under-representation of Indigenous individuals in genomic research has led to a lack of background reference information and less precise genomic interpretation, compounding inequities for those with rare genetic disease. The Silent Genomes Project aims to reduce gaps in access to accurate diagnosis, genomic resources and genomic health parity for Indigenous families through several projects which strive to prioritize Indigenous cultural safety, including a translational diagnostic genome-wide sequencing study only open to enrollment for those identifying as Indigenous, who have not obtained a diagnosis through routine clinical care. A majority of medical genetics clinics across Canada are participating. At the time of submission, 56 families have undergone whole genome sequencing, and results are available for 47 families. Reported candidate variants were categorized according to overall diagnostic relevance after conferencing among multiple clinical experts; categories included Definite, Probable, Uncertain, and Negative Diagnosis, as well as Research Candidate, and Incidental Finding. “Research” variants were typically in genes not yet firmly established as causative of human disease. Eleven families (23%) had no reportable findings (negative). Twenty-four families had only one reported candidate variant, and 12 families had more than one. Overall, we reported 60 candidate variants including 14 definite, 5 probable, 21 research candidates, and 20 uncertain. There were also 3 incidental findings in two families. Sixteen families (34%) received at least one diagnosis. The fact that the majority of reported variants (41 of 60; 68%) introduced uncertain possibilities—either in known or novel disease genes—speaks to the need for improved analytical precision. Deeper knowledge of the frequency of alleles in Indigenous populations should substantially reduce the number of uncertain variants being returned to families.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2251*. A new neurodevelopmental disorder with microcephaly and neural tube defect Loss of NARS1 leads to microcephaly and neural tube defects

Authors:

S. Temel^{1,2}, E. Oruç^{3,4}, A. Kahraman^{3,4}, S. Ozemri Sag¹, E. Eren⁵, M. Ergoren⁶, S. Gul⁷, MarmaRare Group, E. Deniz⁸, A. Fuss^{3,4}; ¹Bursa Uludag Univ., Faculty of Med., Dept. of Med. Genetics, Bursa, Turkey, ²Bursa Uludag Univ., Inst. of Hlth.Sci., Dept. of Translational Med., Bursa, Turkey, ³Bogazici Univ., Dept. of Molecular Biology and Genetics, Istanbul, Turkey, ⁴Bogazici Univ., Ctr. for Life Sci. and Technologies, Istanbul, Turkey, ⁵Bursa Uludag Univ., Faculty of Med., Dept. of Pediatric Endocrinology, Bursa, Turkey, ⁶Near East Univ., Faculty of Med., Dept. of Med. Genetics, Nicosia, Cyprus, ⁷Biotechnology Div., Dept. of Biology, Istanbul Univ., Istanbul, Turkey, ⁸Yale Univ., Sch. of Med., Dept. of Pediatrics Critical Care, New Haven, CT

Abstract Body:

Aminoacyl-tRNA synthetases (ARSs) are enzymes that attach the appropriate amino acid onto its corresponding tRNA and are essential for protein translation. Mutations in ARS genes have been implicated in various human diseases, including neurological, autoimmune, and cancer. One of the thirty-seven ARSs Asparaginyl-tRNA synthetase1 (NARS1) functions in the cytoplasm responsible for asparagine tRNA charging and implicated in neurodevelopmental delay with microcephaly due to toxic gain-of-function and partial loss-of-function effects. **Here**, we report a family with a ten-month-old deceased boy who presented with microcephaly and dysmorphic features. He was the second-born from a consanguineous marriage at 33+5 week gestational age. He presented with resistant epilepsy, neurodevelopmental delay, neonatal diabetes, inguinal hernia, hydrocele testis humoral immunodeficiency, congenital heart disease (VSD, pulmonary stenosis), and multiple cavernous malformations in the brain. The sibling was the third-born female who presented with microcephaly and encephalocele. She also deceased on her first day of birth. Exome sequencing revealed a homozygous missense mutation c.866A>G, (p.Tyr289Cys) in the NARS1 gene, the segregation in the family was confirmed by Sanger sequencing. We turned to animal models to investigate the role of NARS1 in development and bolster the evidence that the variant is disease-causing. We used the frog *Xenopus* model for loss of function assays. We depleted *nars1* using the CRISPR/ CAS9 system targeting three non-overlapping exons. We analyzed brain size and the neural tube defects using optical coherence tomography imaging and showed that *Xenopus* NARS1 mutation leads to microcephaly and neural tube defects recapitulating our patient's phenotype. Next, to test the pathogenicity of the identified NARS1 variant, we used the fly *Drosophila melanogaster* as a model. Using a Gal4 line, we show that the fly orthologue of NARS1, AsnRS, is expressed in the larval and adult central nervous systems and in a subset of neurons. We have generated a mutant for AsnRS and transgenic fly lines expressing wild-type and mutant NARS1.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2252. A novel FAME1 repeat configuration in a European family identified using a combined genomics approach

Authors:

T. Maroilley^{1,2,3,4}, M-H. Tsai^{5,6,4}, R. Mascarenhas^{3,7}, C. Diao^{1,2,3}, M. Khanbabaei⁷, S. Kaya⁸, C. Depienne⁸, M. Tarailo-Graovac^{1,2,3,9}, K. Klein^{2,3,7,9}; ¹Dept. of Biochemistry and Molecular Biology, Cumming Sch. of Med., Univ. of Calgary, Calgary, AB, Canada, ²Dept. of Med. Genetics, Cumming Sch. of Med., Univ. of Calgary, Calgary, AB, Canada, ³Alberta Children's Hosp. Res. Inst., Univ. of Calgary, Calgary, AB, Canada, ⁴Equal, First, AB, Canada, ⁵Dept. of Med. Res., Kaohsiung Chang Gung Mem. Hosp., Taiwan, Taiwan, ⁶Sch. of Med., Coll. of Med., Chang Gung Univ., Chang Gung, Taiwan, ⁷Dept. of Clinical NeuroSci.s, Hotchkiss Brain Inst., Univ. of Calgary, Calgary, AB, Canada, ⁸Inst. of Human Genetics, Univ. Hosp. Essen, Univ. of Duisburg-Essen, Essen, Germany, ⁹Equal, Senior Authors, AB, Canada

Abstract Body:

Familial adult myoclonic epilepsy (FAME) is an autosomal dominant, neurological disease characterized by adult-onset of cortical tremors and infrequent generalized tonic-clonic seizures. Recently, using long-read sequencing, FAME was found to be caused by a pathogenic pentanucleotide repeat expansion, a combination of a TTTCA repeat expansion associated with a polymorphic TTTTA repeat. Thus far, six intronic loci with this pentanucleotide repeat combination have been described, suggesting that the repeat itself plays the main role in the pathogenesis of the disease regardless of the locus/gene where the repeat is located. Common ancestral founder effects may play a role as individual loci expansions have been only associated with specific populations. For example, the *SAMD12* expansions (FAME1) have only been reported in patients of Asian descent. Importantly, a recent detailed analysis performed in the FAME1 families revealed that the repeats have variable size, configuration, and composition. The TTTCA repeats can be very long (> 1000 repeats), but also very short (14 being the shortest identified); and of different configurations and compositions, suggesting a complex relationship between the combination of repeat expansions and disease manifestation. Here, we report two siblings (brother and sister) of European descent with a clinical diagnosis of FAME, yet negative repeat-primed PCR (RP-PCR) test. Using short-read whole genome sequencing (srWGS), we identified a pentanucleotide expansion in intron 4 of *SAMD12*. We used Optical Genome Mapping (OGM) to validate our findings, which confirmed the 4.1kb expansion in *SAMD12*. We also used Cas9-mediated enrichment and long-read sequencing, which confirmed a novel (not previously reported) configuration of the repeat: (TTTTA)₋₈₀₀(TTTCA)₃(TTTTA)₇(TTTCA)₇. Our study is the first to associate the *SAMD12* locus in European FAME patients and currently represents the shortest identified TTTCA expansion, which explains why it was missed using the RP-PCR test. These results also show that we are still learning about FAME and integrating additional genomics approaches (e.g., OGM and srWGS) may help facilitate the discovery of novel configurations and thus mechanisms of this disease.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2253. A rare association of congenital hypohydrotic ectodermal dysplasia, secondary hypogonadism, and severe iron deficiency anemia.

Authors:

D. Selvarajan¹, N. Raman¹, P. Jayapal²; ¹Thiruvarur Med. Ctr., Thiruvarur, India, ²Lucile Packard Children's Hosp., Stanford Univ., Pal Alto, CA

Abstract Body:

A 16-year-old male presented with breathlessness for the past one month. He was short for his age, severely pale, and non- edematous on examination. His secondary sexual characteristics were poorly developed with thin scalp hair, sparse axillary hair, absent pubic hair, micropenis, poorly developed testes, and atrophic nails with grade 1 clubbing. He also had partial anodontia. His bone age was less than chronological age. He gave a history of heat intolerance and absent sweating since childhood. He was born to non-consanguineous parents by full-term delivery. He attained milestones on time. His scholastic performance had been good. His sibling was normal. No other family member had a similar illness. On evaluation, he was found to have severe iron deficiency anemia with Hemoglobin of 6 g/dl. He had low testosterone, FSH, and LH which are suggestive of secondary hypogonadism. A clinical exome test was done to analyze the genes associated with anemia, ectodermal dysplasia, and hypogonadotropic hypogonadism, which showed a heterozygous pathogenic variant involving exon 12 of the EDAR gene indicative of Hypohydrotic ectodermal dysplasia. The other pathogenic genes for hypogonadotropic hypogonadism and iron deficiency anemia were negative. To the best of our knowledge, this is the first case of Hypohydrotic ectodermal dysplasia reported in association with secondary hypogonadism and severe iron deficiency anemia. This combination could be a new syndrome as such.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2254. A recurrent *AFF3* pathogenic variant associated to KINSSHIP syndrome in a Mexican patient

Authors:

L. Flores Gallegos¹, V. GUEVARA SANCHEZ¹, C. JUARISTI MANRIQUE², R. MORALES TOQUERO¹; ¹Hosp. Angeles Puebla, PUEBLA, Mexico, ²CENTOGENE, ROSTOCK, Germany

Abstract Body:

KINSSHIP syndrome was defined in 2021, as an autosomal dominant disorder characterized by a recognizable pattern of anomalies caused by *the novo* missense heterozygous mutations in the degnon of *AFF3* gene within chromosome 2q11. Voisin et al, proposed that this variants are likely to weaken or prevent binding to the ubiquitin ligase motif located in exon 6. [Voisin et al 2021]

Anomalies found in KINSSHIP syndrome are horseshoe kidney, mesomelic dysplasia, seizures, hypertrichosis, intellectual disability and pulmonary involvement as cardinal features. Bilateral fibular agenesis, feet in a marked equinovalgus position, subcortical brain atrophy and ambiguous genitalia are also reported. [Shimizu et al 2019]

We present a 2 year old female mexican patient with bilateral equino valgus position, mesomelic bone dysplasia, sacral dimple, failure to thrive, global developmental delay, camptodactyly, external genital hypoplasia and hypertrichosis. MRI showed cortical atrophy and corpus callosum hypoplasia. Normal karyotype was reported.

A next-generation sequencing exome panel (CentoXome Solo) identified an heterozygous pathogenic variant in *AFF3* gene NM_002285.2:c.691C>T (p.Pro231Ser) that was previously reported. The carrier status test of both parents confirmed a *the novo* event.

To our knowledge this is the first Mexican patient report of KINSHIPP syndrome and one of the few reported in global literature. Interestingly, has the c.691C>T that appears to be a recurrent variant in the degnon of *AFF3*. In our country the access to NGS studies is cost limited, but correct genetic counseling and dedicated patient care could be the difference in the correct diagnosis, management and counseling for the patient and their relatives.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2255. A robust NGS detection workflow for the *IKBKG* gene

Authors:

H-Y. Yen, J. Zdrodowski, I. Zelaya, A. Zhong, M. Fileto, S. Rammohan; Fulgent Genetics, Temple city, CA

Abstract Body:

IKBKG is associated with incontinentia pigmenti (IP), an X-linked (XL) dominant disorder impacting the skin, hair, teeth, nails, eyes, and CNS primarily affecting females and, rarely, males with 47,XXY or low-level mosaicism for the condition. The most distinctive features are evolving skin lesions, starting with blistering at birth and transitioning to wart-like rash, swirling hyperpigmentation, and, in some cases, linear hypopigmentation. Milder variants in *IKBKG* are associated with XL recessive immunodeficiency with or without hypohidrotic ectodermal dysplasia, typically affecting males.

IKBKG (formerly *NEMO*) encodes the regulator gamma subunit of the IKB kinase (IKK) complex. IKK-gamma, in complex, activates NF-kappaB, which protects against the apoptosis induced by tumor necrosis factor alpha, and regulates proinflammatory responses and ectodermal development.

IKBKG analysis is complicated by the presence of a highly homologous pseudogene, *IKBKGP1*, also on the X chromosome. This results in difficulties evaluating the gene by NGS as many reads suffer from multiple-mapping issues, wherein one read matches multiple locations, resulting in a lack of unique reads to call variants. As such, analysis typically includes LR-PCR for a common 11.7-kb deletion and LR-PCR and sequencing for variants in all 10 exons. In order to promote an efficient approach in the laboratory, we propose an NGS detection workflow for *IKBKG*.

We started our analysis with NGS and misalignment analysis for the common deletion, as a screen. Next, we evaluated the multi-mapping BAM file in the regions of high homology (e3-10) for variants with 25% allele fraction (AF) for germline variants in females, 50% for males, or lower AF for potential mosaic variants. Variants identified were confirmed by LR-PCR alone (common del) or LR-PCR and Sanger sequencing targeting the SNV detected. If no variants were identified, no further testing was performed. Eighty-six consecutive orders for *IKBKG* testing submitted to our laboratory were tested using the workflow (70 females). 45.3% had diagnostic results and 3.5% had VUS. Among diagnostic cases, 74.4% were the common deletion (including one mosaic) and 10 were other variants.

Given the management and surveillance options for patients, as well as prognostic and recurrence information, accurate diagnostic testing of *IKBKG* and other genes with highly homologous regions is necessary to avoid false negative and false positive results. We demonstrated a workflow to accurately evaluate the *IKBKG* gene to produce reliable results for patients while improving lab efficiency by utilizing NGS and confirmatory testing by orthogonal methods, as needed.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2256. A variant interpretation framework based on probabilistic graphical modeling that provides continuous high-resolution estimates of variant pathogenicity with improved interpretation accuracy over current discrete five-category frameworks

Authors:

R. Nussbaum, T. Manders, J. Nicoludis, A. Nampally, K. Nykamp, Y. Kobayashi, A. Colavin; Invitae, San Francisco, CA

Abstract Body:

As genetic testing expands to larger panels and exome analysis, laboratories face an increasing burden of interpreting variants while clinicians are forced to sort through expanding lists of variants of uncertain significance (VUS). Current variant interpretation systems rely on rules-based frameworks such as the ACMG/AMP guidelines (Richards et al. *Genet. Med.* 2015) and Sherlock (Nykamp et al. *Genet. Med.* 2017) that use pre-defined heuristic weighting of quantitative and qualitative data of various evidence types to support either a pathogenic or a benign interpretation. While these systems bring consistency and reproducibility to variant classifications, valuable information is lost when rules-based approaches incorporate inherently quantitative data such as population allele frequencies or in silico algorithms as discrete categories. Insufficient or conflicting evidence results in more than 50% of observed variants in disease-associated genes being classified as VUS. This leaves many patients with uncertainty about their disease risk or diagnosis and how those results should influence their care. Here we introduce a novel variant interpretation framework based on probabilistic graphical modeling and machine learning that applies domain expert-informed causal reasoning about the pathogenesis of disease. This framework has several advantages compared to current rules-based systems. First, instead of treating different types of evidence as independent and relying on manually-defined weights, it provides a compact representation of the complete joint probability distribution for the domain to assess the value of information in the context of other information. We demonstrate that this approach can substantially decrease the number of missense VUS observed at our lab by up to 25%, even when using a limited set of distinct data types. Second, the graph structure facilitates incorporation of additional types of quantifiable evidence, such as functional analyses and clinical phenotype information. Third, this approach supports reasoning about the relationship of evidence to interpretation so that the pieces of information that will most effectively increase the certainty of a variant being pathogenic or benign can be identified. Lastly, by estimating quantitative probabilities of pathogenicity instead of relying on a 5-tier classification, the framework provides a continuous high-resolution estimate of variant pathogenicity, which we believe will empower and encourage the medical community to implement more informed, personalized management guidelines that will lead to better patient care.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2257. Accurate Total and Allele-Specific Copy Number Quantification of the Cardiovascular Disease-Associated Lipoprotein (a) Kringle-IV Type-2 Tandem Repeat Using Illumina Sequencing

Authors:

J. Belyeu¹, S. Behera², X. Chen³, N. Q. H. Nguyen⁴, V. K. Menon⁵, C. M. Ballantyne⁶, C. Rodriguez⁷, R. C. Kaplan⁷, G. Metcalf⁵, B. Yu⁸, E. Boerwinkle⁸, M. A. Eberle⁹, F. Sedlazeck⁵; ¹Illumina, Inc., San Diego, CA, ²Baylor Coll. of Med., Houston, TX, ³Pacific BioSci.s, Richland, WA, ⁴Sch. of Publ. Hlth., Univ. of Texas Hlth.Sci. Ctr., Houston, TX, ⁵Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, ⁶Dept. of Med., Baylor Coll. of Med., Houston, TX, ⁷Dept. of Epidemiology and Population Hlth., Albert Einstein Coll. of Med., Bronx, NY, ⁸Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ⁹Pacific BioSci.s, San Diego, CA

Abstract Body:

Transcript length of the human LPA gene, which encodes lipoprotein (a) (Lp(a)), has been associated with increased cardiovascular disease (CVD) risk. The transcript length of LPA is primarily driven by the number of copies of a 5.5kb tandem repeat known as Kringle-IV Type-2 (KIV-2). Quantification of KIV-2 has been challenging with sequencing due to the length of the repeat and its extreme variability, with 2-60+ copies of the region in each allele and greater than 95% heterozygosity in most populations.

Here we present a computational strategy to determine KIV-2 copy number from Illumina whole-genome sequencing (WGS). The method successfully identified total copy number (summed copy number of both alleles) in all 3,202 samples from the 1000 Genomes Project. The approach utilizes SNV markers to determine the phased copy number (individual length of each allele) in ~47% of these samples. We performed an analysis of call accuracy against orthogonal long-read and optical mapping technologies in a subset of 46 samples for total copy number and 39 alleles for phased copy number calls. This analysis revealed high accuracy, with average agreements in total and allelic copy number of 98.2% and 98.6% respectively. We then analyzed KIV-2 copy number in a set of approximately 3,000 additional genomes collected in the USA from European, African, and Hispanic ancestries, with matching Lp(a) protein level measurements. In this analysis we found significant trends of higher Lp(a) levels associated with low KIV-2 copy number across all three ancestral groups.

The LPA calling method, which will be made available on Illumina's DRAGENTM Bio-IT platform*, will enhance characterization of KIV-2 copy number and enable the detection of alleles that increase CVD risk from WGS with vast utility in research applications.

*For Research Use Only. Not for use in Diagnostic Procedures.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2258. Advancing Biobanking in the Genomics Era: Development of Molecular Omics Approaches for Interrogating the Diversity of Population Genetics for Early and Late-Stage Human Disease Prevention

Authors:

M. Sheldon, S. Nahas; Sampled, Piscataway, NJ

Abstract Body:

Precision medicine has demonstrated improvements in the quality of life for people who are at risk for or present with a given disease and its complications. Development of targeted, molecular-based assays and companion treatments that provide early detection and inform effective therapeutic approaches for a number of serious conditions often relies on access to large numbers of high quality biospecimens and associated clinical data. For these reasons, biobanks are poised to play a role in supporting many aspects of clinical research, including clinical trial management, discovery and development of assays and therapies. At Sampled, we taken the next step in the evolution of the modern biobank. The Sampled model is that of a SMART lab, providing a comprehensive suite of sample Storage, Management, Analysis, Research and Transportation services to clients large and small in a range of academic and commercial sectors. In this presentation, we will consider a number of initiatives at Sampled to develop tools for the advancement of the diagnosis and treatment of human genetic disorders. One such example is polygenic risk score (PRS) testing. PRS tests for diseases like Alzheimer's, breast cancer, and even severe COVID have been developed and brought to market by Sampled in collaboration with commercial partners. These tests are designed using Genome Wide Association Study (GWAS) data to identify candidate risk alleles and associated clinical data that can be combined to give a risk score. This score is then be combined with that of the population within the patient's demographics, to assess potential risk for a particular disorder. For example, the Alzheimer's PRS Test can aid in stratification of clinical trials for new treatments or facilitate early detection of subjects at high risk. It includes over 112,000 loci of interest, including the "VIP" APOE loci. Initial applications will be for physicians assessing new patients, with cognitive complaints, to assess their future risk of cognitive decline due to Alzheimer's disease. It can also provide greater insight into the risk of cognitive decline in existing patients. Additionally, the Alzheimer's Risk Test will be used in clinical study recruitment strategies to identify patients most suitable to enter trials of investigational Alzheimer's drugs, including those patients with mild cognitive decline who are most at risk of subsequent decline within the timescale of the study. In addition to working with partners to clinically validate these tests on the ideal platform, Sampled offers the ability to scale up for high throughput testing in its high-complexity CLIA laboratory.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2259. Analysis of mRNA and protein expression levels of MECP2 in missense mutant iPSC and lymphoblast lines against isogenic controls

Authors:

T. Thiruvallur Madanagopal¹, S. Pastore¹, T. Muhammad¹, A. Mikhailov¹, J. Vincent²; ¹Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, ²Ctr. for addiction and mental Hlth., Toronto, ON, Canada

Abstract Body:

Background Mutations in MECP2 (methyl CpG binding protein 2) have found to cause Rett syndrome (RTT), as well as autism spectrum disorder (ASD), schizophrenia, and X-linked intellectual disability (XLID). About 33% of all RTT patients have missense mutations, and mainly in the methyl-CpG-binding domain (MBD) that binds to DNA and forms chromatin associated protein complex. Some of these mutations interfere with the binding and some others disrupt the clustering of chromatin. A similar proportion of RTT patients have nonsense mutations that truncate the MECP2 protein, leading to protein degradation resulting in insufficient levels in the cell. Studies have suggested this degradation might also be true for missense mutations. Hence understanding changes in MECP2 expression in mRNA and protein would shed light on better understanding of the role of MECP2 mRNA and protein degradation towards the clinical manifestations. **Methods** MECP2 mutant and isogenic iPSC (induced pluripotent stem cells) and lymphoblasts were provided by the Rett Syndrome Research Fund or purchased from Coriell Institute for Medical Research, USA. RNA and protein were isolated from these cells. mRNA levels of MECP2 was analyzed using RT-qPCR from complimentary DNA (cDNA) synthesized from RNA. Protein levels were quantified using ELISA and western blot. The mRNA expression levels of MECP2 for the mutants iPSC and lymphoblasts lines were analysed against isogenic controls. **Results** While mRNA levels were not significantly affected, some mutants showed significantly less protein expression compared to its isogenic controls (R225X, R133C). However, other mutants like R306C and R106W did not show a significantly change in protein expression compared to its isogenic control. **Conclusion** The knowledge on the levels of MECP2 protein and mRNA expression could help develop approaches to recover the mutant MECP2 protein and stabilizing it. This would be beneficial to develop personalized therapeutic approaches for MECP2-related clinical phenotypes.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2260. Anthropometric parameters in males with chromosome 9p microdeletion syndrome in the first years of life: preliminary evidence.

Authors:

S. Pajouhanfar, R. Starosta, V. Tate, Y-C. Hou, F. Cole, T. Turner, P. Dickson; Washington Univ. Sch. of Med., Saint Louis, MO

Abstract Body:

Introduction: Chromosome 9p microdeletion syndrome (“9p- syndrome”) is a rare syndrome with a spectrum of expressivity including dysmorphic features, intellectual disability, hypotonia, psychiatric illnesses, metopic craniosynostosis, male-to-female sex reversal, neonatal hypoglycemia, congenital heart disease, and ocular abnormalities. To date, there has been no investigation of the anthropometric data in 9p- patients, and in practice their parameters are plotted in typical growth charts such as the ones published by the World Health Organization (WHO) or the Centers for Disease Control and Prevention (CDC), which have not been validated for this population. Objective: A retrospective study with chart reviews of anthropometric parameters was performed to compare growth curves of males with 9p- syndrome to the general population. Methods: Anthropometric data from clinical visits until age 3 years were collected from all participants. Data were compared to the WHO growth charts for weight, length, and head circumference at 2, 4, 6, 9, 12, 15, 18, and 24 months of age, and to the CDC growth charts for weight, height, and head circumference at 36 months of age. Due to the low sample size, only descriptive statistics were performed. Written informed consent was obtained from the parents or legal guardians of all participants. This study was approved by the Washington University in Saint Louis IRB. Results: Eleven patients (male = 7) with a documented diagnosis of 9p- by karyotype or chromosomal microarray were included in this study. Due to the low female sample size (n = 4) in the study, only males (were analyzed. Mean percentiles for height at 2, 4, 6, 9, 12, 15, 18, 24, and 36 months were 66%, 83%, 86%, 78%, 88%, 81%, 80%, 84%, and 75%, respectively. Mean percentiles for weight at 2, 4, 6, 9, 12, 15, 18, 24, and 36 months were 70%, 92%, 90%, 68%, 68%, 71%, 67%, 68%, and 84%, respectively. Mean percentiles for head circumference at 2, 4, 6, 9, 12, 15, 18, 24, and 36 months were 28%, 71%, 83%, 70%, 67%, 57%, 73%, 73%, respectively. Discussion/Conclusion: Despite the small sample size, on a visual inspection of the data there is an apparent trend toward higher percentiles for height during the first 3 years of life in males with 9p- syndrome compared to normative data. This trend was accompanied by weight during the first year of life, with a relative decrease in weight percentiles afterward. There was no perceived deviation from normality in head circumference values, which may have been confounded by trigonocephaly but may also represent a small degree of relative microcephaly. We are currently enrolling more individuals with a diagnosis of 9p- syndrome to expand this analysis.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2261*. B-allele frequency based approach to detecting absence of heterozygosity using optical genome mapping.

Authors:

A. Raksi, W. A. Sherman, N. Miller, A. R. Hastie, S. Shams; Bionano Genomics, San Diego, CA

Abstract Body:

Introduction: Copy-neutral loss of heterozygosity can be caused by uniparental disomy, associated with meiotic errors resulting in developmental diseases and cancer, or be a result of identity by descent (IBD), when the parents are close relatives or from a bottleneck population. These events have traditionally been identified using SNP microarray and microsatellite analysis. Previously, we described a method to detect the absence or loss of heterozygosity (AOH/LOH) using high-resolution optical genome mapping (OGM) results from the Bionano Genomics Saphyr platform, based on information about structural variant (SV) call zygosity. Here we describe a method for AOH analysis based on the B-allele frequency (BAF) observed at label sites overlapping with known SNP positions, which enables the detection of AOH events at low allelic fractions. **Methods:** DNA was processed on the Bionano Genomics Saphyr instrument and analyzed using the Bionano Solve pipeline. The labels used in OGM attach to a specific nucleotide sequence motif in the genome. SNPs overlapping the sequence will cause a lack of fluorescence at that point. The B-allele frequency is calculated as the ratio of molecules with missing labels to the total number of molecules aligned to the position. Label sites were filtered to those overlapping known SNPs with a minor allele frequency greater than 5% in the population. Data from 180 control samples were used to identify labels for which the BAF values clustered into three well-formed clusters corresponding to the AA, AB, and BB alleles, indicating that heterozygous and homozygous SNPs could be differentiated. Next, BAF values were normalized for the query sample at each SNP position. Finally, AOH regions were called using the NxClinical software's SNP-FASST3 algorithm for segmenting and calling events. **Results:** By filtering data for labels overlapping known SNPs and those that clustered well, 3867 informative label sites were identified using the DLE1 labeling enzyme. The method was validated in a small cohort of constitutional and cancer samples with known AOH/LOH events. We were able to detect a wide range of events ranging from small 20Mb events to as large a whole chromosome. We also demonstrated and validated a 55Mb low level mosaic detection of a CN-AOH at 16% allele fraction. **Conclusions:** Our results show that it is possible to detect AOH/LOH regions at low allele frequency using optical genome mapping alone. This expands the utility of the method to analyzing tumor samples. The software will be made available as part of NxClinical v7.0.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2262. Beyond NGS: RNA-based functional evaluation improves ACMG classification of VUS with potential Loss of function

Authors:

J-L. Blouin^{1,2}, T. RIO FRIO¹, A. VANNIER¹, M. ABRAMOWICZ^{1,2}; ¹Univ. Hosp. of Geneva, Geneva, Switzerland, ²Univ. of Geneva-Sch. of Med., Geneva, Switzerland

Abstract Body:

NGS in clinical practice allows screening an ever increasing number of genes that has improved the portion of elucidated cases, but also creates a dramatic increase of variants of unknown significance. Current variant filtering strategies and interpretation guidelines by the ACMG focus on amino-acid level identify but filter out other non-coding or synonymous variants that may however have a biological and clinical significance. We have implemented an analytical pipeline to assess the variants-of-uncertain-significance (VUS) predicted to impact gene transcript with the aim to decipher their pathogenicity. After selection of pre-filtered variants by an interdisciplinary team, molecular investigations based on techniques of targeted transcript regions that are fast and affordable were performed on RNA extracted from patients blood. Besides confirmation of predicted impact on splicing, sequence or structure of some variants, a number of unexpected effects were also observed in a number of other variants. For example, a variant in an intronic canonical splicing signal of DNAH11 predicted to induce in-frame exon skipping, and, or partial intronic retention ultimately leads to premature polyadenylation of mRNA. Also, a variant in an intronic cis- regulatory splicing element of NEB produces an unexpected broad range of alterations of the sequence, the structure and the balance of alternative isoforms. We also demonstrated the utility of RNA analysis in a prenatal diagnosis in RNA from fetal amniocentesis in the context of ultrasonography abnormalities for a IVS+5 variant of FBN1 that was upgrades to class 5 in the fetus. We demonstrated that such targeted RNA-based analytical approach had the ability to clarify a number of DNA-based interpretation of variants. It was far-more cost-effective than total RNAseq that needs complex analytical process that would not be necessary in a number of such situation where VUS are identified and could explain the phenotype. All variants, within various genes, analyzed through this pipeline so far could thus be re-classified, either upward (pathogenic) or downward (benign) following ACMG guidelines. In addition we found that all the genes assessed until now could be assessed on RNA from blood although some of them were not necessary predicted to be expressed in such easily accessible tissue according to GTEx.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2263. Challenges of characterisation of medically relevant tandem repeats in whole genome sequencing data

Authors:

I. Lojova^{1,2}, M. Kucharik^{3,4}, Z. Pös^{1,4,3}, A. Zatkova¹, E. Tarova^{1,5}, J. Styk^{4,3,6}, J. Budis^{4,3,7}, L. Kadasi^{1,2}, T. Szemes^{4,2,3}, J. Radvanszky^{1,4,2}; ¹Inst. of Clinical and Translational Res., BioMed. Res. Ctr. of the Slovak Academy of Sci., Bratislava, Slovakia, ²Dept. of Molecular Biology, Faculty of Natural Sci., Comenius Univ., Bratislava, Slovakia, ³Geneton Ltd., Bratislava, Slovakia, ⁴Comenius Univ. Sci. Park, Bratislava, Slovakia, ⁵Faculty of Ed., J. Selye Univ., Dept. of Biology, Komarno, Slovakia, ⁶Inst. of Med. Biology, Genetics and Clinical Genetics, Faculty of Med., Comenius Univ., Bratislava, Slovakia, ⁷Slovak Ctr. of Scientific and Technical Information, Bratislava, Slovakia

Abstract Body:

Following the introduction of massively parallel sequencing and dedicated bioinformatic tools, genotyping of tandem repeat motifs (TRs) from these datasets are becoming commonplace too. Effective and reliable TRs characterisation for medical reasons, for example of those connected to repeat expansion disorders (REDs), however, still poses several challenges, specifically when considering the vast variability and complexity of TRs. We aimed, therefore, to characterize some of the potentially relevant TR detection challenges. For this we used whole-genome sequencing (WGS) data (BGI-DNB-SEQ; Paired-End 150bp; PCR-free library preparation) generated for 52 individuals. Altogether, over 50 manually selected and described (according to the actual HGVS nomenclature) medically relevant TR loci were genotyped and characterized from each WGS using the modified version of our previously published tool Dante (Budis et al., 2019). The data set consisted of patients having three different types of REDs (10 DM1, 7 DM2 and 5 HD) and 30 controls. The main challenges identified were: 1) Correct dissection of clinically relevant parts of TR loci and their flanking sequences, especially if they are parts of complex motifs (connected to the use of unified allele nomenclature; e.g.HGVS); 2) Correct allele identification and sizing, affected by the: i) achieved read depth and usable reads; ii) possible complexity of motifs (expected and unexpected complexity, including unexpected sequence interruptions or loss of normally present interruptions); and iii) stutter effects (largely eliminated by PCR-free approaches); 3) Sizing limits, i.e. the maximum allele length determinable from the spanning reads (connected to read-length/motif-length characteristics); 4) Identification of the presence of alleles exceeding the detection limit (expansion detection exploiting partial reads); 5) Correct allele phasing when considering complex motifs consisting of several adjacent simple motifs. Results of our work proved the potential of short-read WGS in the field of TRs characterization, however, with certain limitations, which require further studies, modifications, specific quality control metrics, or at least which should be kept in mind when evaluating TRs in individual patients. The presented work meets the relevant ethical standards and informed written consents were obtained. Financial support: Slovak Research and Development Agency (APVV-18-0319), Scientific Grant Agency (VEGA_2/0167/20, VEGA_2/0040/20) and by the OP Integrated Infrastructure co-financed by the European Regional Development Fund (ITMS: 313011V578, 313011W428).

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2264. Child with Rhabdomyosarcoma and unusual features: suspected Rasopathy.

Authors:

S. Charaya¹, I. Panigrahi², A. Tyagi¹, S. Gupta¹; ¹Post graduate Inst. of Med. education and research, Chandigarh, India, ²PGIMER, Chandigarh, India

Abstract Body:

Case description: 8-month-old female child with intermittent fever episodes for 20 days and gradually progressive left cheek swelling. On examination had failure to thrive, microcephaly, right facial nerve palsy (upper motor neuron) and dysplastic ears. Local examination showed 8cmx8cmx6cm swelling over face, firm in consistency with few nodular surfaces. Skin over swelling- tense, red with raised temperature. Skin examination showed café au lait macule over trunk and lower limb, melanocytic nevi, nevus spilus, Mongolian spots over back, Inflammatory linear verrucas epidermal nevus, scarring alopecia. FNAC of swelling showed predominantly scattered tumor cells with few clusters of many bizarre multinucleated tumor cells s/o malignant neoplasm. Immunohistochemistry from FNAC cells showed desmin, myogenin and MyoD1 positive Rhabdomyosarcoma. MRI Brain showed exophytic heterogeneously enhancing lesion in left masticator space. Exome sequencing studies were performed keeping a possibility of Rasopathy or Genodermatoses.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2265. Chromosomal and molecular investigations of Senegalese children bearing disorders of sex development.

Authors:

H. Younoussa^{1,2}, M. SOUMBOUNDOU³, R. DIAGNE⁴, R. M'kacher⁵, M. GADJI¹; ¹Service of Biological Hematology & Oncology-Hematology (BHOH), Faculty of Med., Pharmacy and Odontostomatology, Cheikh Anta Diop Univ. of Dakar (UCAD), Natl. Ctr. of Blood Transfusion (NCBT), Dakar, BP5002 PtE, Senegal, Dakar, Senegal, ²Cell-Environment, Paris, France, ³Med. Diagnostic Lab., Diarniadjou Hosp., UFR-Santé de Thiès, Iba Der THIAM Univ. of Thiès, Senegal, Thiès, Senegal, ⁴Med. Diagnostic Lab., Diarniadjou Hosp., UFR-Santé de Thiès, Iba Der THIAM Univ. of Thiès, Senegal, Dakar, Senegal, ⁵Cell Environment, Paris, France

Abstract Body:

Disorders of sex development (DSD) are among the most common congenital malformation syndromes in Africa. They are defined as conditions associated with atypical chromosomal, gonadal or anatomical sex development. The incidence and prevalence of these anomalies remain very high in Africa despite of diagnostic difficulties compared to the rest of the world. Indeed, diagnosis and management of these abnormalities in this African continent remain very limited. In order to set up and improve postnatal diagnosis of DSD for a better patient clinical management, we aimed to investigate the epidemiological, genetic and molecular profile of children with DSD in Senegal. A retrospective and prospective study in a multicenter survey of DSD in Senegal was conducted from 2014 to 2021 with 91 included DSD patients. The survey data were entered into Excel and exported to R software version 4.1.3 for statistical analysis. Chromosomal investigations were performed using constitutional cytogenetics (G-banding) and molecular FISH techniques for the search of the SRY gene as well as telomere and centromere labeling followed by M-FISH for the search of other chromosomal aberrations that would be the cause of these abnormalities. The mean age of the patients was 33.16 ± 48.97 with extremes from 2 days to 178 months. The median age was 12 months. The percentage of parental consanguinity was 54%. Karyotype and FISH analyses were performed in the majority of the population. Among the 91 DSD patients, 48.88% had 46,XY karyotypic profile, 31.11% with 46,XX karyotype, 1.1% with 46,XX/45,X mosaic karyotype, 1.1% of 46,XY/45,X karyotype and 17.77% did not have karyotype. The most frequent etiological causes of DSD were congenital adrenal hyperplasia, androgen insensitivity syndrome in the studied population. A few cases of ovotestis and mosaicism were also observed. These preliminary results, allowed us to continue further molecular and genetic investigations using whole genome sequencing in order to fully characterize DSD-patients. This study will pave the way to define the gonadal mechanisms that may be involved in the DSD Senegalese population.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2266*. Chromosome conformation capture to detect structural variation in medical diagnostics

Authors:

S. Holwerda¹, P. Krijger², M. Versteegen², H. Ploos van Amstel¹, W. de Laat², M. van Kempen¹, K. van Gassen¹; ¹Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, ²Hubrecht Inst., Utrecht, Netherlands

Abstract Body:

Genomic aberrations, including structural variations (SVs), are a major cause of human genetic diseases. Routine detection of SVs in clinical diagnostics mostly relies on standard cytogenetic techniques like karyotyping and fluorescent in-situ hybridization (FISH). The current standard-of-care cytogenetic techniques have proven to be highly successful but they are labor intensive, require highly specialized personnel, provide low resolution data and are low throughput. Here, we perform a study to implement SV detection by a chromosome conformation capture technique, HiC, into our diagnostic pipeline. HiC has been used to detect SVs and copy number variation (CNVs) in a genome wide (unbiased) manner, and could be a high throughput / high resolution alternative to SV detection by karyotyping. Its compatibility with NGS, sample requirements and DNA isolation ease allows implementation in current diagnostic flows. Despite some advantages, certain SVs have not yet been reported in literature, and implementation in daily medical diagnostics has not been reported. Our HiC-study initially focusses on the detection of several unreported SVs, such as: Robertsonian translocations and ring-chromosomes. To this purpose we use patient derived cell lines from the Coriell biobank to test our hypothesis: HiC can detect SVs across centromeres. Furthermore we strive to implement HiC with our current diagnostic exome capture, potentially providing SNV and SV calls in one test. Towards a 'one test fits (almost) all', as a bonus, chromosome conformation capture data will provide structural data, such as topologically associated domain (TAD) boundary information, that could explain disease mechanisms behind structural aberration of our genome.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2267. Clarifying genetic disease with long range sequencing: Tipping the balance for structural variant detection in SMAD3.

Authors:

S. Safgren¹, J. Morales-Rosado², R. Aleff¹, M. Bockol¹, J. Smadbeck¹, N. Hoppman¹, Z. Stephens¹, M. Dehankar¹, S. Phillips³, S. Mantia³, E. Wieben¹, K. Wierenga³, E. Klee¹; ¹Mayo Clinic, Rochester, MN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Mayo Clinic, Jacksonville, FL

Abstract Body:

A maternally inherited duplication of uncertain significance in the last 3 exons (7-9) of *SMAD3* was identified by exome sequencing in a patient with a previous clinical diagnosis of Marfan syndrome. No additional relevant variants were identified. While it was clear that this VUS had pathogenic potential, the molecular mechanism was unclear. Exome and genome sequencing accurately identify most single nucleotide variants (SNV), but balanced copy number variants (CNV) or translocations are routinely missed or incorrectly characterized. Mate-pair sequencing was used to determine the orientation and location of the duplication. Interestingly, the Mate-pair results revealed a larger duplication of 4 exons (6-9), as well as a deletion of exon 6. However, it could not be determined if this complex structural variant (SV) was in cis or in trans. Long-read sequencing using PacBio and Oxford Nanopore was used to further characterize this event. This analysis confirmed a deleterious deletion of exon 6 in cis with a non-deleterious tandem duplication of exon 6 through 9. The deletion breakpoints in introns 5 and 6 were manually defined from the long-read data because of SINE repeat sequences in this region. The duplication event was located beyond the 3'UTR of the gene and has the same intron 5 breakpoint observed in the deletion event. In this patient, long-read sequencing aided in understanding the deleterious in-frame loss of exon 6 that was obscured in the exome sequencing by the in cis duplication of exons 6-9 (a balanced duplication/deletion of exon 6 where n=2). The exome sequencing report from the affected mother indicated the same VUS and provided evidence this complex variant was inherited on a single allele. The impact for the patient and family includes a refined diagnosis from Marfan to Loeys-Dietz syndrome - which can have a more severe phenotype, with more extensive surveillance recommendations. This illustrates a unique molecular pathogenic mechanism, which resulted in a certified molecular diagnosis of Loeys-Dietz syndrome, and allowed for molecularly informed care recommendations.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2268. Clinical description of patients with CNV in the 15q11.2 region: What do we know?

Authors:

H. Chaparro Solano¹, M. Arcos-Burgos², L. Rodriguez-Salazar³, G. Ardila Patiño³, P. A. Rueda³, C. Estrada Serrato⁴, O. Londoño⁴, Y. D. Carrillo Rincón³, J. J. LOPEZ RIVERA⁵; ¹Clínica Colsanitas, Bogota, Colombia, ²Inst. de Investigaciones Médicas, Univ. de Antioquia, Medellín, Colombia, ³Laboratorio Especializado, Clínica Univ.ria Colombia, Bogota, Colombia, ⁴Clínica Univ.ria Colombia, Bogota, Colombia, ⁵Clinica Univ.ria Colombia, Bogota, Colombia

Abstract Body:

Introduction: Copy number variants (CNVs) between breakpoints 1 and 2 (BP1 and BP2) of 15q11.2 region, and involving *CYFIP1*, *NIPA2*, *TUBGCP5* and *NIPAI* genes, have been associated with diverse phenotypes and incomplete penetrance. While microdeletions that meet the described characteristics have been associated with Burnside-Buttler syndrome, the absence of phenotype or the presence of variable expressivity in patients with the same, or very similar CNV characteristics, has led to the fact that their interpretation, reporting and follow-up become a challenge. A similar scenario occurs with microduplications contained within this same region. The aim of this study is to describe the clinical characteristics of Colombian patients with this type of CNV, in order to achieve an adequate characterization of the conditions associated with them and to contribute to the delineation of the phenotype and its understanding in prognostic terms.

Methodology: A retrospective cross-sectional study was performed. Clinical and paraclinical variables were collected from Colombian patients (n=26) with deletions or duplications between BP1 and BP2 of 15q11.2 region, identified thru comparative genomic hybridization (CGH). All samples were processed, including its classification, between 2017 and 2021, by a single laboratory. The current study was endorsed by the Research Ethics Committee of Fundación Universitaria Sanitas.

Results: 32% of the CGH were ordered for autism spectrum disorder and 28% for neurodevelopmental delays. Other causes were cardiopathies, immune disorders, obesity, epilepsy and dysmorphism. Sixty-five percent (n=17) of the patients were male and 35% (n=5) were female. The mean age of the patients was 8.54 years (min 2 - max 24 SD 5.41). Among the prenatal history, only 19% (n=5) presented prematurity. In relation to clinical history, 65% (n = 17) registered behavioral disorder, 77% (n = 20) language delay and 58% (n = 15) motor neurodevelopmental delay. Heart disease was present in 23% (n = 6) of the participants and central nervous system malformations in 8% (n = 2). Fifty-four percent of the CNVs identified corresponded to gains and 46% to deletions. The average size of CNVs was 0.51 Mb (min 0.31 - max 0.855 SD 0.19).

Discussion and conclusions: Results are consistent with other scientific evidence, and demonstrate that there is still significant genetic and clinical heterogeneity associated with CNVs in the 15q11.2 region, making their analysis and classification difficult. Further studies, with larger sample sizes, are required to deepen the phenotypes related to this type of variants in order to achieve a better understanding of their pathogenicity.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2269*. Clinical long-read genome sequencing: Analytical performance of germline small variant detection using HiFi genome sequencing

Authors:

N. Hammond¹, L. Liao¹, P. Tong¹, Z. Ng¹, C. Ho¹, Y. Yang^{2,1}, S. A. Scott^{2,1}; ¹Clinical Genomics Lab., Stanford Hlth.Care, Palo Alto, CA, ²Dept. of Pathology, Stanford Univ., Stanford, CA

Abstract Body:

Long-read sequencing has the ability to detect structural variants, interrogate homologous regions, and phase variants; however, short-read sequencing is more commonly implemented for clinical testing based on its higher throughput and accurate detection of single nucleotide variants (SNV) and insertions/deletions (indel). Given the recent advances in long-read single molecule real-time HiFi sequencing chemistry and variant calling practices, we analytically validated HiFi genome sequencing to assess its performance for clinical testing. HiFi genome sequencing (average 32X) was performed on DNA from six GIAB cell lines, and results were compared to short-read genome sequencing data (average 52X). HiFi genome sequencing small variant calling was accomplished using pbmm2 v1.7.0, DeepVariant v1.3.0 and WhatsHap v1.4, and short-read sequencing variant calling using DRAGEN v10.3.4. Small variant calls were benchmarked against the GIAB v4.2.1 truth set using hap.py v0.3.15, including comprehensive interrogation across genomic regions as defined by GIAB v3.0 stratifications. HiFi genome sequencing outperformed short-read genome sequencing on overall SNV accuracy (99.9% recall, 99.9% precision vs. 99.4% recall, 99.8% precision); however, short-read genome sequencing outperformed on overall indel accuracy (99.6% recall, 99.7% precision vs. 99.3% recall, 99.2% precision). Importantly, HiFi genome sequencing significantly outperformed short-read genome sequencing in segmental duplication regions for both SNVs (98.2% recall, 99.0% precision vs. 91.6% recall, 95.6% precision) and indels (98.7% recall, 99.1% precision vs. 92.8% recall, 95.6% precision), and in low mappability regions for both SNVs (98.8% recall, 99.5% precision vs. 90.7% recall, 97.2% precision) and indels (98.6% recall, 99.3% precision vs. 88.3% recall, 95.7% precision). HiFi genome sequencing reproducibility was assessed by measuring non-reference genotype concordance between GIAB data from independent sources, which was >99.9% for SNVs/indels in non-difficult regions. These results indicate that HiFi genome sequencing is highly accurate and robust, outperforming short-read sequencing across the majority of GIAB genomic regions, which supports the use of this technology for clinical diagnostic testing.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2270. Clinical utility of parental testing for the reclassification of likely pathogenic and uncertain copy number and sequence variants in a pediatric neurodevelopmental disorders cohort.

Authors:

C. Bilancia, A. Baxter, M. Martin, A. Ortega; Bionano Genomics, San Diego, CA

Abstract Body:

Genetics professionals frequently recommend parental testing when a child has a variant of uncertain significance identified on genetic testing, with the belief that knowing parental inheritance will impact understanding of the clinical significance. The clinical utility of parental inheritance for variant classification in a pediatric neurodevelopmental disorder population is complicated by variability expressivity, reduced penetrance, and limited parental clinical history. This analysis describes the experience of parental inheritance testing by a commercial lab in a pediatric population for variants identified from clinical chromosomal microarray analysis or whole exome sequencing testing. For cases where parental testing was completed following identification of an uncertain or likely pathogenic copy number variant in the child, 74% of cases had some available clinical or phenotype information for a parent. Of those with clinical information available, 43% had an overlapping feature similar to the child's. Due to complications in assessing parental phenotype and challenges with copy number variants often involving multiple genes, less than 2% of cases received a reclassification following parental testing. In comparison, parental testing for sequence variants resulted in 33% of cases receiving a reclassification, highlighting the importance of trio analysis for whole exome sequencing. This conclusion is further supported by our retrospective analysis showing that 67% of proband-only sequencing cases had a variant of uncertain significance compared to 58% of cases when one or both parents were available for analysis. Examples of chromosomal microarray cases where parental inheritance was informative included novel copy number variants involving 7q36.1 and 19q13.33. For whole exome sequencing, parental inheritance enabled reclassification by identifying *de novo* variants and phasing multiple variants in a gene linked to a recessively inherited disorder. The information presented from this analysis provides guidance for education and counseling about the benefit of parental testing for reclassification of variants associated with neurodevelopmental disorders.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2271. Clinical validation of end-to-end whole genome sequencing and interpretation applied to both panel-based and diagnostic genome testing.

Authors:

A. Salman¹, K. A. Lafferty¹, D. M. Toledo¹, K. C. Lewis¹, E. B. Malolepsza¹, M. V. Harden¹, K. L. Larkin¹, V. Mistry², E. Frise², W. He², J. McCarthy², S. B. Gabriel¹, M. G. Reese², H. L. Rehm^{3,1}, N. J. Lennon¹; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²Fabric Genomics, Oakland, CA, ³Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

The Clinical Research Sequencing Platform (CRSP) at the Broad Institute is a CLIA-licensed and CAP-accredited clinical laboratory that to date has provided clinical whole genome sequencing to partner laboratories and clinical research programs. Currently, customers are provided with raw data (CRAM) or secondary analysis outputs (VCF) for use in their external pipelines. Here, we describe a plan to support and validate panel-based or indication-based interpretation, enabling launch of an end-to-end whole genome service with genomic interpretation.

To enable high-throughput interpretation services, CRSP intends to make use of the Fabric Genomics interpretation platform. Panel-based testing is facilitated by Fabric's AI Classification Engine (ACE), which assists in the auto-application of certain types of evidence according to ACMG/AMP criteria, highlighting potentially pathogenic variants for further review by a variant analyst. The ACMG version 3.1 secondary findings list of 73 genes was used for ACE validation. Indication-based testing is facilitated by Fabric's GEM algorithm, which prioritizes all variants in a genome by their likelihood of causing a given phenotype with subsequent review by a variant analyst and signing geneticist.

To validate this service, we have assembled a cohort of both clinical and reference samples in which variants of interest exist. The variant spectrum is represented by the inclusion of single nucleotide variations (SNVs), small insertions and deletions (indels), and larger copy number and structural variants. Between 20-25 samples of each variant type were identified, using unique samples for the ACE and GEM algorithms, resulting in ~136 validation samples.

We present here the results of the validation study. The validation assessed whether the variant of interest was detected and prioritized for review as expected and where the variant ranked within the GEM algorithm. We also assessed the time for review for each case to understand analyst resources required to support the Fabric Genomics platform. Samples in which the variant of interest was appropriately prioritized for review were considered to have passed this stage of the validation. We also reviewed the automated and manual variant analysis steps for each classified variant to assess accuracy of the automated portion of the classification, as well as time for manual variant review. The algorithm was also assessed for repeatability and the platform was assessed for security. The full outcome of this validation will be presented.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2272. Clinical whole-genome sequencing of oral samples and microbial contamination

Authors:

N. Ameziane¹, A. Kumar¹, V. Skrahina¹, A. Rolfs², G. Oprea¹; ¹Arcensus Diagnostics, Rostock, Germany, ²Albrecht-Kossel-Inst., Univ. of Rostock, Rostock, Germany

Abstract Body:

The oral cavity is an excellent source of biological sampling for clinical genetics and genomics. This is because oral sample collection is quick, easy-to-access, and non-invasive, compared to blood or other tissue collection methods. We have set-up a whole genome sequencing-based clinical genetic testing approach for suspected patients with hereditary disease. During routine molecular diagnostic analysis, some samples suggested the presence of non-human DNA sequences at varying percentages.

We explored the fraction of non-human reads sequenced from buccal swabs and saliva, type of microbial genomes from which they originated, and impact on molecular classification of the subjects. Read sequences not mapped to human reference genome, were aligned to completed reference microbial reference sequences from the National Center for Biotechnology Information's (NCBI) RefSeq database using KRAKEN v2.

A total of 973 samples (860 buccal swabs and 113 saliva samples) were analyzed. For 97% buccal swabs and 74% saliva samples, more than 5% of the sequenced reads were non-human. The most common contaminant was caused by *Streptococcus mitis* and *Rothia mucilaginosa*, which were identified in 288, and 35 of the buccal swabs and in 15 (5,2%) and 5 (14.3%) of the saliva samples, respectively.

Importantly, by comparing contaminated to non-contaminated samples, the presence of non-human DNA contamination did not demonstrate an impact on the diagnostic yield.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2273. Comparing the analytical performance of exome sequencing and traditional panel testing in a cancer population

Authors:

E. Reble¹, J. Sam¹, R. Kodida¹, M. Clausen¹, S. Shickh^{2,1}, C. Mighton^{2,1}, J-M. Capo-Chichi³, E. Greenfeld^{4,2,5,6}, A. Noor^{4,2,5,6}, R. H. Kim^{4,2,3}, J. Lerner-Ellis^{4,2,5,6}, Y. Bombard^{1,2,7}; ¹St. Michael's Hosp., Toronto, ON, Canada, ²Univ. of Toronto, Toronto, ON, Canada, ³Univ. Hlth.Network, Toronto, ON, Canada, ⁴Mount Sinai Hosp., Toronto, ON, Canada, ⁵Dept. of Pathology and Lab. Med., Sinai Hlth., Toronto, ON, Canada, ⁶Lunenfeld-Tanenbaum Res. Inst., Sinai Hlth., Toronto, ON, Canada, ⁷Ontario Inst. for Cancer Res., Toronto, ON, Canada

Abstract Body:

Background: The use of genomic sequencing (GS), including exome and genome sequencing, is increasingly being used as a diagnostic tool over traditional panel testing. With the ability to assess many more genes than traditional panel testing, GS can test for the same well-known genes as traditional panel testing while also testing for more preliminary evidence genes, and may act as an important health tool that can be reanalyzed as new genes and evidence comes to light. However, little is known as to whether GS is a valid and useful diagnostic tool over traditional panel testing. The analytical performance of GS is an important measure of validity that must be determined to inform health technology assessment of GS. **Aim:** To assess the analytical performance of exome sequencing (ES) to detect small single nucleotide variants (SNVs), insertions, and deletions compared to traditional panel testing in a cancer population. **Methods:** We assessed the analytical performance of ES in 277 individuals with a history of cancer or polyposis who received both traditional panel testing and ES. Cancer panels included BRCA1/2 testing (2 genes), Lynch syndrome testing (5 genes), cancer-specific (ex. breast, renal or colon) panels (14-24 genes), or expanded cancer panels (30-70+ genes). For all individuals (n=277), the analytical performance of ES was determined by assessing the concordance rate for all potentially clinically relevant sequence variants previously reported in the cancer panel test was evaluated. For a subset of individuals (n=75) with additional variant data, the sensitivity and specificity of all sequence variants (including likely benign/benign and high frequency variants) identified by ES and the panel in overlapping regions were evaluated. **Results:** 102 variants (VUS or P/LP variants in the region of interest) were identified by cancer panels in the 277 individuals; all 102 of these variants were also detected by ES, resulting in a concordance rate of 100%. For the subset of individuals (n=75) with additional variant data, preliminary analysis in 45 individuals demonstrated high analytical performance of ES compared to panel testing (all variants: sensitivity = 99.69%, specificity = 99.99%; substitutions: sensitivity = 99.69%, specificity = 99.99%; small insertions and deletions: sensitivity = 100%, specificity = 100%). **Conclusions:** Based on preliminary analysis, we have demonstrated high analytical performance of ES compared to traditional panel testing in a cancer population. This is a critical step to allow for ES and GS to be integrated effectively into clinical care as an important diagnostic tool.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2274. Comparison of chromosomal inversions in three different datasets

Authors:

T. Bozkurt¹, M. G. Mehaffey², U. Sezerman¹, C. M. Carvalho^{2,3}, Z. Coban-Akdemir^{4,3}; ¹Acibadem Mehmet Ali Aydinlar Univ., Istanbul, Turkey, ²Pacific Northwest Res. Inst., Seattle, WA, ³Baylor Coll. of Med., Houston, TX, ⁴Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract Body:

Introduction: Inversions (INVs) are underexplored copy-neutral structural variations that involve a change in orientation of a DNA segment often generated by non-allelic homologous recombination. The role of INVs has been well-established in primate evolution and speciation. However, current knowledge of INVs' functional effects involving their role in human disease is still limited. This is mostly due to the fact that it is challenging to identify INVs since they i. present a high rate of false-positive and false-negative errors, ii. are challenging to resolve given their balanced nature, and iii. have breakpoints lying in complex repeat regions. Next-generation sequencing technologies including short-read sequencing, long-read sequencing, Strand-seq, and BioNano optical maps, have provided valuable contributions to the global discovery and catalogues of INVs. However, we are still in need of a cross-comparison of those catalogues involving INVs detected by different technologies. **Method:** Here, we investigate the INVs available from three different datasets: the dataset in the study of Ebert and Audano *et al.* (PMID: 33632895), the Genome Aggregation Database (gnomAD), and the Database of Genomic Variants (DGV). gnomAD and DGV data were directly used in the hg19 version, while INVs were converted to hg19 coordinates for the dataset of Ebert and Audano *et al.* We applied the Bedtools intersect function with the 0.5 fraction option to compare each dataset pairwise. **Results:** There are 737 INVs in gnomAD (59 bp-74.35 Mb, with a median of 1.117 kb), 3,349 INVs in DGV (51 bp-9.73 Mb with a median of 2.639 kb), and 367 INVs in Ebert and Audano *et al.* (300 bp-5.3 Mb with a median of 21 kb). 91.3% and 4.2% of the gnomAD INVs overlapped with the DGV and Ebert and Audano *et al.* INVs, respectively. On the other hand, 48.3% and 16.3% of the DGV INVs intersected with the gnomAD, and Ebert and Audano *et al.* 61.7% and 33.9% of Ebert and Audano *et al.* INVs overlapped with the DGV and gnomAD INVs. Next, we investigated whether INVs disrupt genes associated with any human disease phenotype in Online Mendelian Inheritance in Man (OMIM). Our analysis revealed 54 disease genes (27 AR, 17 AD, 6 AD/AR) overlapping gnomAD INVs (67 bp-0.7 Mb with a median of 711 bp), 172 disease genes (78 AR, 59 AD, and 23 AD/AR) overlapping DGV INVs (53 bp-1.064 Mb with a median size of 1.272 kb), and 23 disease genes (13 AR, 7 AD, and 2 AD/AR) overlapping INVs in the study of Ebert and Audano *et al.* (547 bp-0.110 Mb with a median of 9.562 kb). **Conclusion:** Databases such as gnomAD, DGV, and published datasets provide valuable resources to scientists studying INVs. To our knowledge, we present the first study comparing INVs in these datasets.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2275. Comparison of diagnostic yield between exon-targeted microarray and standard microarray for patients with congenital heart defects (CHD).

Authors:

S. Sulpizio, V. Cawich, A. Morton, A. Dikeman, S. Sackmann, A. Flage, B. Torchia; Allele Diagnostics Inc, Spokane, WA

Abstract Body:

In an effort to increase the diagnostic yield of a standard chromosome microarray (CMA) for patients with congenital heart defects (CHD), we designed a whole genome (CGH+SNP) CMA (the CHD array) with enhanced exon-level coverage in 356 genes associated with CHD. As with our laboratory's standard CMA, process improvements for DNA extraction, labeling, and hybridization were made to provide results in as little as 2 days.

To evaluate the diagnostic yield of the CHD array, 499 deidentified samples from patients with CHD indications were processed. Data was interpreted by an ABMGG Cytogeneticist and classified using ACMG guidelines. Reportable variants (Pathogenic, Likely Pathogenic, VUS) were then categorized based on likelihood of a causative link with the specific type of CHD in the patient and likelihood of being detected by our standard CMA.

In 499 samples, 108 had variants strongly associated with CHD and 30 had variants where the relationship with CHD was uncertain. Therefore, the diagnostic yield of the CHD array was 27.7%. Variants potentially associated with CHD, that would not have been identified by standard CMA, were detected in five samples. These included 6 deletions within autosomal recessive disease genes, ranging in size from 521 bp to 8.2 kb, which were confirmed by dPCR. Thus, for this cohort, there was an increased diagnostic yield of 1.0% for the CHD array over the standard CMA.

A cohort of 8079 samples submitted for diagnostic testing by standard CMA was evaluated against the CHD array design *in silico* using genomic coordinates of the identified variants. Three cases run on standard CMA had pathogenic or likely pathogenic variants that were strongly associated with CHD and would not have been identified by the CHD array. One of these variants, detected in two patients, involved a gene that has a clear but infrequent association with CHD. The other involved a gene only recently associated with CHD, demonstrating the importance of frequent gene curation and periodic updates to the array design.

In conclusion, a CMA platform with enhanced exon-level coverage of genes associated with CHD increased the diagnostic yield of reportable variant types by detecting smaller alterations within the targeted regions. Although a small number of CHD-related alterations may not be detected by an exon-level array, the more targeted design clearly identified additional exon-level alterations. In combination with a rapid turnaround time, the increase in diagnostic yield could significantly impact the care of patients with structural heart defects.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2276. Comparison of positive findings between CMA/exome combo and Whole Genome Sequencing: Aclinical lab experience

Authors:

X. Zhao^{1,2}, X. Luo^{1,2}, J. Lattier², P. Liu^{1,2}, L. Meng^{1,2}, C. Eng^{1,2}, F. Xia^{1,2}, H. DAI^{1,2}, W. Bi^{1,2}; ¹Baylor Coll. of Med., Houston, TX, ²Baylor Genetics, Houston, TX

Abstract Body:

Introduction: CMA and WES have been utilized as genetic diagnostic tests over the last decade for patients with suspected underlying genetic disorders. However, the conventional phenotype-driven and stepwise approach can be time consuming and often lead to inconclusive results. In contrast, WGS offers the potential of a single test that captures nearly all genomic variation in an unbiased manner. As WGS becomes more clinically available, it is often asked whether and when the patients would benefit the most if WGS is ordered as the first-tier diagnostic test. To provide insights into this question, a comprehensive comparison of WGS and WES+CMA is conducted at Baylor Genetics. **Method:** We compared the positive findings, turnaround time (TAT), and diagnostic rate among patients with concurrent orders of CMA and WES (the Combo) and patients who had clinical WGS test in the same period of time. Patients who received a diagnosis with WES or CMA alone and did not proceed with the combo test were not included. CMA was performed using a comprehensive custom designed array with SNP probes and exon-targeted coverage for >4,200 disease genes or candidate disease genes. **Result:** Overall, 206 patients with diagnostic findings were included in this study (145 combo, 61 WGS). The average TAT of the Combo is 33.0 days with a median of 25.4 days. 29.0% of cases received results within 2 weeks (42/145). On the other hand, the average TAT of WGS is 26.3 days. Among these, 37.7% of cases were reported within 5 days (23/61). Of the aberrations detected by WGS, 16% of positive WGS cases (10/61) could potentially be missed. Six findings would definitely be missed by the Combo, including two deep intronic variants, two mitochondria variants, one complex structural rearrangement, and one short tandem repeat (STR) variant. In addition, three of the four <10kb CNVs may be detectable by our comprehensive array because of using exon-targeted array, but all of these four small CNVs may be missed by a clinical array without enhanced coverage for disease genes. Majority of the positive WGS cases were rapid test from NICU patients who have one or more of the following clinical indications (60.7%, 37/61): seizures, congenital heart disease, and respiratory failures. **Conclusion:** We reported here a comparison between the use of WGS versus ordering WES and CMA in clinical settings. In addition to the variants detectable by the combo, WGS may detect small CNVs, deep intronic variants, mitochondrial variants and balanced chromosome rearrangements. In addition, TAT as short as 5 days may be provided by WGS. Our studies support that WGS is a rapid diagnostic tool for diverse genetic aberrations.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2277. Copy Number Variations (CNVs) detected in 14% of patients presenting with peripheral neuropathies in a French cohort.

Authors:

A-S. Lia^{1,2,3}, I. Pyromali¹, F. Miressi¹, N. Benslimane¹, A. Nizou¹, P. Derouault³, P. Chazelas^{1,2}, F. Favreau^{1,2}, F. Sturtz^{1,2}, C. Magdelaine^{1,2}; ¹NEURIT-UR20218-Limoges Univ., Limoges, France, ²Laboratoire Biochimie Génétique Moléculaire - Limoges Hosp., Limoges, France, ³Unité Fonctionnelle de Bioinformatique - Limoges Hosp., Limoges, France

Abstract Body:

Introduction: Inherited Peripheral Neuropathies, such as Charcot-Marie-Tooth Disease (CMT), Hereditary Motor Neuropathies (HMN) and Hereditary Sensitive Neuropathies (HSN) are characterized by damages of motor and/or sensory peripheral nerves. To date, more than hundred genes are known to be responsible for these diseases. Classically, investigations are performed by Next-Generation-Sequencing (NGS) allowing the easy and fast detection of Single Nucleotide Variants (SNVs) or short indels. The goal of this study was to analyze differently these NGS data in order to detect unbalanced Structural Variations (SVs), also called Copy Number Variations (CNVs).

Materials and Methods: DNA from 919 patients presenting with peripheral neuropathies were studied by NGS using either an amplicon-targeted sequencing panel or a capture-targeted panel. NGS data were then analyzed by the bioinformatics tools Cov'Cop and CovCopCan, developed in our laboratory, in order to detect CNVs. To characterize more precisely the pointed out unbalanced SVs, qPCR, CGH and/or Sanger sequencing have been initiated.

Results: We found that 130 of the 919 patients harbor CNVs (14.1%): 87 duplications and 43 deletions. A quarter of these CNVs are large (involving more than three exons). We initiated the molecular characterization of these 130 CNVs. The first results allowed involving a large deletion in *SACS* in Charlevoix-Saguenay spastic ataxia, two *KIF5A* large deletions of respectively 5 and 14 exons, but also the deletion of a single exon in *SH3TC2* in Charcot-Marie-Tooth disease. In addition, we showed the non-pathogenicity of a partial *REEP1* duplication in spastic paraplegia 31, but we could not clearly concluded about the involvement of the complete *AARS* duplication in Charcot-Marie-Tooth disease. The analysis of the other CNVs is in progress and it is currently difficult to estimate the amount of pathogenic CNVs within the 130 CNVs identified.

Conclusions: By analyzing our NGS data with specific bioinformatics tools allowing the detection of CNVs, we identified unbalanced SVs in 14.1% of our patients. Interestingly, the first analyses showed that some of them are pathogenic and can explain the symptoms of the patients. This approach helps us to increase the positive diagnosis rate of our patients presenting with peripheral neuropathies and we suggest molecular biologists and geneticists to use this strategy for all inherited diseases to improve the diagnosis.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2278. CREBBP related Rubinstein Taybi syndrome: Report of 3 patients

Authors:

I. Panigrahi, P. Srivastava; PGIMER, Chandigarh, India

Abstract Body:

Rubinstein Taybi syndrome (RSTS) is a cause of intellectual disability and postnatal growth retardation. The genes associated with this phenotype include *CREBBP* and *EP300*. We report 3 children diagnosed with RSTS and were found to have single nucleotide variants or copy number variant (CNV) involving the *CREBBP* gene. The three children were 1-2 year old and evaluated in the Genetic Clinic. All three presented with developmental delay and dysmorphism, one child had broad thumbs. Two patients were found to harbor single nucleotide variants (SNVs) in the *CREBBP* gene and one was detected to carry heterozygous deletion, a CNV on chromosome 16p13.3 identified on MLPA. Of the SNVs, both variants were truncating variant in exon 10 and exon 22 of the gene respectively. In one patient, MRI brain revealed agenesis of corpus callosum, and in another micro-hemorrhages in left basal ganglia and right cerebellar hemisphere with loss void in straight sinus and right transverse sinus were observed. Thus, RSTS is a relatively common cause of developmental delay or intellectual disability. We discuss the clues and approach to early diagnosis in RSTS. Appropriate evaluation can help in early diagnosis and appropriate genetic counseling in the affected families.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2279*. Deciphering the genetic etiology of idiopathic male infertility by low-pass mate-pair genome sequence analysis.

Authors:

J. Qian¹, T. Law^{1,2,3}, M. Chau^{1,2,3}, Y. Cao^{1,2,3,4}, S. Tong¹, Y. Kwok^{1,2,3}, T. Leung^{1,2,3,5}, J. Chung^{1,2,3,4}, K. Choy^{1,2,3,4,5}, Z. Dong^{1,2,3,4}; ¹Dept. of Obstetrics and Gynaecology, The Chinese Univ. of Hong Kong, Hong Kong, China, ²Shenzhen Res. Inst., The Chinese Univ. of Hong Kong, Shenzhen 518057, China, ³Hong Kong Hub of Paediatric Excellence, The Chinese Univ. of Hong Kong, Hong Kong, China, ⁴The Fertility Preservation Res. Ctr., Dept. of Obstetrics and Gynaecology, The Chinese Univ. of Hong Kong, Hong Kong, China, ⁵The Chinese Univ. of Hong Kong-Baylor Coll. of Med. Joint Ctr. for Med. Genetics, Hong Kong, China

Abstract Body:

Objective: The genetic etiology of idiopathic male infertility contributed by copy-number variant (CNV), structural variant (SV), and absence of heterozygosity (AOH) is yet to be established. We aimed to investigate the genetic contributions in male infertility with negative results from routine genetic tests by applying our in-house low-pass mate-pair genome sequencing (GS). **Methods:** This is a retrospective cohort study of 101 patients (mean age = 36) (63 with non-obstructive azoospermia and 38 with severe oligospermia) with primary infertility and negative results from Y-microdeletion detection and karyotyping analysis. Low-pass mate-pair GS (4X) was performed in each subject for identifying clinically significant CNVs, SVs, and regions with AOH using in-house analytical pipelines. High read-depth GS (40X) was subsequently performed in five cases with intragenic deletions affecting autosomal recessive (AR) genes that related to male infertility or cases with multiple regions of AOH. **Results:** Candidate molecular etiologies were identified in 20 of 101 cases (19.6%). Firstly, 12 likely clinically significant CNVs were identified and subclassified into four groups: (1) three pathogenic deletions/duplications showed direct evidence to explain the infertility; (2) three pathogenic microdeletion/duplication syndromes with limited evidence of male infertility (22q11.2 duplication syndrome, 16p13.11 microdeletion syndrome, and terminal deletion of 5p); (3) two pathogenic deletions in AR manner (genes *IFT74* and *LDLR*); and (4) four variants of uncertain significance involving male infertility related genes. In addition, three inversions and six complex insertions involving OMIM disease-causing genes or genes curated by GeneReviews were reported. For instance, *DMRT1*, deletion of which leads to nonsyndromic 46, XY disorders of testicular development, was directly disrupted by a cryptic insertion. Furthermore, multiple regions with AOH were identified in three cases indicating parental consanguinity (n=2) and uniparental disomy (UPD; n=1). Lastly, high read-depth GS excluded any pathogenic mutations in two cases with intragenic deletions affecting AR genes and one with suspected UPD, but indicated pathogenic homozygous point mutations in two cases suspected parental consanguinity. **Conclusions:** Our study provides a landscape of various types of chromosomal abnormalities in idiopathic male infertility by applying our in-house low-pass mate-pair GS. Most importantly, the clinically significant variants potentially contributed to male infertility in approximately 20% of the cases, underappreciated by the standard care.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2280. Deep exome sequencing to detect mosaic variants in craniofacial microsomia

Authors:

N. Parmalee¹, J. Gustafson¹, D. Jensen¹, V. Dmyterko¹, R. Bly¹, J. Perkins¹, M. Cunningham², D. Luquetti³, K. Sie¹, C. Heike³, J. Bennett⁴; ¹Seattle Children's Res. Inst., Seattle, WA, ²Univ. of Washington and Seattle Children's Res. Inst., Seattle, WA, ³Univ. of Washington, Seattle, WA, ⁴Seattle Children's Res. Inst., Seattle, WA

Abstract Body:

We conducted deep exome sequencing to a depth of >1000x in multiple tissues from probands with craniofacial microsomia (CFM), an isolated birth defect. We sequenced multiple tissues, as well as blood from the proband and blood or saliva from both parents. Our goal was discovery of mosaic variants in causative genes in which variants are present in affected tissues and either absent from proband blood or present at lower levels in blood and absent from both parents. Deep sequencing allows us to detect variants that would be missed with a standard exome.

Isolated structural birth defects are congenital anomalies that affect the morphology of a single region or tissue as opposed to complex or syndromic anomalies that may span multiple tissues or involve multiple systems. We focused on CFM in which craniofacial features including the mandible and maxilla are underdeveloped and the ear may be underdeveloped or absent. A constellation of other morphological signs may be present; CFM has a high degree of phenotypic variability. Presentation is generally asymmetrical with the right side affected more often. The prevalence of craniofacial microsomia is 1:3500-1:4000 births. CFM is thought to arise in early embryogenesis as a defect in neural crest cell migration into the 1st and 2nd pharyngeal arches. The cause for this is unknown.

We are investigating postzygotic somatic mutations that lead to genetic mosaicism and may be causative in isolated structural birth defects. We hypothesize that some percentage of unsolved genetic conditions may be due to mosaicism in which the mutation is not present in the blood and thus not detected with a standard exome. In this study we selected CFM as a model; while some patients have family history, in large part CFM cases are sporadic. We hypothesize *de novo* variation and we are investigating the hypothesis of postzygotic mutation that leads to developmental defects in patterning. It is not known what tissue may harbor a causative mutation. We selected CFM as a model in part because multiple tissues are discarded during reconstructive surgeries and are available for sequencing, including preauricular tags, cartilage, bone, teeth, and soft tissue.

Based on expertise in the lab with mosaicism in other conditions we expected that the variant allele frequency (VAF), the percentage of sequencing reads supporting a true mosaic variant, may be very low in an affected tissue. We routinely detect mosaic variants in known causative genes with a VAF of less than 1% that are confirmed with orthogonal assays. We report results from our search for mosaic variants at low VAF that would not be detected with a standard coverage exome that present in tissue and not in the blood.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2281. Development and analytical validation of an innovative 397 gene genome-based inherited kidney disease panel

Authors:

K. Weymouth¹, V. Charu², N. Hammond¹, S. White¹, R. da Silva Francisco, Jr.², J. Kohler³, P. Tong¹, Z. Ng¹, C. Ho¹, K. Dumas¹, D. Fisk¹, C. Hale¹, T. Nguyen Dolphyn¹, C. Reavey¹, E. E. Smith¹, W. Qiao^{1,2}, Y. Yang^{1,2}, N. Kambham², S. A. Scott^{1,2}; ¹Clinical Genomics Lab., Stanford Hlth.care, Palo Alto, CA, ²Dept. of Pathology, Stanford Univ., Stanford, CA, ³Dept. of Pediatrics, Stanford Univ., Stanford, CA

Abstract Body:

Chronic kidney disease (CKD) is an important cause of morbidity and mortality, which can progress to end stage renal disease necessitating dialysis and renal transplantation. Monogenic causes of CKD are increasingly recognized, as evidenced by diagnostic yields in adults and children that range from ~10-30%, supporting the implementation of genetic testing to guide management and transplant considerations. Diagnostic panel and exome testing often employ enrichment-based sequencing; however, technical advantages using genome sequencing include unbiased coverage, copy number variant (CNV) detection, and flexible interrogation of clinical regions of interest (ROI). Therefore, we developed an innovative inherited kidney disease (IKD) panel based on genome sequencing (NovaSeq 6000; $\geq 40X$) that includes 397 clinically significant genes implicated in a spectrum of monogenic kidney disorders and selected syndromic diseases with abnormal nephrology phenotypes. Clinically relevant non-coding variants (ClinVar, HGMD) were also included in the ROI, and the panel was subsequently validated by measuring performance with detecting single nucleotide variants (SNVs), insertions/deletions (indels), and CNVs. Small variants (SNV/indel) were called using DRAGEN v3.9.5 and benchmarked for accuracy against the GIAB v4.2.1 truth set using hap.py, which resulted in $\geq 99.9\%$ recall and precision across the high confidence regions of the IKD genome panel. Moreover, small variant precision across all assessments (inter-run, intra-run, inter-instrument, inter-side, inter-flow cell) were 100% in the non-difficult regions of the IKD genome panel, indicating that the assay was robust and precise. Clinical specimens (blood, saliva, assisted saliva, fibroblast) were also validated using paired samples and measuring non-reference genotype concordance across the IKD genome panel, which resulted in SNV/indel concordance $\geq 99.9\%$. Clinical performance was assessed by testing ≥ 30 samples with known clinically significant small variants and/or CNVs, which previously had undergone orthogonal testing using other methods. Taken together, these results support the validation and implementation of this novel IKD genome panel for clinical testing among adults and children with abnormal nephrology phenotypes.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2282. Development and clinical implementation of genome-based germline exome testing.

Authors:

K. Dumas¹, D. Fisk¹, C. Hale¹, N. A. Hammond¹, C. Ho¹, L. Liao¹, Z. Ng¹, T. Nguyen Dolphyn¹, Y. Park¹, C. Reavey¹, H. Shi¹, E. E. Smith¹, T. Tan¹, P. Tong¹, N. Watson¹, K. S. Weymouth¹, S. White¹, L. Hudgins², E. Ashley³, J. A. Bernstein², E. Spiteri^{1,4}, W. Qiao^{1,4}, S. A. Scott^{1,4}, Y. Yang^{1,4}; ¹Clinical Genomics Lab., Stanford Hlth.Care, Palo Alto, CA, ²Dept. of Pediatrics, Stanford Univ., Stanford, CA, ³Stanford Ctr. for Inherited Cardiovascular Disease, Stanford, CA, ⁴Dept. of Pathology, Stanford Univ., Stanford, CA

Abstract Body:

Exome sequencing is widely used as a diagnostic test for children and adults with suspected Mendelian disease, and typically employs enrichment-based sequencing targeted to all coding exons; however, limitations of this approach include the additional steps of enrichment, fixed test content, and interpretation limited to coding region and splice variants. Given the quality advantages and potential utility of using genome sequencing for clinical exome testing, we developed a novel genome-based exome assay that was validated and implemented as a clinical test. The Stanford genome-based exome reportable range was defined by coding regions of human RefSeq genes, GENCODE, CCDS, as well as all clinically relevant non-coding variants, as defined by ClinVar likely pathogenic/pathogenic, and HGMD disease- and likely disease-causing. Validation of the genome-based exome was accomplished by sequencing >90 samples (NovaSeq 6000; ≥40X) that were strategically selected to measure accuracy, precision, clinical performance, and other required elements of assay development. Small variants (single nucleotide variants (SNVs) and insertions/deletions (indels)) were called using DRAGEN v3.8.4 from seven GIAB samples and benchmarked for accuracy against the GIAB truth set using hap.py, which resulted in >99.9% recall and precision across the high confidence regions. Moreover, small variant precision across all assessments (inter-run, intra-run, inter-instrument, inter-side, inter-flow cell) were >99% in GIAB non-difficult regions, indicating that the assay is robust and precise. Clinical specimens (blood, saliva, assisted saliva, fibroblast) were also validated using paired samples and measuring non-reference genotype concordance, which resulted in SNV/indel concordance of ~99%. Clinical performance was assessed by testing 29 samples with known clinically significant small variants, which resulted in a clinical accuracy of 100% (95% CI: 99.7-100%). To date, the genome-based exome has been performed on 86 patients (trios and probands), which resulted in a positive diagnostic rate of 20.9% across indications. These results support the validation and implementation of this novel genome-based exome assay for clinical testing among children and adults with suspected Mendelian disease.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2283. Development of a comprehensive whole genome sequencing test for cardiovascular disease patients.

Authors:

L. Amendola¹, A. Coffey¹, S. Strom¹, A. Kesari¹, J. Avecilla¹, J. Lowry¹, S. Thacker¹, C. Brown¹, K. Golden-Grant¹, M. Brown¹, B. Milewski¹, J. Belmont², D. Perry¹, R. Taft¹; ¹Illumina Inc, San Diego, CA, ²Genetics & Genomics Services, Inc, Houston, TX

Abstract Body:

Guidelines for genetic assessment in individuals with cardiovascular disease (CVD) are inconsistently followed, in part due to the complexity of the recommended testing regime. In an effort to build a single scalable, and future-proof, molecular testing platform that can meet the needs of CVD patients, we have developed a clinical whole genome sequencing (WGS) test that includes an in-silico CVD gene panel, CVD risk alleles, PRS for coronary artery disease and pharmacogenomic findings. Here we focus on the selection and curation of gene-disease relationships and variant types related to CVD as a fundamental component of robust molecular test development. The broad primary CVD indications we considered included aortopathy, arrhythmia, cardiomyopathy, dyslipidemia, hypertension, and thrombophilia. Genes and genomic variation associated with these CVD indications were identified through comprehensive literature review, assessment of commercially available panels, expert interviews, and review of relevant ClinVar and HGMDTM variants. Gene-disease relationships curated as Definitive or Strong were included on the CVD gene panel, and an internal knowledge base of relevant variant classifications was developed to reduce curation burden at the time of testing. WGS performance was assessed across all genes and variants of interest. Of the 309 gene-disease pairs identified, 132 had been previously curated as having a Definitive or Strong relationship and were included in the in-silico CVD gene panel. All remaining genes (n=177) were prospectively curated based on the ClinGen framework, leading to the inclusion of an additional 57 genes, and a final in-silico panel size of 189 genes. Potential or known pathogenic genomic rearrangements were identified in 155 genes, with 37 having more than one such variant smaller than 10 kb. Short tandem repeats were included for 3 genes (*DMPK*, *CNBP* and *FXN*) where they are an established mechanism for conditions with CVD phenotypes. WGS performance was assessed across all 189 genes and associated variants, including repeat expansions, which showed both high sensitivity and specificity and overall performance acceptable for a broad population screen. Variants identified as major contributors to genetic CVD were manually curated for the knowledge base. A CVD-focused WGS test may be able to address challenges related to the underutilization of and variability in CVD genetic testing by providing a single multifaceted molecular diagnostic platform. These data demonstrate that the development of an evidence-backed in-silico WGS CVD testing panel component of such a test is tractable.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2284. Diagnosis of genetic causes of intellectual disability and multiple congenital abnormalities in center of excellence for human genetics in Egypt

Authors:

A. Mohamed¹, A. kamel¹, M. Eid², O. eid¹, M. Mekkawy¹, M. Zaki¹, E. Ashaat¹, G. Abdelsalam¹, M. Aglan¹, M. issa¹, S. Housein¹, s. Gharib¹, M. Elruby¹; ¹Natl. Res. Ctr., Cairo, Egypt, ²Natl. research center, Cairo, Dokki, Egypt

Abstract Body:

Aim: the aim of this study is the application of different cytogenomic technique in diagnosis of genetic causes of intellectual disability (ID)/multiple congenital abnormalities (MCA), and genotype/phenotype correlation. This work was done in the Centre of Excellence of Human Genetics, National Research Centre, Egypt. The center is funded by the STDF (project 5253). **Methods:** The total referral patients (ID/MCA) were 2170 through years 2016 to 2020, karyotype was don for all patients, FISH, multiple ligation probe amplification (MLPA) and array CGH technique were performed not to all patients due to high cost of the kits. **Results:** chromosomal abnormalities found in 6.6% of the referred patients. FISH was performed for 324 patients, 226 patients diagnosed as microdeletion syndromes,98 had chromosomal abnormalities or marker chromosome. The MLPA technique was performed for 160 patients who had ID/MCA, we used probe mix for MR, subtelomere and microdeletion syndromes. 26 patients (16%) had positive results. Array CGH were performed for 90 patients. Out of the 90 sampled there are 44 samples (49%) had large copy number variations, deleted and/or duplicated segments. Some of these copy number variations involved two chromosomes copy number variation or complex rearrangements which involved more than two chromosomes. Negative patients need whole exome or genome sequencing. The array CGH is the 1st tier for the diagnosis of genetic cause of ID/MCA and for genotype/phenotype correlation.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2285. Diagnostic approach of transthyretin amyloidosis in a Colombian patient

Authors:

J. Estela-Zape¹, L. Moreno Giraldo^{2,3}, D. Arturo-Terranova², J. Satizabal-Soto²; ¹Fundación Univ.ria María Cano, Cali, Colombia, ²Univ. del Valle, Cali, Colombia, ³Univ. Libre, Cali, Colombia

Abstract Body:

Transthyretin amyloidosis (ATTR) is characterized by the deposition of amyloid fibrils, especially affecting cardiomyocytes and peripheral nerves. In Colombia, to date 43 cases of ATTR have been confirmed, and due to its low prevalence it has been included in the list of orphan diseases, according to resolution 5265 of 2018, however, its clinical presentation is heterogeneous and the delay in diagnosis does not allow a timely approach. We present the case of a 5-year-old patient, product of the third pregnancy of non-consanguineous parents, who from the age of 20 months presented regression of developmental milestones, muscular atrophy, hypotonia, gait disturbance, cardiomyopathy, and gastrostomy. Given the complexity of the clinical picture and in order to determine the etiology of the disease, whole exome sequencing was performed. The heterozygous variant in the transthyretin gene (TTR) was identified: c.424GxA, (p.Val142Ile) (classically Val122Ile), with pathogenic clinical significance. This variant is commonly reported in patients with ATTR. Functional studies have shown that this variant renders the TTR complex unstable, leading to protein cleavage and decreased stability in retinol transport. New treatments have been developed in order to stop or delay amyloid deposition by transthyretin, stabilizers of the TTR gene molecule that reduces the rate of amyloid deposition, drugs that act by inhibiting hepatic expression of transthyretin by interfering ribonucleic acid or by antisense oligonucleotides that block the transcription of the messenger RNA of the TTR gene, drastically reducing hepatic production of TTR. In addition, advances in treatment show the use of CRISPR technology to edit the human genome in vivo and produce a chronic silencing of TTR expression in hepatocytes through a single intravenous administration. All these advances will allow timely therapy to be addressed in those diagnosed for the disease. The timely identification of the disease allows to provide an adequate genetic counseling, recognizing available treatments for the management of the pathology.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2286. Diagnostic Utility and Lessons Learned from Deep Sequencing Vascular Malformations

Authors:

C. Myers¹, C. R. Paschal^{2,3}, M. Gillentine^{4,1}, D. Sternes^{5,6}, Z. Nelson⁷, J. Perkins^{5,8}, J. Bennett^{9,10,2}; ¹Dept. of Lab., Seattle Children's Hosp., Seattle, WA, ²Dept. of Lab., Seattle Children's Hosp., Seattle, WA, ³Dept. of Lab Med. and Pathology, Univ. of Washington, Seattle, WA, ⁴Univ. of Washington, Seattle, WA, ⁵Seattle Children's, Seattle, WA, ⁶Patient-centered Lab. Utilization Guidance Services (PLUGS), Seattle Children's Hosp., Seattle, WA, USA, Seattle, WA, ⁷Seattle Children's Hosp., Seattle, WA, ⁸Dept. of Otolaryngology/Head and Neck Surgery, Univ. of Washington, Seattle, WA, ⁹Seattle Children's Res. Inst., Seattle, WA, ¹⁰Div. of Genetic Med., Dept. of Pediatrics, Univ. of Washington, Seattle, WA

Abstract Body:

Now recognized to be primarily caused by tissue-restricted somatic mutations, disorganized morphogenesis of blood vessels results in vascular malformations. Mutation identification increasingly drives targeted medical therapies but requires specialized diagnostic approaches. Here we report results on over 300 individuals using VANSeq, a clinical sequencing assay for vascular malformations. Over two years, we performed 319 clinical tests on 306 individuals. The most common clinical indication was isolated vascular malformation (lymphatic or venous), followed by vascular malformation with overgrowth. Of the 319 samples tested, 57% had pathogenic or likely pathogenic findings, 2% had variants of uncertain significance, 37% were negative (no variants reported), and 4% of testing could not be completed due to insufficient or poor-quality DNA. DNA was extracted from formalin-fixed paraffin-embedded (FFPE) tissue (41%), fresh-frozen tissue (33%), peripheral blood (22%), or saliva (4%). Diagnostic yield was highest when affected tissue was tested. Increasing the number of samples tested increased the diagnostic yield. Nine of 13 patients that had repeat testing had diagnostic results after the first test was non-diagnostic. The variant allele frequency (VAF) was less than 5% in nearly half of samples with a mosaic pathogenic variant identified (43%, n=143). Across the cohort, variants were reported in 25 of the 44 genes tested; *PIK3CA* accounted for the majority of positives (n=86), followed by *TEK* (n=18), *GNAQ* (n=10), *MAP2K1* and *RASAI* (n=8, each). Nearly half of patients with a *TEK* pathogenic variants had co-occurrence of a second somatic *TEK* variant (n=8/18). Four individuals had coexisting somatic activating mutations in two separate genes (ex: *BRAF* and *GNAQ*). Concomitant mutations frequently coexist with driver mutations in cancer, but this has not been well-documented for vascular malformations. Our experience highlights several considerations unique to clinical testing of vascular malformations.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2287. Diagnostic utility of the targeted next generation sequencing panel test for suspicious genetic glomerular diseases

Authors:

J. Kim, H. Lee; Dept. of Internal Med., Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract Body:

Background: A substantial proportion of glomerular diseases has genetic background, but there is no solid guideline concerning genetic work-up. The targeted next-generation sequencing (NGS) gene panel can identify possible mutations causing genetic glomerulopathies. We evaluated the diagnostic utility of targeted NGS panel for glomerular diseases. Methods: Patients who received targeted NGS panel from 2017 to 2021 were included. Our NGS panel covered 23 variants causing genetic glomerulopathies including collagenopathies (COL4A3, COL4A4, COL4A5, and MYH9), genetic nephrotic syndrome (ACTN4, ADCK4, ANLN, COQ2, COQ6, EMP2, INF2, NPHS2, NUP107, PLCE1, and TRPC6), and other syndromic diseases (NPHS1, LAMB2, LCAT, LMX1B, PAX2, SMARCAL1, WDR73, and WT1). Variants were classified according to American College of Medical Genetics and Genomics 2015 guideline. Diagnostic yield was calculated as positivity rate of pathogenic or likely pathogenic variants. Results: A total of 111 patients were included. The median age of disease onset and of receiving test was 9.0 [3.0;22.5] and 17.0 [7.5;33.0] years. Among them, 50 had family history of kidney disease and 3 had congenital anomalies. Seventy-three patients received percutaneous kidney biopsy for their pathologic diagnosis. Overall diagnostic yield of targeted NGS panel was 36% and higher in patients with younger onset age (<18 years) than those older [44.7% vs 17.1%, p=0.03]. It was associated with previous kidney biopsy (adjusted odds ratio (aOR), 6.23; 95% confidence interval (CI), 1.21-32.06), hematuria (aOR, 4.68; 95% CI, 1.32-16.56), systolic blood pressure (aOR, 0.94; 95% CI, 0.88-0.99) and absence of edema (aOR, 0.05; 95% CI, 0-0.66) even after covariate adjustment. Among 40 patients with positive gene test results, 16 changed their diagnosis from initial pathologic diagnosis and 13 patients received new genetic diagnosis. Conclusion: Targeted NGS panel test for genetic glomerular diseases was useful in genetic diagnosis with a modest diagnostic yield of 36%. Previous kidney biopsy, hematuria, low systolic blood pressure and absence of edema were associated with higher positivity rate.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2288. Diagnostic yield of trio and individual Whole Exome Sequencing in a group of colombian patients with a suspected monogenic disease.

Authors:

E. Rodriguez-Alvarino¹, S. Bello Uyaban², D. Otero Rodríguez², M. Galvez²; ¹Univ. Natl. de Colombia, Bogota, Colombia, ²Gencell - Genuino Group, Bogotá, D.C, Colombia

Abstract Body:

Introduction: During last years, Whole Exome Sequencing (WES) has had great acceptance for the evaluation of genetic disorders. Due to the access to next-generation sequencing (NGS), bioinformatics tools, and professionals' experts, these molecular tests represent a first-tier diagnostic tool for many genetic diseases. **Objective:** To compare the diagnostic yield of singleton and trios WES in a Colombian cohort with clinical suspicion of a genetic disease, as well as the yield per phenotype groups, between June 2020 to May 2021. **Methodology:** A cross-sectional descriptive, retrospective study, approved by an ethics board, was carried out. Medical records were reviewed to obtain phenotypic data and trio WES was performed. Variant calling was done using *GATK* best practices, with subsequent annotation and filtering using Human Phenotype Ontology (HPO) terms and variants were classified using the American College of Medical Genetics and Genomics (ACMG) 2015 guidelines. Analysis as singleton exome and trio exome were performed in every case. Cases were considered positive when pathogenic (P) or likely pathogenic (PP) variants were identified, with a mechanism of inheritance explaining the patient's phenotype. Additionally, cases were stratified by groups according to the main clinical phenotype, to evaluate the specific yield of the test. **Results:** A total of 100 cases were included. The average age was of 8,6 years, and seven cases were born of consanguineous parents. The diagnostic yield was 26% for singleton and 31% for trio WES. Among positive cases, 18% had an autosomal dominant disease, 8% an X-linked, 4% autosomal recessive, and 1% a digenic inheritance disorder. Two cases had incidental findings in *PMS2* and *MSH6* genes, and two other cases had concomitant monogenic conditions. The most prevalent phenotype was neurodevelopmental delay in 47% individuals, with a yield of 36,1%. The highest yield was found for paralysis (71,3%), followed by hypoacusis (55,5%) microcephaly (50%), pulmonary hypertension (50%) and hematologic alterations (50%). Abnormal movements had the lowest diagnostic rate (10%). **Conclusion:** This study showed a 5% better yield for trio exome in comparison to singleton. This is mainly explained by the identification of *de novo* variants and confirmed compound heterozygous mutations, which cannot be established in singleton exomes. WES showed the best performance in neurological phenotypes, especially paralysis.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2290. Does a larger gene panel size change the mutation detection rate in hypertrophic or dilated cardiomyopathy ?

Authors:

S. Fokstuen¹, P. Meyer², E. Hammar¹, M. Beghetti³, F. Masclaux¹, M. Guipponi¹, J. Wacker³, A. Py², J. Schwaiger⁴, M. Abramowicz¹, S. Suchet¹, C. Gruner⁴, J-L. Blouin¹; ¹Genetic Med., University Hospitals of Geneva, Switzerland, ²Cardiology, University Hospitals of Geneva, Switzerland, ³Pediatric Cardiology, University Hospitals of Geneva, Switzerland, ⁴Cardiology, University Hospitals of Zürich, Switzerland

Abstract Body:

Implementation of next generation sequencing (NGS) has led to a rapid expansion in the number of genes included in diagnostic genetic testing for cardiomyopathies. This tendency is changing with the evidence-based assessment of genes conducted by ClinGen. Our aim was to evaluate the mutation detection rate and the involved genes in patients with isolated hypertrophic or dilated cardiomyopathy (HCM and DCM) according to the gene panel size. Our NGS approach consists of targeted exome sequencing. Before September 2019, we have used in-house gene panels including 65 resp. 78 cardiomyopathy genes. Since October 2019, we switched to Genomics England PanelAPP, including 144 resp. 140 high-evidence based genes causative for the different subtypes of cardiomyopathy. A total of 141 patients with non-syndromic HCM (109) or DCM (32) underwent molecular testing. Our mutation detection rate was 47 % (35/74) with the in-house panels, 39 % (26/67) with the PanelAPP panels and 43 % (61/141) overall. The use of the larger PanelAPP panels did not show any difference in the genes found with causative variants. All the genes with pathogenic or likely pathogenic variants identified with the larger PanelAPP panels were already included in our smaller in-house panels. Our results confirm the lack of major clinical benefit of large panels in isolated HCM or DCM and support the recommendations of the 2021 European Society of Cardiology heart failure guidelines, which suggest to use small evidence-based panels and to consider additional large panels only if there is a clear family history or a specific phenotype.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2291. *DST* variants are responsible for neurogenic arthrogryposis multiplex congenita confirming the large clinical spectrum of type VI hereditary sensory autonomic neuropathy

Authors:

Y. Capri^{1,2}, L. BOURMANCE², C. DUPONT³, M-H. SAINT-FRISON⁴, F. Guimiot^{5,4}, S. GROTTTO⁶, A. LAQUERRIERE⁷, Y. CHITRIT⁸, J. Melki²; ¹Genetic Dept.-Univ. Hosp. APHP-Nord Robert Debré, Paris, France, ²INSERM-UMR1195, University-Paris Saclay-Le Kremlin Bicêtre, France, ³Cytogenetic Unit-APHP-Hosp. Robert Debre, Paris, France, ⁴Foetopathology Unit-Univ. Hosp. APHP-Nord Robert Debré, Paris, France, ⁵INSERM UMR1141, University-Paris Nord, France, ⁶Univ. Hosp. APHP-Ctr. Hosp. Cochin, Paris, France, ⁷Normandie Univ.-INSERM UMR1245, Rouen, France, ⁸Obstetric Dept.-Univ. Hosp. APHP-Nord Robert Debré, Paris, France

Abstract Body:

Arthrogryposis multiplex congenita (AMC) is a developmental condition characterized by multiple joint contractures resulting from reduced or absent fetal movements. Through whole-exome sequencing combined with high resolution array-CGH from DNA of a fetus presenting with early onset AMC, we identified biallelic loss-of-function mutations in *DST* with a stop gain variant (NM_001144769.5:c.12208G to T: p.Glu4070*) on one allele and a 149 kb microdeletion including exons 26 to 83 of the *DST* gene (arr[GRCh37] 6p12.1(56323554_56472573) x1) on the other allele. Transmission electron microscopy of the sciatic nerve revealed abnormal morphology of the peripheral nerve with severe hypomyelination associated with dramatic reduction of fiber density which highlight the critical role of *DST* in peripheral nerve axonogenesis during development in human. Mutations in *DST* cause hereditary sensory and autonomic neuropathies (HSAN) type VI which has been reported in several unrelated families with highly variable age of onset from fetal to adult onset. Our data enlarge the clinical spectrum and disease mechanisms of neurogenic AMC.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2292. Evaluate using artificial intelligence and machine learning in to predict deletion detection possibility in a cohort of chromosome 22q11.2 deletion syndrome cases

Authors:

I. Alabdulkareem¹, M. Addam², A. Alkhalidi², M. Albalwi³; ¹SDAIA, Riyadh, Saudi Arabia, ²King Abdulaziz Med. City, Riyadh, Saudi Arabia, ³Natl. Guard Hosp, Riyadh, Saudi Arabia

Abstract Body:

Chromosome 22q11.2 deletion syndrome (DiGeorge syndrome) is a neurodevelopmental illness that has been intensively investigated to elaborate more on the connection between chromosomal microdeletion, brain development, cognitive function, and psychiatric signs. It is a rare genetic disorder that is characterized by a wide range of clinical manifestations presented with distinctive facial features, common congenital heart disease, immune deficiency, and learning difficulty. The majority of 22q11.2 deletion cases are due to de novo heterozygous deletion with the minority of familial inherited deletion. Cytogenetically, large deletion cases may be detected easily with G-banding high-resolution karyotyping and FISH analysis in 85%. Here, we evaluate using artificial intelligence and machine learning in to predict deletion detection possibility in a cohort of chromosome 22q11.2 deletion syndrome cases. The analysis pipeline is built on AI/ML image-based diagnostics undertaking the 450-655 banding level of chromosome 22 using analyzed CytoVision image karyotypes of the subjects followed by recalling and linking DiGeorge FISH probe signal images for the same subject(s). A subsequent AI algorithm was used composed of chromosomes images and FISH images obtained with high sensitivity in predicting Chromosome 22q11.2 deletions. Our finding looks promising in supporting and helping in both fast and urgent results and to the naïve cytogenetic laboratories who are constrained with providing advanced approach technology such as CGH array and provide a crucial cytogenetic diagnostic decision on their patients.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2293. Evaluating the potential for using internal data in assigning benign classification to variants of uncertain significance.

Authors:

J. Wang, E. Choo, C. Lo, H. Tadepally, T. Smart, B. Gall, R. Munch, N. Sanapareddy, Y. Sha, V. Souter; Natera, Inc., Austin, TX

Abstract Body:

Variants are classified as variants of uncertain significance (VUS) when current evidence is not sufficient to assign either a benign or pathogenic classification to them. However, since a definitive clinical interpretation is not available for common VUS, their frequent reevaluation often increases the workload for clinical labs. When a variant is seen in healthy individuals homozygous for a recessive disorder, a strong benign criterion (American College of Medical Genetics and Genomics (ACMG) BS2) can be applied to potentially reclassify it as likely benign (LB), thereby minimizing its frequent reevaluation. Our study objective was to evaluate the potential for using internal data in reclassifying VUS occurring in presumably healthy individuals homozygous for these variants. Following ACMG/Association for Molecular Pathology guidelines, we evaluated the potential for reclassifying VUS to LB using presumably healthy individuals homozygous for these variants who underwent carrier screening with the Horizon™ 274 gene panel between March 2020 to March 2021 (38401 cases). Only genes associated with early-onset recessive disorders (defined by MedlinePlus) were included in this study. Genes associated with autosomal dominant and X-linked phenotypes were excluded. Based on our screening and internal data curation, 58 VUS variants across 44 genes were identified in presumably healthy individuals homozygous for these variants. Of these, 12 variants across 11 genes could be reclassified as LB based on ACMG BS2 criterion. These 12 variants were found in a total of 2647 cases processed by our lab since April 2018. All our proposed classifications due to homozygosity on carrier screening were consistent with current benign (B)/LB classification in ClinVar. In contrast, only 7 out of 58 variants were reported for their homozygosity in healthy individuals in the public database gnomAD. None of the 58 VUS variants were submitted as pathogenic or likely pathogenic in ClinVar. Our results demonstrate that homozygosity data from presumably healthy individuals retrieved from an internal variant database may be useful for VUS reclassification of genes associated with early-onset recessive disorders. Additionally, findings here highlight the importance of internal databases carrying homozygosity data of variants from healthy individuals in improving the reporting efficiency of clinical labs. However, while internal data may be utilized to improve reporting efficiency, careful review of clinical data for these individuals should be conducted first.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2294. Evidence review of the use of first-line genome sequencing to diagnose rare germline disorders

Authors:

K. Wigby^{1,2}, D. Brockman³, G. Costain⁴, C. Hale⁵, S. Taylor⁶, J. Belmont⁷, D. Bick⁸, D. Dimmock², S. Fernbach⁹, J. Greally¹⁰, V. Jobanputra¹¹, S. Kulkarni⁹, E. Spiteri⁵, R. Taft⁶, Medical Genome Initiative; ¹Univ. of California San Diego, San Diego, CA, ²Rady Children's Inst. for Genomic Med., San Diego, CA, ³Massachusetts Gen. Hosp., Boston, MA, ⁴The Hosp. for Sick Children, Toronto, ON, Canada, ⁵Stanford Hlth.Care, Palo Alto, CA, ⁶Illumina Inc, San Diego, CA, ⁷Genetics & Genomics Services, Inc, Houston, TX, ⁸Genomics England Ltd., Gig Harbor, WA, ⁹Baylor Coll. of Med., Houston, TX, ¹⁰Albert Einstein Coll. of Med., Bronx, NY, ¹¹New York Genome Ctr., New York, NY

Abstract Body:

Background: Genome sequencing (GS) has evolved as clinically validated assay with robust capabilities to identify rare genetic disorders. Recently published guidelines and health technology assessments have sought to identify patient populations that would benefit most from exome or genome sequencing. Questions remain, however, regarding when to apply GS as a first-line diagnostic test for rare germline disorders. To address this question, the Medical Genome Initiative conducted a focused literature review as part of a framework to develop patient selection recommendations.

Approach: The purpose of the literature review was to appraise the evidence focused on the use of first-line GS for the diagnosis of rare germline disorders. Studies published from January 2011 - December 2021 that reported on the diagnostic yield (DY) or clinical utility of GS were included. Studies on fetal, microbiology, oncology and healthy cohorts were excluded. Study characteristics, DY, and clinical utility were abstracted by two independent reviewers. To assess study quality, we adapted criteria for diagnostic studies from the American College of Radiology which grades study design and measures to reduce sources of bias.

Results: Sixty-three studies met inclusion criteria, comprising over 2700 patients who received GS in one of the following settings: acute care pediatric inpatients, pediatric outpatients, adult outpatients, or mixed. GS was the first-line genetic test in 41% of studies analyzed. For pediatric acute care studies, the proportion of studies using GS as a first line test was considerably higher at 78.5%. The proportion of high quality first-line GS studies was greater than non-first line GS studies (23% vs 2%, p=0.017). Most studies (61.5%) of first-line GS involved mixed phenotype patient populations. The mean DY of first-line GS was 45.2% (14-73%) compared with 39.3% (6-86%) in studies with cohorts that received prior genetic testing and 33.5% (9-60%) in exome-negative cohorts. Clinical utility was assessed in 33% of studies, the majority of which used GS as a first-line test.

Discussion: These studies suggest that first-line GS is appropriate in certain patient populations, but additional high-quality studies are needed to further assess the use of first-line GS for indications beyond pediatric acute care patients. Heterogeneity among study design and variant reporting limits comparability of DY across studies and underscores the need for additional metrics to assess utility of GS. The learnings from this literature review combined with the clinical expertise of the expert panel will be used to develop recommendations for first-line GS.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2295. Examining the factors impacting molecular diagnosis during clinical exome sequencing re-analysis

Authors:

A. Chan¹, K. Wood¹, J. Knoblauch¹, K. Pechter², K. Cao³, J. Wang⁴, J. Murrell⁵, L. Conlin¹; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²The Children's Hosp. of Philadelphia, Philadelphia, PA, ³Children's Hosp. of Philadelphia, Philadelphia, PA, ⁴Children's Hosp. Philadelphia, Philadelphia, PA, ⁵The Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

The use of clinical exome sequencing (CES) has been increasing due to its comprehensive and diagnostic capacity. However, most individuals who receive CES remain undiagnosed. With the rapid increase in gene-disease associations, evolving clinical presentations and analysis pipelines, the diagnostic utility is expected to increase over time, making it crucial to reanalyze non-diagnostic CES results. We conducted a retrospective study to examine the factors impacting molecular diagnoses during CES reanalysis. Re-analyses of 182 CES were ordered at the Children's Hospital of Philadelphia between July 2018 and May 2022, of which 20 CES were diagnostic after reanalysis (11%). Of these 20 diagnostic cases, the initial CES result was negative in six, diagnostic in one, and non-diagnostic in 13. CES reanalysis was ordered on the diagnostic case to identify a molecular cause for the other features not explained by the original variant. In the majority of cases (80%, n=16), a molecular diagnosis was made after CES re-analysis due to new published information at the gene (10 new disease or candidate disease genes), variant (five with additional case reports), and phenotype levels (one phenotype expansions). This newly published information resulted in additional variants being reported at the time of reanalysis in 12 of the 16 cases, upgrade in variant classification in three cases, and no change in variant classification for one case. Four of the 16 cases with newly published information also had evolving clinical presentations that facilitated phenotype correlation. The elapsed time between the original CES and the newly published information that impacted interpretation ranged between 3.2 months and 3.6 years, with the median time elapsed being 2.1 years. In the remaining four exomes, diagnostic variants were identified in two cases (10%) as a result of an updated pipeline; another had an upgraded classification of a previously reported variant; and the remaining reanalysis reported the same diagnostic variant as the original report with additional information. In summary, this study highlights the utility of CES reanalysis for previously non-diagnostic results. Factors leading to diagnostic findings include new information about genes, variants or disorders, updated patient-specific clinical information, and technical improvements, such as an updated bioinformatics pipeline and analysis strategy. Exploring other outcomes from CES reanalysis (e.g., inconclusive CES that became negative after re-analysis) will help to determine the circumstances which reanalysis is most useful.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2296. Gabriella Miller Kids First Data Resource Center (DRC): Collaborative Platforms for Accelerating Pediatric Research

Authors:

D. Higgins¹, J-P. Thibert², M. Mattioni³, J. DiGiovanna³, R. L. Grossman⁴, S. L. Volchenboum⁴, D. M. Taylor^{5,6}, V. Ferretti², A. P. Heath¹, A. Resnick¹, Gabriella Miller Kids First Data Resource Center; ¹Ctr. for Data Driven Discovery in Biomedicine, Children's Hosp. of Philadelphia, Philadelphia, PA, ²CHU Sainte-Justine Res. Ctr., Montreal, QC, Canada, ³Seven Bridges, Cambridge, MA, ⁴Univ. of Chicago, Chicago, IL, ⁵Dept. of Pediatrics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ⁶Dept. of BioMed. and Hlth.Informatics, Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Children born with certain structural birth defects have an increased risk of developing cancer before they turn 18 years old. Structural birth defects and pediatric cancer share a context of altered developmental biology and their co-occurrence suggests underlying genetic pathways responsible for both of these. Investigating pediatric cancer and structural birth defects together will help facilitate discoveries into their genetic causes and spur advancement in prevention, detection, and therapeutics which can improve patient outcomes.

The Gabriella Miller Kids First Pediatric Research Program (GMKF) is an NIH Common Fund initiative focused on providing large-scale clinically annotated genomic data for pediatric cancer and structural birth defect cohorts, including tumor- and germline whole genome sequencing (WGS), trio based joint-genotyping, and paired RNA sequencing of somatic tissues. The GMKF Kids First Data Resource Center is charged with generating these datasets and empowering collaborative discovery on its data resource platforms. 24 Kids First studies are released on the Kids First Data Resource Portal, representing more than 20,000 participants and more than 1.0 PB of data, with additional datasets being released yearly. Here we present an overview of the Kids First Data Resource Center's data and platforms. Our harmonized clinical and genomic datasets and platforms are interoperable with other NIH-supported consortia, allowing investigators to easily incorporate Kids First data into their research projects. The Portal's Variant Search feature allows users to query variants or genes of interest present in Kids First datasets quickly, guiding them toward specific studies and participants. The Portal's integration with the CAVATICA cloud analysis platform allows users to run large-scale bioinformatic workflows as well as notebook-based analyses in R Studio and Jupyter Lab, all within their browser window. Overall, we show how the Kids First Data Resource can help investigators of all types access and analyze genomics-scale pediatrics data as part of their research.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2297. Generic genome sequencing: one lab flow for all

Authors:

G. Schobers^{1,2}, R. Derks¹, E. Bosgoed¹, D. Hellebrekers³, N. de Leeuw¹, A. Stegmann³, E-J. Kamsteeg¹, A. Paulussen³, M. Ligtenberg¹, A. van den Wijngaard³, C. Gilissen^{1,4}, R. Blok³, H. Brunner^{1,2,3}, H. Yntema^{1,2}, M. Nelen¹, L. Vissers^{1,2}; ¹Radboudumc, Nijmegen, Netherlands, ²Donders Inst. for Brain, Cognition and Behaviour, Nijmegen, Netherlands, ³MUMC+, Maastricht, Netherlands, ⁴Radboud Inst. for Molecular Life Sci., Nijmegen, Netherlands

Abstract Body:

Background Genetic laboratories maintain numerous workflows to diagnose the full spectrum of hereditary and congenital diseases, including traditional approaches and advanced technologies. A single generic workflow would increase efficiency quite dramatically. We therefore assessed whether genome sequencing (GS) can replace all existing workflows supporting germline genetic diagnoses. **Methods** We performed GS (NovaSeq6000TM; 37x mean coverage) on 1,000 cases with 1,271 clinically relevant variants, selected from 1 year's diagnostic yield in a tertiary referral center, identified through 15 different workflows. Variants were binned by size and type: small variants (SNVs and indels <50 bp), large variants (CNVs and repeat expansions) and other variants (SVs and aneuploidies). VCFs were queried per variant and assessed in Trusight Software Suite (DRAGEN Germline Pipeline, TSS, Illumina). **Results** Overall, 93.9% (1,194/1,271) of variants were detected with GS. Detection rates differed per type, with small variants detected in 95.2% (825/867), large variants in 91.9% (328/357), and other variants in 87.2% (41/47). Importantly, variants were identifiable through routine clinical interpretation strategies, including disease-based clinical filters or gene-specific searches in TSS. Variants that remained undetected were located in homologous/repetitive regions, or mosaic. **Conclusion** GS is an efficient generic workflow to capture clinically relevant germline variants in a 'one-test-fits-all-strategy'. No new challenges in variant detection were identified, besides those known for short-read sequencing. GS can therefore not only replace exome sequencing, but also >99% of Sanger sequencing, smMIP, MLPA, and array analysis, allele specific PCRs, and cytogenetic analyses including karyotyping and FISH. These results provide perspective on how genetic laboratories will evolve.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2298. Genetic Testing Interpretation Consult (GTIC): A Novel Electronic Health Record (EHR) Support Tool to Further Characterize Variants of Interpretation for the Nongenetic Provider to Support Their Clinical Practice.

Authors:

J. Talbert¹, J. Cogan², **R. Hamid**³; ¹Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Vanderbilt Univ, Nashville, TN, ³Vanderbilt Univ. Sch. of Med., Nashville, TN

Abstract Body:

Introduction: Exponential growth in next-generation sequencing has fueled an increase of genetic testing beyond clinical genetics and made it accessible to nongenetic specialties. Inherently the progress has resulted in variant classification discrepancies between laboratories despite ACMG guidelines and a glut of variants of uncertain significance (VUS) leading to ambiguous results, leaving nongenetic providers especially perplexed. The clinical impact of VUS' include missed diagnoses, lack of follow up testing or over testing, all leading to poor outcomes for patients and provider frustration. To support our nongenetic providers and deter from "curbside consults" for interpretation of VUS', we created a genetic testing interpretation consult (GTIC) via EHR Econsult with the goal of providing next steps for the clinician to follow. **Methods:** We designed an Econsult tool integrated into our EHR (EPIC). Nongenetic providers were trained to use this tool utilizing a web-based training module. Nongenetic providers requested further interpretation of genomic test results via this Econsult as an orderable. A dedicated team of genetic counselors, MD and PhD geneticists received the order along with limited clinical and family history and the molecular test report. The VUS' in the report were analyzed via a customized protocol utilizing variant specific bioinformatic tools, literature reviews, phenotyping databases, VUMC biobank data and phenotyping databases. A report was generated highlighting discrepancies, directing specialists for clinical correlation based on gene/phenotype with suggested next steps to further clarify etiology. The report was returned via the EHR to the requesting provider. **Results:** From 14 unique nongenetic specialties, 57 Econsults were ordered between Feb 2021-May 2022, variants analyzed totaled 195. Of the 57 cases, 24 resulted in alternate interpretations. Of the 24 cases, 19 variants were changed to VUS-possibly pathogenic and 25 were deemed VUS-possibly benign, revealing a 22.6% rate of discrepancy with the performing lab (44/195). **Conclusions:** We were able to provide further clarity for the VUS' in question for 24 cases (42%) and recommend specific next steps. With Econsult in place providers felt more confident and supported in ordering genomic testing.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2299. Genetic testing of *STRC* for hearing loss: a clinical lab's approach to pseudogenes

Authors:

M. Luo¹, M. Diaz-Miranda¹, T. Hartman¹, E. Bedoukian¹, I. Krantz¹, L. Conlin¹, J. Balciuniene²; ¹The Children's Hosp. of Philadelphia, Philadelphia, PA, ²PerkinElmer Genomics, Pittsburgh, PA

Abstract Body:

Pathogenic *STRC* variants cause DFNB16, a common type of autosomal recessive childhood-onset hearing loss (HL). *STRC* has a highly homologous pseudogene, *STRCP1*, located nearby that renders this region prone to non-allelic homologous recombination, leading to copy number (CN) changes, gene conversions, and fusion genes. The most common pathogenic *STRC* findings are large deletions, while disease-causing sequence variants (e.g. SNVs) are rarely reported due to analytic challenges caused by *STRCP1* on conventional sequencing tests. Therefore, a subset of patients with *STRC* diagnoses are missed and the scope of *STRC* pathogenic variants is not well explored. We use exome-based NGS and array CGH platform to screen a panel of 121 genes associated with pediatric ostensibly non-syndromic HL. This testing also includes an *STRC*-specific module which uses long range PCR followed by NGS and droplet digital PCR for integrative analysis of SNVs and CN variants in the *STRC* locus. Our testing of 479 unrelated probands with HL yielded diagnostic findings in 45% (216/479). *STRC* positive findings were identified in 5.6% (27/479), accounting for 12.5% of all diagnoses. Only 33% (9/27) of the *STRC* diagnoses were due to biallelic *STRC* deletions. A heterozygous *STRC* deletion and a hemizygous SNV accounted for 26% (7/27) and compound heterozygous SNVs accounted for 22% (6/27) of *STRC* diagnoses. In 7 of these patients, the SNVs were of *STRCP1* origin indicating focal or multi-exonic gene conversion events. More complex *STRC* findings were identified in the remaining 18.5% (5/27) patients. Two individuals carried a deletion on one allele and a non-functional *STRC-STRCP1* fusion on another allele. In three individuals, an apparently homozygous pathogenic variant was identified by NGS; however, the integrated *STRC* data were consistent with the presence of *STRC* deletion on one allele, and *STRC* duplication on the other allele with one *STRC* copy carrying a pathogenic SNV, and the second copy representing a non-functional *STRC-STRCP1* fusion. While some pathogenic variants in *STRC* can be detected by standard NGS and array CGH, the inclusion of this *STRC*-specific testing resulted in a doubling of the number of *STRC*-related diagnoses. Without the *STRC*-specific testing, diagnoses would have been completely missed in 5 individuals, and 8 individuals would have yielded inconclusive results. This integrative *STRC* testing approach uncovered signatures and potential mechanisms of complex genomic changes in the *STRC-STRCP1* locus and can serve as a model of cost-effective testing strategy for other similar pseudogene-containing genes.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2300. Genetic testing of the *ATP7B* gene for patients with suspected Wilson disease.

Authors:

M. Kyriass, H. Cox; PreventionGenetics LLC, an Exact Sci. company, Marshfield, WI

Abstract Body:

Wilson disease (WD) is an autosomal recessive inborn error of copper metabolism. Clinical disease occurs due to pathogenic genomic variants in the *ATP7B* gene, which encodes the transmembrane ATPase Cu(I) transporting β polypeptide. This leads to a disruption of copper homeostasis, which results in the accumulation of copper in the liver, nervous system, corneas, and other tissues (Weiss 2016. PubMed ID: 20301685; Chanpong and Dhawan. 2022. PubMed ID: 35042319; Dev et al. 2022. PubMed ID: 35586338). Typical features of WD include liver disease, neurologic abnormalities, and psychiatric disturbances. Kayser-Fleischer rings, which occur due to a high level of copper storage throughout the body, are often present in the eyes (Weiss 2016. PubMed ID: 20301685; Chanpong and Dhawan. 2022. PubMed ID: 35042319). Biochemically, low serum ceruloplasmin concentrations and low serum copper levels may be observed, while 24-hour urinary copper levels may be elevated. Additionally, copper levels in the liver are generally elevated (Weiss 2016. PubMed ID: 20301685; IEMbase v2.0.0 at <http://www.iembase.com/disorder/201>). Diagnosis is based on the combination of clinical and biochemical features, and/or biallelic pathogenic *ATP7B* variants identified by molecular analysis (Weiss 2016. PubMed ID: 20301685). However, the clinical picture and age of onset for WD patients can be highly variable. This clinical variability, along with the fact that other diseases and environmental factors can lead to a WD-like phenotype, can complicate diagnosis (Weiss 2016. PubMed ID: 20301685; Sánchez-Monteagudo et al. 2021. PubMed ID: 34572285). Our lab has been analyzing the *ATP7B* gene for over a decade in patients with a suspected diagnosis of Wilson disease, as well as in at-risk family members. During that time, over 200 variants with an interpretation of uncertain significance, likely pathogenic, or pathogenic have been observed. Of the nearly 800 single gene tests performed analyzing *ATP7B*, we reported a diagnostic molecular result in over 20% of cases. In addition to the clearly positive cases, an indeterminate outcome was observed in an additional ~12% of cases while nearly 70% of cases were reported as negative. Overall, these results highlight the benefit of molecular genetic testing for diagnosis of WD patients, while also emphasizing the fact that biallelic *ATP7B* pathogenic variants may not be identified in all patients with a clinical suspicion of WD. Further molecular, laboratory, and/or clinical analysis may be warranted in patients with a clinical suspicion of WD but without any pathogenic variants identified in the *ATP7B* gene.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2301. Genome sequencing enables parent of origin analysis of the X chromosome in probands with Klinefelter syndrome

Authors:

B. Seifert¹, R. Duncan¹, S. Li¹, R. Gore¹, L. S. Schaffer², E. N. Torres², M. N. Similuk¹, W. Bi³, H. Dai³, A. J. Oler¹, R. Ghosh¹, A. Raznahan², M. A. Walkiewicz¹; ¹Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, ²Natl. Inst. of Mental Hlth., NIH, Bethesda, MD, ³Baylor Coll. of Med., Houston, TX

Abstract Body:

Background: Klinefelter syndrome (47,XXY) is characterized by tall stature, hypogonadism, delayed pubertal development, lack of secondary sexual characteristics, and varying neurodevelopmental phenotypes such as learning difficulties and social challenges. Among individuals with a 47,XXY karyotype, prior studies have shown maternal meiotic nondisjunction accounts for approximately 51% of cases, and paternal meiotic nondisjunction accounts for approximately 49% of cases. Whether the majority of sex chromosome nondisjunction occurs in meiosis I or meiosis II is less well understood. Here, we analyzed the parent of origin of the X chromosome via genome sequencing (GS) in individuals with a known diagnosis of Klinefelter syndrome. **Methods:** Trio Parentage/UPD Studies (TRIPS) analysis was conducted on 40 trios undergoing GS and clinical oligonucleotide SNP chromosomal microarray analysis (SNP-CMA). All probands of each trio had a 47,XXY karyotype previously diagnosed through chromosome analysis. Parents of each proband were included for trio GS analysis. Genome sequencing was performed by short-read sequencing and clinical SNP-CMA performed by comprehensive high-resolution array that detects uniparental isodisomy (UPiD), regions of homozygosity (ROH), and copy number variants (CNVs) underlying neurodevelopmental disorders. **Results:** Among 40 trios undergoing GS and clinical SNP-CMA analysis, 17/40 probands (42.5%) showed X chromosome aneuploidy consistent with paternal meiotic nondisjunction. Twenty-one of 40 (52.5%) probands showed X chromosome aneuploidy consistent with maternal heterodisomy. Two of 40 probands (5.0%) showed X chromosome aneuploidy consistent with maternal isodisomy, suggestive of a maternal meiosis II or post-zygotic mitotic error. The results showing maternal isodisomy were confirmed by clinical SNP-CMA. Within the individuals with maternal heterodisomy of the X chromosome, 1 individual (2.5%) showed marker patterns suggestive of segmental isodisomy within the short arm of the X chromosome, which is currently under investigation. **Conclusions:** Similar to previous studies, our analysis showed ~52% of probands with a 47,XXY karyotype to result from maternal meiotic nondisjunction and ~43% to result from paternal meiotic nondisjunction. Our data demonstrate the power of genome sequencing in determining the parent of origin of the extra X chromosome in individuals with Klinefelter syndrome. Ongoing work aims to correlate neurodevelopmental phenotypes in probands with the parent of origin of the additional X chromosome in these probands with Klinefelter syndrome.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2302. Genome Sequencing Enables Precision Clinical Care in Genetic Leukoencephalopathies

Authors:

K. Muirhead¹, J. L. Hacker¹, F. Gavazzi¹, S. Woidill¹, O. Sherbini¹, G. Helman¹, J. Kramer-Golinkoff¹, G. H. Velasquez², A. Pizzino¹, J. Schmidt¹, R. Taft³, A. Vanderver¹; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Univ. of California San Francisco, San Francisco, CA, ³Illumina Inc, San Diego, CA

Abstract Body:

OBJECTIVE: Leukoencephalopathies are rare, heritable white matter disorders with variable presentations that encompass more than thirty distinct phenotypes. While recent studies have shown the efficacy of genome sequencing (GS) in this population, the impact of a GS diagnosis on clinical care has not yet been investigated. **METHODS:** The cohort consisted of 165 individuals with clinically ascertained leukoencephalopathies who were subsequently enrolled in the IRB-approved Myelin Disorders Biorepository Project and received clinical GS. Clinical data from pre- and post-GS office notes was extracted and assessed for etiologic-specific screening, treatment and referrals to specialist providers after achieving a GS diagnosis to assess for outcomes of care. **RESULTS:** As a result of GS, 68 individuals (41%) achieved a positive test finding, 28 individuals (17%) had suspicious variants in a gene(s) of interest reported and 69 individuals (42%) had negative or non-diagnostic results. Of the 68 individuals who achieved a definitive diagnosis, 57 had sufficient clinical data available to investigate the impact of a GS diagnosis on clinical care. Within this group, 54 (95%) were referred to specialist providers for disease monitoring, 38 (67%) received additional targeted screening and 12 (21%) were eligible for disease-specific treatment. **CONCLUSIONS:** Individuals with clinically ascertained leukoencephalopathies who achieved a diagnosis via GS received precision clinical care inclusive of specialist referral, screening and condition-specific treatment. These data suggest that GS should continue to be considered as a first-line test for this population. Future implementation studies may be necessary to understand barriers to adoption and widespread deployment.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2303. Genome-wide investigation of potentially pathogenic copy number variants & mechanisms fomenting their origins.

Authors:

Z. Dardas¹, R. Duan¹, D. Marafi², D. Calame¹, J. Fatih¹, M. Dawood¹, H. Du¹, C. M. Grochowski¹, J. E. Posey¹, C. M. B. Carvalho³, J. R. Lupski¹; ¹Baylor Coll. of Med., Houston, TX, ²Kuwait Univ., Safat, Kuwait, ³Pacific Northwest Res. Inst., Seattle, WA

Abstract Body:

Deletions and duplications of chromosomal segments (copy number variants, CNVs) are a major source of variability between personal genomes and are a significant contributor of pathogenic rare variants to genomic disorders, as well as to both human gene & genome evolution. Despite these significant roles, the contribution of CNV to human disease and their underlying mechanisms are not fully understood. Herein, we aim to uncover the contribution of CNVs to a variety of rare Mendelian diseases in a large cohort of patients (>13,000) and better unfold the molecular mechanisms driving their formation. Moreover, as there is a possibility that genes in a family may have derived from one ancestral gene which became duplicated into different copies (paralogs) during evolution, we also aim to identify CNV in gene families including the tubulin superfamily. Using bioinformatics tools for CNV prediction from exome data (XHMM and HMZ/HTZ Dup/Del Finder), we identified a total of 50 potentially pathogenic novel exonic CNVs in neurodevelopmental disease (NDD) cohort (homozygous (16) duplications, *de novo* heterozygous (20) deletions and (14) duplications). A custom designed high-density array comparative genomic hybridization (aCGH) was utilized to further validate the presence of the predicted CNV and better delineate its precise genomic interval (start and stop positions) and breakpoint junction sequence. Of note, 9 candidate novel NDD-associated genes were identified among the identified CNVs including *SYNRG*, *DDX15*, *HNF1B*, *DUSP14*, *CYP4F11*, *CYP4F2*, *NID1*, *NEDD4*, and *C21orf57*. Interestingly, several genes that are members of gene families were detected within the candidate CNVs including *DDX15*, *HNF1B*, *CYP4F11*, *CYP4F2*, *KIF1A*, and *TUBB4A*. Additionally, multiple *de novo* large deletions and duplications were predicted by XHMM on chromosomes 5 and 18, respectively, in an unsolved NDD case. Array CGH revealed a terminal 14.4 Mb deletion in Chr5p and a terminal 21.5 Mb duplication in Chr18q suggesting a novel unbalanced translocation between Chr5p and Chr18q. Moreover, we fully characterized the CNV structure in three cases in which we confirmed to be tandem duplications with blunt fusion suggesting non-homologous end joining (NHEJ) as their driving mechanism. We also detected a complex genomic rearrangement (CGR) involving an apparent quadruplication-quintuplication event, which warranted further investigation by other technologies including long-read whole genome sequencing and optical mapping. This work has helped to reveal new potentially disease-causing genes, the disease pathogenesis, and provide insight into their associated SV mutagenesis mechanisms.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2304*. Genomic sequencing generates less uncertainty than panel-based testing: Results of over 1.5 million tests

Authors:

H. Rehm¹, J. T. Alaimo², S. Aradhya³, P. Bayrak-toydemir⁴, H. Best⁴, J. G. Buchan⁵, L. Conlin⁶, S. Das⁷, D. del Gaudio⁷, C. DiVincenzo⁸, L. Guidugli⁹, M. Hammer⁹, S. Hofherr¹⁰, V. Jobanputra¹¹, I. D. Karbassi⁸, M. A. Kelly¹², M. L. Kluge¹³, C. R. Marshall¹⁴, Y. Meng¹⁵, C. Neckelmann¹⁵, K. J. Rasmussen¹³, E. Spiteri¹⁶, C. Stanley¹⁷, B. Thomas¹⁸, M. Towne¹⁹, R. J. Taft²⁰, H. Zouk²¹; ¹Massachusetts Gen. Hosp./Broad Inst., Boston, MA, ²Children s Mercy Hosp., Kansas City, MO, ³Invitae, San Francisco, CA, ⁴Univ. of Utah/ARUP Lab., Salt Lake City, UT, ⁵Washington Univ, St. Louis, MO, ⁶Children s Hosp. of Philadelphia, Philadelphia, PA, ⁷Univ. of Chicago, Chicago, IL, ⁸Quest Diagnostics, Marlborough, MA, ⁹Rady Children's Inst. for Genomic Med., San Diego, CA, ¹⁰Gene Dx, Gaithersburg, MD, ¹¹New York Genome Ctr./Columbia Univ. Irving Med. Ctr., New York, NY, ¹²HudsonAlpha Clinical Services Lab, Huntsville, AL, ¹³Mayo Clinic, Rochester, MN, ¹⁴The Hosp. for Sick Children, Toronto, ON, Canada, ¹⁵Fulgent Genetics, Temple City, CA, ¹⁶Stanford Med., Stanford, CA, ¹⁷Variantyx, Framingham, MA, ¹⁸Illumina, San Diego, CA, ¹⁹Ambry Genetics, Aliso Viejo, CA, ²⁰Illumina Inc, San Diego, CA, ²¹Mass Gen. Brigham/Harvard Med. Sch., Boston, MA

Abstract Body:

A growing fraction of the population is receiving genetic testing, yet DNA sequencing often uncovers variants of uncertain significance (VUSs). Physicians and patients are often ill-prepared to manage VUSs and insurers are concerned about downstream costs. To understand the contribution of panel-based and genomic (exome and genome) testing to the generation of inconclusive results, data on rates of VUS return and diagnostic yield were collected from over 1.5 million tests from 19 clinical labs from Jan 2020 - Dec 2021. Results were delineated by panel size (2-10, 11-25, 26-50, 51-100, 101-200, >200 genes) or trio vs. less-than-trio for genomic testing, and for a subset, disease area across 12 broad indications. We found a larger and statistically significant difference ($p < 0.0001$) in the rate of inconclusive test results due to VUSs from multi-gene panel tests (32.4%; 470,834/1,452,897) compared to the rate from genomic tests (30.1%; 11,100/36,845). For panel tests, the rate of inconclusive results correlated with panel size ranging from 5.9% for 284,953 panel tests of 2-10 genes to 76.2% for 76,800 panel tests >200 genes. The rate of inconclusive results due to VUSs from less-than-trio genomic testing (36.8%; 5,903/16,049) was statistically significantly ($p < 0.0001$) higher than the rate from trio-based genomic testing (25.0%; 5,197/20,796). Although an increased rate of inconclusive results were observed in genome tests (32.4%; 1,646/3,432) compared to exome tests (29.8%; 9,454/22,313), this correlated more strongly with testing laboratory as most only ran one platform (exome or genome) and inconclusive rates per lab ranged from 16.2% to 72.5%. Although inconclusive rates for most disease areas directly correlated to panel size, rates were higher than expected for panel size for hematology/rheumatology/immunology tests (61%; average 21 genes) and lower than expected for panel size for cardiology tests (36%; average 68 genes). In contrast, little correlation to panel size was observed for diagnostic yield which ranged from a low of 7% for germline cancer (92,560/1,236,963; average 21 genes) to a high of 36% for dermatology (195/536; average 4 genes). In summary, the largest source of inconclusive results was multi-gene panel testing and was lower when deploying genomic testing. This is best explained by current practices of obligatory reporting of all VUSs in panel-based testing compared to genomic testing where correlation with phenotype constrains which VUSs are reported. These results may inform future reporting practices and payer coverage in genetic and genomic testing and a heightened appreciation for genomic testing interpretation methods.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2305. GMCK-RD a success story: Implementation of genomic medicine in Stockholm healthcare region, update on first 10,000 samples

Authors:

A. Lindstrand¹, H. Stranneheim², K. Lagerstedt Robinson¹, N. Lesco¹, M. Kvarnung¹, D. Nilsson², M. Engvall¹, J. Eisfeldt², H. Malmgren¹, C. Chiara², A. Jemt², S. Vonlanthen¹, P. Marits¹, M. Johansson Soller¹, A. Nordgren¹, V. Wirta², A. Wedell¹, GMCK-RD; ¹Karolinska Inst.t, Stockholm, Sweden, ²Karolinska Inst.t / SciLifeLab, Solna, Sweden

Abstract Body:

In healthcare clinical genetics is transitioning to clinical genomics and especially rare disease diagnostics is increasingly done through panel, exome and whole genome sequencing. At Karolinska we have formed Genomic Medicine Center Karolinska Rare Diseases (GMCK-RD), a joint unit between healthcare and academia, enabling large-scale genome sequencing of patients. Our transition to genomic medicine has been a success with over 10,000 individuals across a broad spectrum of rare diseases analyzed, enabling a diagnosis as well as personalized prediction, prognosis, and treatment in thousands of patients. GMCK-RD brings together experts from various medical disciplines with clinical geneticists, bioinformaticians and researchers. To facilitate implementation, we have developed a number of bioinformatic tools and processes covering steps as variant calling, workflow management, variant prioritization and interpretation, data sharing and quality assurance. Challenges include big data processing, ethical and legal issues as well as rigorous quality control while at the same time enabling continuous development of analysis pipelines and tools. Our analysis covers detection and interpretation of SNVs, INDELS, uniparental disomy, CNVs, balanced structural variants, and short tandem repeat expansions. Ongoing developemet projects include: (i) detection of mobile elements, (ii) Implementation of RNA-seq for diagnostic multiomic analyses and (iii) exploring novel reference genomes. The rapid increase in genomic testing also brings along a high need for scaling up interpretation, including through recruitment and training of highly specialized “genomicists” and development of new support solutions. We also need integrated units where highly specialized clinicians work closely together with laboratory experts, enabling patient selection, correct interpretation and validation of findings, and rapid translation to individualized treatment. This interactions have also enabled a large number of new disease gene discoveries. By overcoming these issues, we at GMCK-RD have moved healthcare in our region towards precision diagnostics and precision medicine.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2306. High diagnostic yield from clinical genome sequencing supports genome sequencing as first-tier genetic test: Evidence-based from 2002 index cases

Authors:

F. Guo, R. Liu, A. Mathur, C. Da Silva, B. Nallamilli, C. Collins, L. Bean, N. Gurujju, X. Chen-Deutsch, R. Yousaf, Z. Ma, J. Balciuniene, M. Hegde; PerkinElmer Genomics, Pittsburgh, PA

Abstract Body:

Genome sequencing (GS) is one of the most comprehensive tests that interrogate single nucleotide variants (SNV), copy number variants (CNV), mitochondrial variants, repeat expansions, and structural variants in one single assay. Despite the clear technical superiority, few studies are available regarding the diagnostic utility of its clinical application. In this study, we systematically evaluated 2002 consecutive clinical GS cases performed in our laboratory since 2017 to explore the diagnostic utility of clinical GS. Among these 2002 clinical GS cases, there were 1429 pediatric patients (633 singleton and 796 trios) and 573 adult patients (393 singletons and 180 trios). Whole blood, saliva, isolated DNA, and dry blood spot (DBS) samples accounted for about 25.6%, 48.2%, 12.0%, and 14.1% of samples, respectively. Saliva was the most common sample type, accounting for 41% and 55.4% in pediatric and adult cases, respectively. In contrast, DBS was the most common sample type for the pediatric cohort with 36.5% and 17.3% in pediatric singleton cases and all pediatric cohorts, while 6.7% of the adult samples were DBS. Overall, more than 30% of cases were previously tested by another laboratory but still undiagnosed. The overall assumed diagnostic yield was 38.1% (762/2002), with 40.3% (255/633) in pediatric singleton cases, 40.5% (322/796) in pediatric trio cases, 30.3% (119/393) in adult singleton cases and 36.7% (66/180) in adult trio cases. In addition, a relatively higher diagnostic yield of 40.4% in pediatric cases versus 32.3% in adult cases was observed in our cohort. Among the diagnostic findings, 82.3% of cases had SNVs only, including 83.9% in the pediatric cohort and 75.7% in the adult cohort; 13.2% of cases had a CNV ranging from a single exon deletion to aneuploidy in 12.3% of the pediatric cohort and 15.7% of the adult cohort. Eighteen cases had one SNV and one CNV in the same gene associated with recessive diseases consistent with the patient's clinical findings. Fifteen cases had diagnostic findings in the mitochondrial genome consistent with mitochondrial disease. We also detected two pediatric cases with SMA and three pediatric cases with positive repeat expansion screening consistent with the patient's clinical findings, which would have been missed by other sequencing platforms like exome sequencing. Importantly, around 3% of diagnostic cases had findings consistent with multiple different clinical diagnoses.

In summary, we demonstrate the clinical utility of GS within a large-to-date clinical GS cohort along with consistency in the use of DBS cards which are widely used in newborn screening.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2307. Identification of a small duplication in the upstream noncoding enhancer of *SOX9* in a family with 46,XX disorder of sexual development by clinical whole genome sequencing.

Authors:

S. Sajan¹, C. Brown¹, L. Davis-Keppen², K. Burns², E. Royer², A. Kesari¹, D. Perry¹, R. Taft¹; ¹Illumina Inc, San Diego, CA, ²Sanford Children's Hosp., Sioux Falls, SD

Abstract Body:

Patients with rare undiagnosed genetic disease often undergo years of testing before receiving definitive molecular diagnoses. Here, we describe a family with multiple individuals with *SRY*-negative 46,XX testicular disorder of sexual development (DSD) referred for clinical whole genome sequencing (cWGS) through the Illumina's iHope program, comprising the affected proband, the clinically unaffected mother, and an affected maternal uncle. Both affected individuals identify as male. The proband is a 9-year-old male with ambiguous genitalia at birth, elevated testosterone levels, proximal scrotal hypospadias, phallus length of 2.6 cm, and palpable gonads. Also noted were urinary infections, short stature, ADHD and anxiety, whereas development was normal and intelligence in the superior range. A previous 64-gene panel test for ambiguous genitalia was negative. The mother's 30-year-old brother also has *SRY*-negative 46,XX testicular DSD with features including perineal hypospadias, bifid scrotum, palpable gonads with dysgenetic testes, mullerian structures in the scrotum, as well as hypogonadism and short stature. Trio case analysis for this particular family structure did not identify any clinically relevant variants using the standard clinical laboratory protocol. However, a detailed literature review revealed noncoding copy number variants upstream of the *SOX9* gene known to cause DSD that were excluded from the original analysis. We therefore manually inspected the cWGS data upstream of *SOX9* which resulted in the identification of a small ~3.5 kb duplication of 17q24.3, seq[GRCh37]dup(17)(q24.3)chr17:g.69480152_69483631dup, located about 634 kb upstream of the *SOX9* gene impacting the well-known XYSR enhancer of this gene. The smallest duplications of this enhancer that had previously been reported were 23.9 kb and 24.2 kb in two patients whose minimal critical region of overlap was ~5.2 kb containing experimentally validated binding sites for the testis-specific transcription factors SOX9 and SF1. The 3.5 kb duplication reported in our proband is entirely contained within this minimal critical region and is the smallest such duplication reported to-date in patients with 46,XX DSD. This duplication was present in everyone in the trio including the unaffected mother, which is consistent with one previous report of incomplete penetrance. These data show that cWGS as a first-tier test can benefit patients with suspected genetic disorders and also underscores the importance of analyzing cWGS noncoding variants.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2308. Identification of Balanced and Unbalanced Complex Chromosomal Rearrangement Involving Chromosomes 2, 4,5,18 and 21 in a family

Authors:

A. Venkateshwari, S. A, S. M; Inst. of Genetics and Hosp. for Genetic Dis, Hyderabad, India

Abstract Body:

Introduction: Complex Chromosomal Rearrangements (CCR's) are the rare structural chromosomal abnormalities, which involve more than two non-homologous chromosomes with two or more breakpoints. If it is a balanced rearrangement, there will be no significant phenotypic abnormalities detected in the individuals. Many clinical symptoms related with CCR's are congenital malformations, dysmorphism, mental retardation, infertility, recurrent pregnancy loss, bad obstetric history etc. **Material and methods:** A consanguineous (uncle-niece) couple with a history of secondary recurrent pregnancy loss with two spontaneous abortions and the third one was a full-term delivery of female child who died at 7 months of age was referred to institute for genetic analysis. Couple karyotyping revealed a complex chromosomal balanced translocation in female partner with five chromosomes involving five breakpoints with der2,t(2,18)t(4,5,18)t(21,4) karyotype and the male partner karyotype was normal with 46, XY chromosomal constitution. Further, Spectral karyotyping analysis was performed to confirm the CCR. Microarray analysis revealed no gain or loss of chromosomal material in the female partner. The child was presented with dysmorphic features, developmental delay, severe respiratory distress, abdominal pain and hepatomegaly at the time of birth. Her cytogenetic analysis revealed an unbalanced 46, XX, der(2) chromosome and her microarray analysis showed a deletion of 3474.756kbp at cytoband 2q37.3 which was denoted as pathogenic. **Conclusion:** The present case study highlights a novel familial translocation with CCR in a couple having bad obstetric history. The transmission of CCR's in fetus may result in either abortions or child with chromosomal defects and result in congenital anomalies leading to death of the child. To the best of our knowledge this is the first case to be reported with these breakpoints involving 5 different chromosomes from our ethnic group. Hence, the couple was advised for genetic counselling and pre-natal diagnosis in future pregnancies.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2309. Identification of somatic Neurofibromatosis 1 mosaicism using targeted next-generation sequencing of café-au-lait-macules

Authors:

E. Korpershoek¹, R. R. van Minkelen¹, M. Nellist¹, W. N. M. Dinjens², M. van Vliet¹, A. L. Mooyaart², T. J. van Ham¹, Y. van Ierland¹; ¹Erasmus MC, Univ. Med. Ctr. Rotterdam, Dept. of Clinical Genetics, Rotterdam, Netherlands, ²Erasmus MC, Univ. Med. Ctr. Rotterdam, Dept. of Pathology, Rotterdam, Netherlands

Abstract Body:

Purpose Detection of somatic mosaicism in blood can be challenging, even with a sensitive variant detection technique such as Next Generation Sequencing (NGS). For individuals with (segmental) neurofibromatosis type 1 (NF1), mutation analysis of tumors can be an alternative strategy for the identification of somatic NF1 mosaicism, although this is not possible in individuals without syndrome-related tumors. Here we describe a patient with a segmental phenotype, presenting solely with Café-au-lait macules (CALMs) and freckles, in which no NF1 mutation was identified in DNA isolated from peripheral blood. **Methods** DNA isolated from blood and four independent CALMs was investigated by targeted Ion Torrent-based NGS. In addition, cells were cultured from a CALM and from normal skin. RNA isolated from these cultured cells was subjected to transcriptome analysis (RNAseq). **Results** Targeted NGS revealed that three of the four investigated CALMs showed the pathogenic variant NM_000267.3(NF1):c.3916C>T, p.(Arg1306*), which was absent in DNA derived from blood. RNAseq confirmed expression of the NM_000267.3(NF1):c.3916C>T, p.(Arg1306*) variant in RNA from the cultured fibroblasts from a CALM (variant allele frequency of 8%), while no expression was seen in RNA from the cultured fibroblasts from normal skin. **Conclusion** Our results indicate that targeted NGS of independent CALMs can assist in the detection somatic mosaicism in patients with segmental NF1 for whom no variant was identified in peripheral blood DNA.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2311. Importance of the reclassification of the clinical significance of new genetic variants in non-classical mucopolysaccharidosis type IV A

Authors:

R. Gomez, L. Moreno; Univ. Libre, Cali, Colombia

Abstract Body:

In mucopolysaccharidosis IV A, there is a deficiency of lysosomal hydrolase, N-acetylglucosamine-6-sulphate sulphatase enzyme, leading to the accumulation of keratan sulfate and chondroitin-6-sulfate. It affects men and women with equal frequency, the prevalence ranges between 1/40,000 and 1/200,000 births. In Colombia, records are scarce; a frequency of 0.68/100,000 live births is reported. The c.1431G>A variant is the most frequent with 60% of the alleles, followed by the benign p.H36= variants and the intronic variant 634-19G>A with 41% and 34%, respectively. Children have no clinical findings at birth, progressive bone and joint involvement leading to short stature and disabling pain. Spinal cord compression is a common complication that results in neurological deterioration. A male patient with a clinical picture of muscle weakness in the lower limbs and difficulty walking is described. MRI of the spine with alteration in the configuration of the vertebral bodies from T9 to L5 with a tendency to platyspondyly and anterior wedging with a cortical peak at the L2 level. Blood drop collected on GALNS filter paper < 0.03 $\mu\text{mol/l/hr}$ (normal \geq 0.39 $\mu\text{mol/l/hr}$). With subsequent confirmation in leukocyte sample: Galactose-6-sulphate sulphatase: 0.10 (normal = 2.6 - 35.9). Molecular study of the GALNS gene, two variants: One of pathological significance: c.901G>T, Protein change: p. Gly301Cys, heterozygous, Inheritance: autosomal recessive and another variant of uncertain significance: c.1088T>C, Protein switch: p. Ile363Thr: Heterozygous, Inheritance: Autosomal recessive. A bioinformatic study of in silico technology is carried out, including the use of prediction software such as Sorting intolerance from tolerant, mutation taster, UMD-predictor, functional analysis through hidden Markov models, polymorphism phenotyping V2, and protein variation effect analyzer. With a change in clinical significance from a variant of uncertain significance: c.1088T> C to pathogenic. Clinical, paraclinical, enzymatic, and molecular correlation and its phenotype, endotype, and genotype correlation are performed.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2312. Inadequate parental sequencing depth burdens African ancestry patients with inaccurate *de novo* inheritance calls.

Authors:

S. Choi¹, P. White^{1,2}, B. P. Chaudhari^{1,2}; ¹Nationwide Children's Hosp., Columbus, OH, ²The Ohio State Univ., Columbus, OH

Abstract Body:

The rapid genome sequencing (rGS) study is a pilot study which seeks to validate and understand the feasibility of internal rGS compared to standard of care clinical testing (send-out rapid exome or genome sequencing; rES/rGS). The rGS study has enrolled 120 cases including 92 trios (64 clinical rES and 28 clinical rGS comparators). Clinically reported variants and inheritance are compared to research rGS findings to inform test development.

All variants reported as *de novo* on clinical rES were concordant with research rGS (66 variants). When clinical standard of care transitioned to rGS, we noted, in 14 consecutive trios, 8/25 clinically reported *de novo* variants appeared to be inherited from an unaffected parent on research rGS. Sanger sequencing confirmed inheritance in all variants tested (n=6). These variants appeared more commonly in probands of self-reported African (4/5) vs. European (1/9) ancestry (p=0.023, Fisher's exact test). We sought to identify the cause of these inaccurate inheritance calls as well as the apparent inequity by ancestry.

We found that parental depth in clinical rGS was low (median 13.6x, range 5.6x-23.4x) compared to research testing (median 35.6x, range 23.4x-44.8x). We used the distribution of heterozygous, phenotype-related, rare, deleterious variants in genes known to cause autosomal dominant disorders, along with the estimated probability of failing to call the alternate allele in a heterozygous parent, conditional on sequencing depth (assuming a minimum of 3 reads are required to call a variant) to estimate the burden of inaccurate *de novo* inheritance calls when sequencing parents at low depth. At a parental depth of 10x, the positive predictive value (PPV) for phenotype-associated *de novo* variants is estimated at only 47%. At 19x, the PPV becomes >99.9%. The number of rare, phenotype-related, inherited variants in genes known to cause autosomal dominant disorders that could be inaccurately called as *de novo* was significantly greater in those of self-reported African ancestry (median 43, range 29-69) vs. European ancestry (median 20, range 9-25) (p=0.0001, M-W test).

The accuracy of variant calling is directly affected by sequencing depth. Low parental sequencing depth leads to missed variant calls in parents, inaccurate identification of "*de novo*" variants in probands, and, potentially, incorrect variant interpretation. As the absolute number of variants inaccurately called *de novo* is also a function of the number of rare variants, individuals of African ancestry are most affected. Based on our findings, the clinical lab updated their procedures and no incorrect calls of *de novo* inheritance have been reported since.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2313. Increased diagnostic yield from negative whole genome-slice panels using automated reanalysis.

Authors:

S. Berger¹, A. Cohen¹, J. LoTempio², E. Andrew³, G. Mas⁴, S. Marmolejos¹, J. Albert¹, A. Kahn-Kirby⁴, E. Delot⁵, V. Fusaro⁴, E. Vilain⁶; ¹Children's Natl. Med. Ctr., Washington, DC, ²Children s Natl., Washington, DC, ³Children's Natl., Washington, DC, ⁴Invitae, San Francisco, CA, ⁵Children s Natl. Res. and Innovation Campus, Washington, DC, ⁶Children s Natl. Med. Ctr., Washington, DC

Abstract Body:

Purpose: Panel testing, sequencing predetermined lists of genes selected for specific clinical indications, is standard practice in genetic medicine. These tests often are not diagnostic because the relevant gene was not included on the panel. The process to add new genes to traditional hybridization capture panels for clinical use is slow and laborious. Recently, many clinical laboratories have been using next-generation sequencing to offer more flexible panels by validating an exome or whole genome platform and then report on panel regions using bioinformatic analyses. As additional clinical information presents it is possible to reanalyze the existing data without collecting a new sample from the patient or performing additional wet lab work. Typically, the type of genetic test ordered is dictated by clinical guidelines, institutional priorities, payor policies, and patient costs. This often results in ordering the cheapest and smallest panel that meets an immediate diagnostic need. We sought to assess the incremental diagnosis rate for cases lacking a conclusive diagnosis after panel-based testing that was performed using whole-genome sequencing.

Methods: We obtained 15 negative clinically generated whole-genome data cases referred to the Pediatric Mendelian Genomics Research Center, an NHGRI-funded GREGoR site. Previous panel-based testing was performed at Children's National Clinical Molecular Laboratory. Whole-genome variant call files were analyzed with Moon software, an AI phenotype/variant prioritization tool.

Results: In 33% of the cases (5 out of 15) a reportable variant was highly ranked by Moon. Of the five cases, one variant was originally missed, despite having the gene on the original panel, because it was a complex structural rearrangement with deep intronic breakpoints falling outside the reporting regions. In the remaining four cases, actionable variants were discovered in genes not included in the original panel.

Conclusion: Whole-genome or exome-based panels offer the ability to improve diagnostic yield by performing reanalysis as guidelines are updated or as new phenotypic information becomes available. As whole-genome sequencing costs continue to decrease, we anticipate this will become a more tractable option for clinical laboratories.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2314. Increased Diagnostic Yield With Long Read Sequencing In Patients With Undiagnosed Neurodevelopmental Disorder

Authors:

S. Jamuar¹, J. Goh¹, J. Teo², A. Wang¹, S. Ng³, M. Chin⁴, K. Chua⁵, S. Au⁴, X. Chan⁴, J. Lim³, S. Kam³, H. Law¹, A. Wenger⁶, W. Lim²; ¹KK Women's and Children's Hosp., Singapore, Singapore, ²SingHealth.Duke-NUS Inst. of Precision Med., Singapore, Singapore, ³KK Women's and Children's Hosp., Singapore, Singapore, ⁴Pacific BioSci.s, Singapore, Singapore, ⁵PacBio, Singapore, Singapore, ⁶Pacific BioSci.s, Menlo Park, CA

Abstract Body:

Background: Singapore Undiagnosed Disease Programme, established in 2014, has performed whole exome/genome sequencing using short read technology in over 600 patients so far, achieving a diagnostic yield of 38.8%. In a sub-cohort of 82 patients with neurodevelopmental disorder, we achieved a diagnostic yield of 43%.

Aim: Given the limitations of short read sequencing in detecting structural variants, repeat expansion disorders, etc., we aimed to study the potential increase in diagnostic yield in our cohort with the use of long read sequencing.

Method: We identified 10 families where a genetic disorder was strongly suspected. This included a personal history of severe developmental delay/ intellectual disability and/or presence of growth abnormalities and/or multiple malformations; or a family history of affected relative(s). We performed HiFi long read sequencing using Sequel IIe system on high molecular weight genomic DNA extracted from peripheral leucocytes. We performed de novo assembly to identify structural variants, as well as alignment to human reference genome for variant calling using established bioinformatic pipelines. Variants were reviewed at a multidisciplinary team meeting, and candidate variants were selected for validation through orthogonal technologies.

Results: Out of the 5 samples analysed so far, we have identified a candidate variants in 2 (40%) of these families. One is a 1.83kb deletion of exon 1 of *DYRK1A* (MIM# 600855) in a child with global developmental delay and facial dysmorphism. In another child with progressive neurodegenerative disorder, gaze palsy and ataxia, a 7kb repeat expansion in *BEAN1* gene (MIM# 612051), associated with spinocerebellar ataxia 31, was identified. Analysis on the remaining 5 samples is ongoing.

Conclusion: Our initial pilot of long read sequencing on patients with undiagnosed neurodevelopmental disorder further increased our diagnostic yield by identifying variants that are typically missed on traditional short read sequencing.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2315. Increasing global access to clinical whole genome sequencing for under-resourced undiagnosed individuals

Authors:

S. Terry¹, J. L. Ortega¹, R. J. Taft^{2,1}, S. Moosa³, J. Green¹, V. Rajan¹; ¹Genetic Alliance, Damascus, MD, ²Illumina Inc, San Diego, CA, ³Stellenbosch Univ., Cape Town, South Africa

Abstract Body:

Background: The search for a diagnosis by individuals with undiagnosed conditions, sometimes called the diagnostic odyssey, is burdensome for individuals, families, communities, and society. This burden is compounded when the resources necessary to end the diagnostic odyssey are not available, as is the case in the majority of low and middle-income countries (LMICs). Genetic Alliance's iHope™ Genetic Health (iGH) is a global program designed to help patients from LMICs and/or low- and moderate-income households achieve a diagnosis through clinical whole genome sequencing (cWGS), education, and support. Leveraging contributions from Illumina, AWS, LunaPBC, and others, iGH connects patients and their clinicians with laboratories offering cWGS free of charge. **Objective:** Genetic Alliance issued a Request for Information (RFI) to determine needs of patients, laboratories, clinical sites that care for rare disease patients. We aimed to understand the opportunities and challenges for cWGS for under-resourced individuals and their clinicians, the global capacity for cWGS, and identify novel care pathways. **Methods:** Genetic Alliance created an RFI with questions tailored to various stakeholder groups, including laboratories, researchers, clinicians, families, and advocacy organizations. The RFI was sent by direct email to genetic testing labs, professional societies, and academic and patient organizations. In addition, it was distributed through Genetic Alliance's social media channels. Responses were collected during March 2022. We reviewed the responses and synthesized a summary categorized by major stakeholder groups. **Results:** We received 28 responses from the stakeholder groups from nine countries. The overall themes from all stakeholders are similar: the main barriers to providing cWGS are cost, coverage policies and reimbursement, and clinician and patient education. Cost: the high costs of cWGS make routine utilization difficult. Reimbursement: cWGS is rarely covered by payors. Education: patients require cultural-background appropriate genomic education, particularly prior to testing; medical practitioners are often unaware of both the availability and potential impact of genomic testing; and the availability of specialist staff, including medical geneticists, genetic counselors and bioinformaticians, is limiting uptake and availability. **Discussion:** This information has helped define iGH's near-term goals and enable resource investment where it will have the most impact. iGH will release its first round of RFPs in 2022, with initial disbursements of grants to participating laboratories in LMIC shortly thereafter.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2316. Infinium Global Screening Array for CNVs detection.

Authors:

L. Vieira¹, G. Carvalho¹, Y. Gasparini², V. Almeida¹, B. Wolff¹, A. Nascimento¹, L. Kulikowski³; ¹Faculdade de Med. da Univ.e de São Paulo, São Paulo, Brazil, ²Univ.e de Sao Paulo, Sao Paulo, Brazil, ³Univ.e de Sao Paulo, São Paulo, Brazil

Abstract Body:

Copy number variations (CNVs) play important role in the development of many complex phenotypic traits. Cytogenomic CNVs analyses using Infinium CytoSNP-850K bead chips (Illumina) providing comprehensive genome-wide coverage across a screen of 850,000 probes and commonly applied at routine laboratories. However there are other platforms for genotyping assays, such as the Global Screening Array BeadChip (GSA), which uses the same beadarray technology to detect single-base polymorphisms (SNP) with 650,000 probes. In this sense, we compared the effectivity of CNVs detection by GSA platform with CNVs results obtained from with samples previously genotyped by the gold standard Infinium CytoSNP-850K techniques and also Multiplex Ligation-dependent Probe Amplification (MLPA) assays. We studied 21 samples previously evaluated by gold standard platforms for CNV detection, 7 by MLPA assays and 14 by CytoSNP-850K. All samples were submitted to the GSA assay and analyzed by the CNV partition plug-in of the Genome Studio software provided by Illumina. The genomic results were associated with the clinical phenotypes. The CNVs obtained by both platforms, GSA and CytoSNP, were fully compatible, showing only expected variations in breakpoints at regions with different probe coverage. The GSA assay showed higher resolution compared with samples genotyped by MLPA, as the coverage of probes in deleted and/or duplicated regions is greater than that of MLPA probes. Thus enabling to suggest the breakpoints of the deletions and the amount of genes affected. In addition to the CNVs, it was possible to visualize clinically relevant regions such as the ROH regions. Therefore, the GSA assay is effective for the detection of CNVs when compared with MLPA assays for genome screening, despite its lower resolution. The use of GSA can also represent an excellent cost-benefit compared to CytoSNPs, especially considering the search for pathogenic CNVs in the laboratory routine.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2317*. Integrating population-scale data into structural variant visualization

Authors:

M. Chowdhury, R. Layer; Univ. of Colorado, Boulder, CO

Abstract Body:

Structural variation (SV) has been linked to various diseases, but use of SV information remains low in the clinical setting which relies mainly on short-read sequencing. The main barrier to adoption is the overabundance of false positive calls made by SV detection tools. Given the uncertainty of SV callers, many tools have been developed to help validate SV calls by visualizing the genomic regions around potential SVs.

Visualization tools are of great importance in the validation of SV callsets. It is often possible to accept or reject a potential SV by examining the sequence alignments in and around the reported genomic region. Viewing raw alignments in a tool like IGV is difficult and time consuming, however. To alleviate this issue we developed Samplot, a tool for visualizing genomic regions that contain an easy to understand view of genomic coverage as well as a compact representation of paired/split read alignments. To enhance the utility of these images for manual SV curation, we developed SVPlaudit, a cloud based framework for collaborative annotation of putative SV callsets. Given the ease of human based curation with Samplot images, we developed Samplot-ML, a machine learning model trained to classify Samplot SV regions, to enable automated filtering of potential false positive SVs.

We now seek to extend the utility of Samplot into the realm of population level data using our recently developed structural variant index (STIX). STIX extracts discordant/split read alignments (often used by short read SV callers to make their predictions) from huge cohorts of whole genome sequencing data and allows one to determine the density of SV evidence across the indexed population given a simple interval query. The new integrated visualization displays the original sample level information with a population view of discordant/split read alignments returned from a STIX query of the SV breakpoints. This enables us to visually compare sample level alignments in a genomic region against the background distribution in one or more STIX populations. For example: if we are interested in calling somatic SVs in a tumor sample, we can use the presence of STIX evidence in a normal tissue cohort to reject a SV as likely germline.

We see the development of our visualizations and automated curation tools as an important two-pronged approach to the SV problem. Determining elements of useful visualizations has informed the development direction of tools like Samplot-ML and vice versa. We foresee this new integrated view of sample and population level data will help improve the quality of SV callsets in the short term and serve as a base for future automated SV curation tools.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2318*. Investigating the frequency of parental gonosomal mosaicism in Neurodevelopmental Disorder cohorts to improve the genetic diagnostic rate

Authors:

N. Sandran^{1,2}, D. Webber¹, K. Harper², J. Berry¹, M. Corbett^{1,2}, A. MacLennan², J. Gecz^{1,2}, C. van Eyk^{1,2}; ¹Neurogenetics Res. Program, Adelaide Med. Sch., Univ. of Adelaide, Adelaide, Australia, ²Australian Collaborative Cerebral Palsy Res. Group, Robinson Res. Inst., Univ. of Adelaide, Adelaide, Australia

Abstract Body:

Neurodevelopmental disorders (NDDs) are a heterogeneous group of predominantly genetic disorders like intellectual disabilities (ID), autism spectrum disorders (ASD), epilepsy or cerebral palsy (CP), among others. Exome or genome sequencing fails to resolve up to 70% of select NDDs. We have identified parental mosaicism as one of the possible factors, not yet systematically investigated, which can increase NDD diagnostic rates. We consider parental mosaicism to be either (i) germ line restricted, i.e. germline mosaicism (GeM), or (ii) gonosomal (GoM), i.e. present in a subset of somatic tissues plus germ cells. Our study focuses on GoM present at <0.3 variant allele frequency (VAF) in an unaffected or mildly affected parent which may lead to filtering of that variant from the subsequent analysis/prioritisation, i.e. false negative finding. We hypothesised that the detection of parental GoM from the start of the trio analyses could resolve a proportion of unresolved individuals with NDDs. We investigated the Australian Cerebral Palsy (CP) Biobank (discovery cohort, n=155 parent-child trios) and Simons Simplex Collection exome datasets (replication cohort, n=200 parent-child trios) and developed a pipeline incorporating multiple variant callers (Mutect2, MosaicHunter and MosaicForecast) to detect GoM variants in parents prior to confirming the presence of those variants as constitutional germline variants in the affected child. Our preliminary analyses of the CP cohort show that, on average, each parent carries 56 suspected mosaic variants, and transmits two (approximately 4% of total mosaic variants). Therefore, each individual carries two constitutional germline variants which were transmitted from a GoM parent (average: 2.12 maternal variants, 2.14 paternal variants). These GoM variants are present at an average 0.25 VAF in the parents. Approximately 10% of GoM variants in our cohort are rare deleterious (gnomAD population frequency < 1x10⁻⁴, CADD phred >=20) exonic or canonical splicing variants and we have prioritised seven potentially disease-causing variants. While preliminary, our approach and supporting data show that considering parental mosaicism at the outset of trio exome or genome studies have the potential to discover a sizeable fraction (~ 4 variants/affected child) of otherwise undetected variants which may explain at least a proportion of the unexplained NDDs

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2319*. Is 22q11.2 deletion syndrome truly less common in African American patients?

Authors:

D. McDonald-McGinn^{1,2}, T. Crowley¹, D. McGinn³, K. Gaiser^{1,2}, V. Giunta¹, L. Lairson¹, O. Tran¹, M. Share⁴, K. Valverde², S. Pastor¹, B. Emanuel^{1,2}, E. Zackai^{1,2}; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Perelman Sch. of Med. of the Univ. of Pennsylvania, Philadelphia, PA, ³Children's Hosp. of Philadelphia and Hosp. of the Univ. of Pennsylvania, Philadelphia, PA, ⁴ECRI Inst., Philadelphia, PA

Abstract Body:

Background: In 2005, our Center reported underrepresentation of African Americans in our 22q11.2DS cohort (45/379 [12%]) compared with our hospital population (42%) postulating a paucity of dysmorphic features as a possible explanation and expressing concern that late/missed diagnoses impacted remediation and recurrence risk counseling. 17 years later only 11% of our patients (172/1620) are Black, suggesting that, as has been inferred when examining 22q11.2 low copy repeats (LCRs), the 22q11.2DS is truly less common in Blacks. Here we examined associated features and geocodes to determine if Black patients are being identified. **Methods:** We performed a retrospective IRB approved chart review on 1620 patients with 22q11.2DS. Deletion size, age at diagnosis, indication for testing, congenital anomalies particularly presence, absence, and severity of congenital heart disease (CHD), additional co-morbidities, demographics including geocodes, familial status, and year of birth were abstracted for analysis to compare patients self-reported as Black or White. In addition, Face2Gene technology was used to compare recognizable facial features in 125 Black patients with 22q11.2DS (Mean age 2.5 years) compared with White control patients with 22q11.2DS. **Results:** The mean age at diagnosis (MAAD) was 5.1 years for the entire cohort (N=1620); 5.4 years for White patients (N=1228); 5.2 years for Black patients (N=172); 3.7 and 2.6 years for White and Black probands respectively. Controlling for presence and type of CHD: MAAD where there was no CHD was 6.7 and 7.9 years; simple CHD was 4.3 and 1.5 years; moderate CHD was 2.6 and 1.8 years; complex CHD was 2.6 and 0.2 years, for White and Black patients respectively. Additional co-morbidities did not change MAAD but geocodes were found to be significant. Familial deletions were identified in 8% and 16% of White and Black probands respectively. Face2Gene identified 22q11.2DS associated craniofacial features in both cohorts. **Conclusions:** Controlling for CHD, timing of 22q11.2DS diagnoses for Black and White patients was similar. Co-morbidities were minimally different. Geocodes revealed significantly lower MAAD for White and Black patients traveling >100 miles to our Center (3.8 and 1.3 years respectively). White patients were more likely to travel (47% v 16%). Familial deletions were significantly more common in Black patients (16% v 8%). Face2Gene successfully recognized 22q11.2DS in both Black and White patients, suggesting the lack of craniofacial features is not a limiting diagnostic factor. Thus, we believe 22q11.2DS is truly less common in Black Americans perhaps explainable by differences in 22q11.2 LCRs.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2320. *KANSLI*, mosaic or not?

Authors:

M. I. A. Weerts¹, M. A. van Slegtenhorst¹, M. Alders², K. E. Stuurman¹; ¹Dept. of Clinical Genetics, Erasmus MC, Univ. Med. Ctr. Rotterdam, Rotterdam, Netherlands, ²Dept. of Clinical Genetics, Amsterdam Univ. Med. Ctr., Amsterdam, Netherlands

Abstract Body:

Background: Koolen-De Vries syndrome (OMIM #610443) is caused by heterozygous pathogenic variants - including 17q21.31 microdeletions and sequence variants - in the *KANSLI* gene. The syndrome is a clinically heterogeneous multisystemic condition with characteristic facial features, developmental delay, intellectual disability, hypotonia, epilepsy and congenital malformations. Here we present a patient with clinical features of Koolen-De Vries syndrome, where a common benign copy number variant complicated molecular confirmation of the clinically suspected diagnosis. **Case presentation:** A 9-year old girl presented with an IQ of 57, selective mutism, hearing loss and craniofacial features including a course face, brachycephaly, low frontal hairline, upslant of the eyes, broad mouth with widespread teeth, lowset ears, and long and slender hands and feet including broad first digits. Although clinically the diagnosis of Koolen-De Vries was suspected, genetic testing by microarray analysis and whole exome sequencing using a gene panel for intellectual disability did not confirm the suspected diagnosis. Subsequent whole exome analysis with extra emphasis on the *KANSLI* gene revealed a *de novo* pathogenic frameshift variant in exon 2 of the *KANSLI* gene (NM_001193466.1:c.1213_1216dup) with a skewed allele frequency of 20% - confirmed by Sanger sequencing analysis - suggesting mosaicism. Because the clinical presentation did not fit mosaicism, additional genetic testing was performed. Re-evaluation of the microarray data showed that the skewed allele frequency of the sequence variant was likely to be caused by a common benign copy number gain including *KANSLI* exons 1-3. Episignature analysis revealed a methylation profile distinct for Koolen-De Vries syndrome, confirming the clinical suspicion and pathogenicity of the sequence variant. It was concluded that the pathogenic frameshift variant NM_001193466.1:c.1213_1216dup in the *KANSLI* gene was present in a non-mosaic state and the molecular confirmation of the full clinical phenotype of Koolen-De Vries syndrome in this girl. **Discussion:** Our case highlights the pitfall of neutral copy number variants complicating genetic testing, the added value of Episignature methylation profile analysis to further interpret genetic findings, and the importance of clinical phenotyping.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2321. Leveraging genomic advances to ascertain the cause of congenital limb defects in Indian subcontinent.

Authors:

N. Langeh¹, M. Ansari¹, S. Khan¹, M. Faruq², M. Chowdhury¹, M. Kabra¹, N. Gupta¹; ¹All India Inst. of Med. Sci. (AIIMS), New Delhi, India, ²CSIR Inst. of Genomics and Integrative Biology (IGIB), New Delhi, India

Abstract Body:

Introduction: Limb reduction defects can range in severity from missing fingers to the partial or complete absence of a limb. The genetic etiology and pathogenesis for approximately one-third of limb reduction defects remain unidentified till date. In the present study, we explored the genomic variants involved in the pathogenesis of longitudinal limb defects: radial ray defect (RRD), and split-hand/foot malformation (SHFM). The aim of this study was to better define the strategy for genetic testing of limb defects and genotype-phenotype correlations in the Indian cohort. **Method:** Patients with clinical features suggestive of either RRD or SHFM were recruited. Following clinical and radiological evaluation, targeted Sanger sequencing of the gene was carried out for known syndromes. Genomic evaluation was performed using Chromosomal microarray (CMA) alone; CMA followed by whole exome sequencing (WES) analysis; WES analysis alone based on the clinical phenotype and a negative targeted molecular testing. **Results:** Ninety-one of the 144 subjects with radial ray defect (n=62) and split-hand/foot malformation (n=29) were recruited for genomic evaluation. Of 91, 16% (15/91) had consanguinity and 10% (9/91) had a positive family history. Extensive phenotyping suggested known syndromes including Holt-Oram syndrome (14/62) and thrombocytopenia-absent radius syndrome (4/62). Out of 91, genomic results (CMA in 38; WES in 43; targeted Sanger sequencing in 2) were available for 73 families with a global diagnostic yield of 23% (17/73). The yield of CMA in 18 SHFM and 20 RRD was 17% (3/18) and 20% (4/20) respectively, whereas the diagnostic yield of WES was 19% (3/16) SHFM and 15% (4/27) RRD. Monoallelic, novel pathogenic variant in *TBX3*, and biallelic, novel pathogenic variants in *CEP57*, *ROR2*, and *ESCO2* were found. **Conclusion:** Limb defects are commonly encountered; however, the overall yield of genomic evaluation remains low. We provide a preliminary basis resulting in the delineation of causative genes and improved molecular diagnosis of limb defects in clinical practice. Our study supports the importance of screening copy number variants in the diagnostic testing of SHFM.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2322. Long-read HiFi genome sequencing reveals a 2.7 kilobase intronic insertion in *NR5A1* as a cause of 46,XY disorder of sexual development

Authors:

G. Del Gobbo¹, X. Wang¹, M. Couse², C. Lambert³, S. Zhang³, S. Dhillon³, C. Fanslow³, W. Rowell³, C. Storer³, C. R. Marshall⁴, K. D. Kernohan^{5,1}, K. M. Boycott^{6,1}; ¹Children's Hosp. of Eastern Ontario Res. Inst., Univ. of Ottawa, Ottawa, ON, Canada, ²Ctr. for Computational Med., The Hosp. for Sick Children, Toronto, ON, Canada, ³Pacific BioSci.s of California, Inc, Menlo Park, CA, ⁴The Hosp. for Sick Children, Toronto, ON, Canada, ⁵Newborn Screening Ontario, Ottawa, ON, Canada, ⁶Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada

Abstract Body:

Short-read exome and genome sequencing have shown immense utility in diagnosing rare genetic disorders, however nearly two-thirds of rare disease patients still remain without a molecular diagnosis following sequencing and analysis. While short reads provide highly accurate detection of small variants, their application to detect complex structural variants and variation in highly repetitive or GC-rich regions of the genome is limited. Long-read genome sequencing addresses these limitations and enables accurate detection of structural variation, tandem repeat expansions, and improved sequencing in regions typically inaccessible by short reads.

We used long-read genome sequencing to investigate a family with a multigenerational autosomal dominant 46,XY disorder of sexual development where extensive clinical and research-based genetic testing had failed to identify a diagnosis. This condition was limited to males, incompletely penetrant, and showed phenotypic variability, with external genitalia abnormalities ranging from mild hypospadias to female genitalia, dysgenic testes, and reduced fertility. Previous testing included karyotyping, microarray, targeted *NR5A1* and *DHH* sequencing, and short-read exome, genome, and RNA-sequencing.

DNA from four affected family members was sequenced on the Sequel IIe system, generating PacBio HiFi reads, as part of the Care4Rare Canada research project. All four individuals were found to carry a heterozygous 2,752 bp insertion in intron 4 of *NR5A1*. *NR5A1* is located at 9q33.3, and pathogenic loss-of-function variants in this gene are associated with autosomal dominant 46,XY disorders of sexual development. The insertion is absent from control databases and segregates with the disorder in the family. Functional studies in fibroblasts derived from affected gonadal tissue support reduced *NR5A1* expression, and RNA-sequencing shows significant allele-specific expression of coding variants on the haplotype without the insertion.

This study provides a promising example of the utility of long-read sequencing in the context of unexplained rare diseases and highlights the contribution of undiscovered non-coding variation to Mendelian disorders.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2323. Low-level paternal mosaicism of novel pathogenic *DNM1L* variant confirms EMPF1 diagnosis in proband, highlighting importance of deep sequencing when parental carrier status is ambiguous.

Authors:

G. Akler^{1,2,3}, V. Dolgin⁴, O. Birk^{4,5}, S. Glover⁶; ¹TOVANA Hlth., Houston, TX, ²Geneyx Genomex Ltd., Herzliya, Israel, ³Precision Med. Insights, P.C., Great Neck, NY, ⁴The Morris Kahn Lab. of Human Genetics, Natl. Inst. for Biotechnology in the Negev and Faculty of Hlth.Sci., Beer-Sheva, Israel, ⁵Soroka Med Ctr, Beer-Sheva, Israel, ⁶Univ. of Mississippi Med. Ctr., Jackson, MS

Abstract Body:

Encephalopathy due to defective mitochondrial and peroxisomal fission-1 (EMPF1) is caused by biallelic pathogenic variants in *DNM1L*, encoding dynamin 1-like, a key regulator of mitochondrial fission or division. EMPF1 is characterized by variable phenotype, onset and severity, the main features of which include delayed psychomotor development or neurodevelopmental regression, dystonia or hypotonia, and refractory seizures consistent with epileptic encephalopathy, with some reported cases leading to death. Herein we report on the case of a boy affected by neurodevelopmental delay and regression, one sided hypotonia, involuntary movements, optic atrophy, hereditary alpha tryptasemia and immunological dysfunction, carrying two likely pathogenic compound heterozygous novel *DNM1L* variants: c.270C>G (p.Asn90Lys) and c.1393G>A (p.Val465Ile). Segregation analysis using Sanger sequencing showed maternal carrier status of the c.1393G>A variant, but no parental source of the c.270C>G variant. Suspicion of low-level mosaicism of the father was entertained given the classic symptoms of the proband. Deep sequencing was performed on multiple tissue types collected from the father with findings demonstrating mosaic carrier status of the father (7% in saliva, 12% in dermal fibroblast tissue, 5% in blood WBC), confirming the diagnosis in the proband. The findings emphasize the importance of follow-up studies using deep sequencing in cases with high index of suspicion for a genetic diagnosis despite inconsistent results in segregation analysis, enabling unraveling of parental mosaicism.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2325. Minimizing false negatives in NGS-based diagnostic testing by optimizing variant filtering strategy

Authors:

B. Guan, A. Naik, E. Ullah, D. McGaughey, A. Agather, B. P. Brooks, R. Hufnagel; Natl. Eye Inst., NIH, Bethesda, MD

Abstract Body:

Minimizing false-negative variant calling and filtering in next-generation sequencing (NGS) is critical for diagnostic testing and, ultimately, patient care. Sources of false-negative findings from variant calling are well-recognized, including insufficient coverage of certain medically relevant genes and failed identification of structural variants by short-read technologies. However, the contribution of variant filtering strategies toward false negatives is less well-understood. Variant allele frequencies (AF) and maximum subpopulation AF (popmax AF) in the general population are often used to filter out variants considered too common to cause a rare disease in a “hard-filtering” process. However, the allele frequencies are known to vary among different public population databases. We thus hypothesized that the widely disparate AFs in public reference databases could lead to false negatives in NGS-based genetic testing. We first examined the extent of AF differences for the variants located in the OMIM Gene Map in the gnomAD v2.1.1 exome and genome datasets. Among these 4701 disease-relevant genes, we found 6991 variants in 2977 genes with popmax AF < 0.01 in one dataset but > 0.01 in the other, whose overall AFs were 1.5 to ~1e5 fold higher in one dataset than the other. Of these, 48 are predicted loss of function (pLOF) variants. We thus established a variant filtering strategy that considers possibly pathogenic variants observed with AF > 0.05 in reference databases. As a result, from an exome cohort of 220 patients with inherited retinal dystrophies, we identified two likely-pathogenic variants in *CEP164* in a patient that were not reported in genetic testing by two clinical laboratories, despite the gene being included in the analysis. One *CEP164* exonic variant (gnomAD exome: AF, 0.064, popmax AF, 0.069; gnomAD genome: AF, 0.001, popmax AF: 0.002) in a homopolymer region was classified as benign by two clinical laboratories as reported in the ClinVar database. Detailed examination of other 47 pLOF variants with substantial AF differences in gnomAD exome and genome datasets implicates such variants in medically relevant genes may have frequent false negative NGS calls, such as variants in *MYO15A*, *NFI*, and *RNPC3*. This type of error from hard-filtering can be remedied by incorporating variant filter flags for low quality calls, coupled with relaxing AF cutoffs in the analysis pipeline. We expect that this strategy to address the false-negatives resulting from disparate AFs in reference databases and to resolve missing inheritability in a substantial number of patients in NGS-based genetic testing.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2326. Mitochondrial variants and short tandem repeat expansions detected by clinical whole genome sequencing

Authors:

K. Schulze¹, L. Vossaert¹, P. Liu¹, C. Qu², V. Nguyen², M. Santana², L. Meng², C. Eng¹, H. Dai¹, F. Xia¹; ¹Baylor Coll. of Med., Houston, TX, ²Baylor Genetics Diagnostic Lab., Houston, TX

Abstract Body:

Whole genome sequencing (WGS) allows the detection of many different types of genetic variation that have traditionally been evaluated by other tests and methodologies. Aside from genomic single nucleotide, copy number, and structural variants, WGS can also detect mitochondrial variants and short tandem repeat (STR) expansions. We have reviewed the mitochondrial variants and STR expansions reported in our laboratory among all clinical WGS cases.

Using PCR-free WGS, we reported 7 mitochondrial variants, found in 6 individuals, that had $\geq 1,000\times$ read coverage, $\geq 1.5\%$ heteroplasmy, and were classified as pathogenic, likely pathogenic, or variant of unknown significance favoring pathogenic. The average depth of these variants was 3,663.6 reads (range 1,601-5,412 reads) with mean heteroplasmy of 22.8% (range 1.8%-100%). All variants occurred only once, with the exception of m.14453G>A, which was reported twice. In addition, we reported 2 STRs with expansions that were within the range of a full mutation, one in the *DMPK* gene and another in the *PHOX2B* gene. Of note, our validation studies have shown that the accuracy of calling STR repeat expansions by WGS decreases as the true number of repeats increases; this means that while WGS can detect STR expansions beyond the normal range for most repeat expansion disorders, the short read technology often limits the determination of the exact number of expanded repeats. In summary, we have encountered multiple examples of pathogenic mitochondrial variants and STR expansions in our experience with clinical WGS and can confirm the wide scope of genetic variation that can be detected by WGS.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2327. Molecular characterization of a rare case of chromosome 20p duplication syndrome in an Egyptian male patient with multiple congenital anomalies

Authors:

S. Abd Allah, M. K. Mekkawy, E. Asem, A. M. Mohamed, S. Ismail, A. Kamel; Natl. Res. Ctr., Cairo, Egypt

Abstract Body:

Introduction Chromosome 20p duplication syndrome is a rare chromosomal disorder that occurs due to duplication of variable segments of the short arm of chromosome 20. The severity and manifestations of the condition depend on the size and location of the duplication and if it is associated with another chromosomal abnormality. Very few cases of chromosome 20p duplication have been reported with no recurrent breakpoints. The majority have partial duplications and have occurred as part of a translocation (along with a deletion in another chromosome). Precise genotype-phenotype correlations remain elusive due to the small number of patients and heterogeneity of breakpoints. The clinical manifestations reported in people with chromosome 20p duplication are markedly variable ranging from completely normal individuals to severe clinical disability. Symptoms and signs include intellectual disability (ID), developmental and speech delay, poor coordination, dental problems, spinal bone abnormalities, distinctive facial features, and heart defects. **Case Report** A male patient presented to the Human Genetics Clinics at the National Research Centre, Cairo, Egypt with dysmorphic features and developmental delay. He was 7 years old of a non consanguineous parents. Detailed clinical examination revealed distinctive facial dysmorphic features, mild scoliosis, delayed language development and mild intellectual disability with poor coordination. Pelvic ultrasonography showed hydronephrosis and urethral duplication, while brain MRI showed ventriculomegaly. Pedigree analysis revealed similar clinical manifestations in two family members. Karyotyping for the patient and his parents was performed using GTG banding technique on peripheral blood lymphocytes and showed 46,XY,der(14) in the patient resulted from a balanced paternal translocation: 46,XY,t(14;20) (p11.1;q11.1), while the mother had a normal karyotype (46,XX). Molecular characterization of the breakpoints was conducted for the patient using Affymetrix Genome-Wide Human SNP Array 6.0 and identified duplication of the whole short arm of chromosome 20 without associated abnormalities in chromosome 14 : arr[GRCh38]20p13q11.1 (80927_30190264)x3 Fluorescent insitu hybridization (FISH) was performed for the patient and his father using whole chromosome paint probes for chromosomes 14 and 20 and confirmed the origin of the add material in the proband and the balanced translocation in the father. **Conclusion** We report the first Egyptian patient with rare pure 20p duplication, inherited from a balanced paternal carrier with molecular characterization and phenotype/ genotype correlation.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2328. Monozygous twins with distal 5p duplication: a case report.

Authors:

D. Contreras, **R. Lebel**, J. Black; SUNY Upstate Med. Univ., Syracuse, NY

Abstract Body:

Duplications of the short arm of chromosome 5 are very rare, and their severity is highly variable. The few patients reported with duplication distal to 5p13.3 had milder psychomotor delays and short stature compared to those with duplications proximal to 5p13.3, the latter group being characterized by developmental delays, failure to thrive and seizures.

We saw monozygotic diamniotic twin girls born to a 25-year-old primigravida mother and 28-year-old father for developmental delays, behavioral concerns, and multiple malformations (umbilical hernia, loose jointed fingers, pectus excavatum and Duane anomaly). SNP arrays revealed 2.9Mb duplication of 5p15.33p15.32 in both girls. In the DECIPHER database and one previously published case report, one finds similar duplications (ranging 1.07 to 2.42 Mb) that overlap in five genes: *MRPL36*, *NDUFS6*, *IRX4*, *IRX2* and *IRX1*. Though all patients have global developmental delays, the twins uniquely present with pectus excavatum and Duane anomaly. Their duplication exclusively contained the genes *LSINCT5* and *C5orf38*. Further, their family history suggests inheritance with incomplete penetrance. In conclusion, the overlapping five genes in all these cases may be related to their phenotype.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2329*. Multiple molecular diagnoses identified by clinical whole genome sequencing in an international cohort of more than 1,200 individuals

Authors:

E. Thorpe, A. Malhotra, T. Kalista, D. Perry, R. Taft; Illumina, San Diego, CA

Abstract Body:

Clinical whole genome sequencing (cWGS) is increasingly utilized to support the molecular diagnosis of patients with rare genetic disorders. From June 2016 to May 2022, 1,233 individuals with a suspected rare genetic disease pursued cWGS through the Illumina iHope Program, a philanthropic effort to provide cWGS to children with resource limitations impeding access to molecular testing. Here, we describe individuals who received multiple molecular diagnoses (MMD).

A positive test outcome (PTO) was reported in 491 individuals (39.8%), defined by at least one reported variant consistent with the indications for testing (IFT), found in the molecular state expected to cause disease, and classified as pathogenic or likely pathogenic by ACMG variant classification criteria. MMD were reported in 46 individuals (3.7% of cases; 9.4% of positive reports). Molecular diagnoses were associated with two (n=42 individuals), three (n=3 individuals) or four (n=1 individual) distinct conditions. MMD were specific to the IFT in 39% (18/46), related to the phenotype in addition to a secondary finding in one of 59 genes as defined by ACMG v2.0 in 20% (9/46), or in addition to an incidental finding (IF) in a gene outside of ACMG recommendations but which met laboratory reporting criteria due to clinical guidelines or actionability criteria in 41% (19/46).

Across the 46 MMD reports, 114 variants were reported. A majority (65%) were SNVs, which included two mitochondrial SNVs, followed by small indels (18%) and CNVs (17%), which included two chromosomal aneuploidies and one UPD. Eight cases had multiple variant types contributing to the MMD on a single report.

Variants related to the IFT were associated with 78 unique diseases, each reported in a single individual except for Cornelia de Lange syndrome (n=3), *CACNA1A*-related disorders (n=2) and *TRAF*-related syndrome (n=2). Twenty-six (33%) disorders related to the IFT followed dominant or X-linked inheritance associated with heterozygous or hemizygous *de novo* variants, with an additional 12 (15%) suspected to be *de novo* but both parents were not available for testing. Eighteen (23%) disorders related to the IFT followed autosomal recessive inheritance. Disorders reported as a secondary finding all followed dominant inheritance with a majority inherited from a parent. Nearly half (9/19) of IFs were associated with dominant cancer predisposition syndromes, although glucose-6-phosphate dehydrogenase deficiency was the single most common IF (n=5).

In summary, identification of MMD was observed in nearly 10% of positive reports, which supports the benefit of comprehensive testing approaches, including cWGS, in rare disease cohorts.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2330. Mutation Spectrum of Polycystic Kidney Disease-Toronto Genetic Epidemiology Study

Authors:

A. Haghghi^{1,2}, I. Bari², S. Khowaja², N. He², I-A. Iliuta², X. Song², M. Lanktree³, A. Paterson⁴, J. Lerner-Ellis⁵, Y. Pei²; ¹Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, United States, Boston, MA, ²Toronto Gen. Hosp., Univ. Hlth. Network, Univ. of Toronto, Toronto, ON, Canada, ³St. Joseph's Hlth.care Hamilton, Hamilton, ON, Canada, ⁴SickKids Res. Inst., Toronto, ON, Canada, ⁵Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada

Abstract Body:

Background: Autosomal dominant polycystic kidney disease (ADPKD) is the most common hereditary kidney disease worldwide. Mutations in *PKD1* and *PKD2*, respectively, account for 85% and 15% of the genetically resolved cases in clinical series enriched with high-risk patients. Here, we report the ADPKD mutation spectrum of a large cohort of relatively unselected patients from a single geographical region. **Methods:** We performed mutation screening in 2,171 patients from 1,606 different families from the Greater Toronto Area (population 6.8 million) using NGS targeted sequencing and multiplex ligation-dependent probe amplification of *PKD1* and *PKD2*, as well as NGS of a panel of 50 cystic disease genes. Standard algorithms for sequence alignment, base calling, and QC filtering were applied to identify rare (MAF $\leq 1\%$) deleterious variants as predicted by multiple algorithms. **Results:** We detected *PKD1* and *PKD2* mutations in 1205 (75%) families, non-*PKD1* and *PKD2* (i.e. *ALG8*, *ALG9*, *PKHD1*, *GANAB*, *PRKCSH*, *SEC63*, *LRP5*, *WFS1*, *TSC1-2*, *COL4A1*, and *COL4A3-5*) rare putative pathogenic variants in 120 (10%) families, with no mutations detected in 281 (15%) families. Among the *PKD1* and *PKD2* genetically resolved families, 916 (76%) and 289 (24%) were due to mutations in *PKD1* and *PKD2*, respectively. Adjusted for exon size across all 46 exons in *PKD1*, we found an enrichment of truncating mutations in exon 44. We also found over 100 recurrent mutations in ≥ 2 different families (haplotype analysis in progress). **Conclusion:** We found extensive genetic and allelic heterogeneity in ADPKD with a higher prevalence of *PKD2* mutations than reported in the clinical series. We also found non-*PKD1* and non-*PKD2* cystic disease mutations in 10% of families, while 15% of the families remained genetically unresolved.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2331. Non-invasive detection of *TEK* mutations in a cohort of children with isolated venous malformations using cell-free DNA

Authors:

N. Au^{1,2}, D. M. Jensen¹, V. Dmyterko¹, T. Cook¹, K. Zenner³, J. Perkins^{4,3}, J. T. Bennett^{1,5}; ¹Seattle Children's Res. Inst., Seattle, WA, ²Univ. of Washington, Seattle, WA, ³Dept. of Head and Neck Surgery, Univ. of Washington Sch. of Med., Seattle, WA, ⁴Seattle Children's Hosp., Seattle, WA, ⁵Div. of Genetic Med., Dept. of Pediatrics, Univ. of Washington Sch. of Med., Seattle, WA

Abstract Body:

Venous malformations (VeM) are the result of abnormal vascular morphogenesis. VeM are caused by post-zygotic activating mutations within *TEK*, a tyrosine kinase receptor within the *PI3K-MTOR-RAS-MAPK* pathway. *TEK p.L914F* is among the most common pathogenic variant in VeM, contributing to ~45% of cases. Currently, molecular diagnosis of VeMs requires surgically resected tissues because these somatic variants are present in a small fraction (1-20%) of cells in the lesion. Cell-free DNA (cfDNA) is an emerging diagnostic analyte that is already in use in cancer diagnosis and prenatal genetic screening. The critical role mutant endothelial cells play in the development of VeM and their proximity to blood led us to ask whether plasma-based cfDNA could be used to detect the mutant variant. We detected *TEK* mutations in the cfDNA from 7/19 patients with isolated VeM using droplet digital polymerase chain reaction (ddPCR). Plasma obtained directly from lesions was more likely to have a mutation detected (3/4) than peripherally derived plasma samples (2/15). Mutations were detected in plasma-derived cfDNA but not in the blood genomic DNA. Our results will reduce the need for invasive procedures and the time required to determine the pathological variant in VeM patients, improving their chances for targeted pharmacologic therapies.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2332. Novel potential biomarkers for *DICER1* syndrome

Authors:

M. Wu¹, R. D. Lopez², M. Warren^{3,4}, R. Shah^{1,4}, P. Neviani⁵, J. F. Amatruda^{1,4}; ¹Div. of Hematology-Oncology, Cancer and Blood Disease Inst., Children's Hosp. Los Angeles/Univ. of Southern California, Los Angeles, CA, ²Children's Hosp. Los Angeles/Univ. of Southern California, Los Angeles, CA, ³Dept. of Pathology and Lab. Med., Children's Hosp. Los Angeles, Los Angeles, CA, ⁴Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ⁵Saban Res. Inst., Children's Hosp. Los Angeles, Los Angeles, CA

Abstract Body:

MicroRNAs (miRNAs) are small RNA molecules that play a critical role in regulating gene expression. The *DICER1* enzyme is a key component of the microRNA biogenesis pathway. Reflecting this central role, patients who inherit or acquire mutations in the *DICER1* gene have an increased risk of developing a spectrum of benign and malignant tumors —thereby earning the recognition as a pediatric cancer predisposition syndrome known as *DICER1* syndrome (OMIM 601200). *DICER1* syndrome is definitively diagnosed by DNA sequencing demonstrating presence of a loss of function mutation on one copy of *DICER1* plus a “hotspot” mutation on the other copy of *DICER1*. As tumor and germline DNA sequencing can pose logistical and/or cost difficulties, a complementary diagnostic method based on marker expression is valuable for both diagnostics and surveillance. Biomarkers that are candidate therapeutic targets are also appealing as these may offer a more tailored approach for therapy. As there are no approved biomarkers for identification of *DICER1*-associated tumors, we performed RNAseq and miRNAseq in an engineered murine mesenchymal cell model that incorporates a specific *DICER1* hotspot mutation to identify candidates. We validated candidate biomarkers at the RNA (*Eya2*, and miR-30d-3p) and protein (*Eya2*) levels *in vitro*, to provide proof of principle data for their eventual clinical utility. The miRNA candidate biomarkers levels were evaluated in both total and small extracellular vesicle-enriched fractions. These novel data pave the way for additional tools for diagnosis and surveillance in patients with *DICER1* syndrome.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2333. Optical genome mapping and whole genome sequencing in a case of multiple chromosomal rearrangements

Authors:

J. Levy¹, M. Rachid¹, C. Dupont², A. Rapinat³, D. Gentien³, A. Verloes⁴, Y. CAPRI⁵, A-C. Tabet⁵; ¹AP-HP, Paris, France, ²APHP-Hosp. Robert Debre, Paris, France, ³Inst. Curie, Paris, France, ⁴Robert-Debré Univ. Hosp., Paris, France, ⁵Robert Debré Univ. Hosp., Paris, France

Abstract Body:

Few studies have shown the potential of using short-reads whole-genome sequencing (WGS) approaches and optical mapping (OM) for the detection of clinically relevant structural variants (SVs) in neurodevelopmental disorders. We presented a 7-year-old female patient with macrocephaly (+3.5 SD), moderate valvular stenosis with patent foramen and developmental delay. SNP array analysis revealed several genomic rearrangements: a paternal uniparental isodisomy (UPD)pat of the 11pter region and four *de novo* duplications: a 800 kb-3q23 duplication; a 730kb-12q13.12 duplication, a 256 kb-15q15.1 duplication and a 342 kb-19q13.42 duplication. Whereas (UPD)pat explained the macrocephaly, the 4 duplications were considered as variations of uncertain significance. WGS allowed specifying that the 3q23 and 15q15.1 duplications were in tandem and revealed structural variants involving both chromosomes 12 and 19. Short-reads WGS could not map precisely any of the rearrangement's breakpoints and position that lie within SVs. We used optical genome mapping (OGM) to determine the orientation and the position of the duplicated segments at 12q13.12 and 19q13.42, allowing us to propose a mechanism and classify this rearrangement. This study highlights the complementarity of techniques for the detection of clinically relevant structure variants.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2334. Optical genome mapping identified a likely pathogenic *POLR3B* variant in an undiagnosed male with ataxia, hypotonia, and cerebellar atrophy

Authors:

A. Ortega¹, J. Yu¹, A. Baxter¹, J. Rosenfeld², K. Worley², S. Ketkar², D. A. Scott², Undiagnosed Diseases Network, A. Hastie¹, A. Chaubey¹; ¹Bionano Genomics, San Diego, CA, ²Baylor Coll. of Med., Houston, TX

Abstract Body:

Recent advances in genomic technology have significantly increased the identification of causative variants in subjects with genetic disorders. However, a large portion of subjects with a suspected genetic etiology remain undiagnosed after numerous advanced genetic tests. Here we describe a research case referred to a Bionano postnatal clinical study through the Undiagnosed Diseases Network (UDN). The subject was a 6-year-old male presenting with a normal early development, who then presented with neurological symptoms including ataxia, hypotonia, cerebellar atrophy and speech delay beginning at one and a half years of age. Prior genetic tests including chromosomal microarray, ataxia panel, and whole-exome sequencing (WES) were not diagnostic. Using optical genome mapping (OGM), the subject's ultra-high molecular weight (UHMW) genomic DNA was isolated, labeled, and imaged across nanochannel arrays. An automated bioinformatics pipeline was then used to identify structural variations in the genome. In this case, we identified a heterozygous 5.5 kilobase (kb) deletion located in *POLR3B* which is associated with autosomal recessive leukodystrophy (MIM: 614381). This 5.5kb deletion is within a 17.9kb window (chr12:106,446,738-106,464,666, hg38) including exons 20 to 23 and is likely to affect one of the coding exons of *POLR3B*. Importantly, the subject's prior WES data identified a maternally inherited likely pathogenic missense variant (c.1244T>C; p.Met415Thr) in *POLR3B*. After reviewing trio WES data (subject and his parents), a paternally inherited exon 20 deletion was suspected. In addition, the subject's RNA sequencing data showed exon 20 skipping in some *POLR3B* transcripts, supporting the finding of an exon 20 deletion. Biallelic missense and deletion variants in the proband were later confirmed by whole-genome sequencing and targeted Sanger sequencing. In conclusion, we used OGM in combination with other genetic testing technologies to identify likely pathogenic, biallelic *POLR3B* variants in a previously undiagnosed subject with ataxia, hypotonia, and cerebellar atrophy. This case demonstrates the challenges and limitations of current genetic analysis tools and highlights the need to use complementary technology, like OGM and NGS sequencing, to comprehensively identify candidate disease-causing variants and provide comprehensive answers in research on rare disease.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2335*. Optical genome mapping identifies double parental paracentric inversions as risk factor for atypical monocentric recombinant chromosomes in offspring

Authors:

P. Kuentz¹, A. Vitobello², Y. Duffourd³, M. Chevarin⁴, B. Keren⁵, C. Schluth-Bolard⁶, D. Sanlaville⁷, M. Rossi⁸, A-L. Mosca⁴, N. Marle⁴, C. Thauvin-Robinet⁴, L. Faivre⁴, P. Callier⁹; ¹CHU BESANCON, BESANCON, France, ²CHU Dijon Bourgogne, Dijon, France, ³CHU DIJON - INSERM U1231, DIJON, France, ⁴CHU Dijon, Dijon, France, ⁵APHP, Paris, France, ⁶CHU Strasbourg, Strasbourg, France, ⁷HCL, CBPE, BRON Cedex, France, ⁸CHU Lyon, Lyon, France, ⁹CHU Le Bocage, Dijon, France

Abstract Body:

Individuals with heterozygous paracentric inversions are not thought to have a higher risk of abnormal offspring as compared to the general population so that prenatal diagnosis should not be offered systematically. Using a combination of conventional, molecular and next-generation cytogenetics, we characterized atypical unbalanced recombinant monocentric chromosomes identified in three unrelated individuals with syndromic intellectual disability, arisen from parents with balanced double paracentric inversions. In the affected individuals, chromosomal microarray identified a common pattern composed of an interstitial gain, a variable copy-neutral region and an interstitial loss in all affected individuals. Optical mapping (OM) revealed that the duplicated segment was inserted in an inverted orientation within the deleted region. In parents, OM identified large paracentric inversions (ranging from 17.1 to 23.6 Mb) encapsulating, in two out of three cases, a smaller inversion (of 1.8 and 4.8 Mb), corresponding to the neutral regions found in the affected individuals. In the third case, neither technique could identify an inversion in the copy-neutral region because of its small size (16.4 Kb) and of its composition in segmental duplications. Finally, using OM and fluorescent in situ hybridization, we demonstrated that at least in one case, the two encapsulated inversions were in cis. Overall, our data support either a double inversion loop or U-type exchange as possible meiotic mechanisms as a cause of the atypical rearrangements identified in affected kindred. Our observations may guide future recommendations for genetic counseling in case of known parental paracentric inversions.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2336. Optical genome mapping improves clinical interpretation of constitutional copy number gains.

Authors:

G. Raca, J. Han, C. Fong, J. Ji, R. Schmidt, A. Dharmadhikari; Children's Hosp. Los Angeles, Los Angeles, CA

Abstract Body:

Genomic location and orientation of copy number gains, which are critical for their clinical interpretation, cannot be determined from chromosomal microarray (CMA) data. Consequently, copy number gains detected by CMA are often classified as variants of uncertain clinical significance. Sequencing studies have shown that the majority of copy number gains are positioned head-to-tail (in direct orientation) adjacent to the original locus, likely preserving the reading frame of the affected gene. If this structural information was readily available, most copy number gains could be classified as likely benign, thus avoiding clinician and patient concern and costly follow up parental studies. In contrast, those gains that are inverted, triplicated, or inserted into a distant locus can disrupt or fuse genes in a manner that is not revealed by CMA, resulting in missed diagnoses. Optical genome mapping (OGM) allows genome-wide detection of CNVs with the resolution that supersedes most clinically used CMA platforms, but it also provides detailed structural information about altered chromosome regions. We hypothesized that OGM will allow improved interpretation of the functional and clinical significance of copy number gains as compared to CMA. By retrospectively searching our clinical database we identified 22 cases in which CMA testing detected copy number gains that included either the 5' or the 3' portion of a disease-associated gene. In all the cases the gains were reported as variants of uncertain clinical significance. OGM not only detected the same copy number gains detected by CMA, but also provided information about their genomic location and orientation. Gains involving the following genes implicated in autosomal dominant or X-linked genetic disorders were found to represent direct in situ duplications that preserved an intact copy of the coding region: *PAFAH1B1*, *MID1*, *PLCB4*, *PDK3*, *EXT1*, *TRIP12*, *RAD51*, *TTC21B* and *SLC6A1*. In contrast, more complex rearrangements were identified involving the *GLI3* and *HCNI* genes. The structural information obtained by OGM allowed to predict the effect of the detected gains on gene function, and to reclassify the gains as likely benign or likely pathogenic. Our study demonstrates that OGM not only shows concordant results to CMA, but that it also represents a superior assay for constitutional CNV testing since it allows improved interpretation of the clinical significance of copy number gains.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2337. Pallister-Killian syndrome: Characterization of the isochromosome 12p by fluorescent in situ hybridization: A Case Report.

Authors:

S. Lavu, L. Vemulakonda, F. Shakil, H. Islam, C. KUMAR; Westchester Med. Ctr., Valhalla, NY

Abstract Body:

Pallister-Killian syndrome (PKS) is a rare sporadic developmental disorder usually with a tissue-specific mosaic distribution of an additional isochromosome 12p [i(12p)] without gender predilection. Typically, children with this diagnosis are characterized by craniofacial dysmorphism, hypotonia, intellectual disabilities, skin pigmentation, seizures, heart defects, diaphragmatic hernia, sparse hair and other phenotypic features. As such, it can be detected as early as in utero or in infancy and childhood. The signs and symptoms of PKS vary in severity. Here, we present a case of 4-day old male baby born to a 37-year-old at 34 weeks with a left sided congenital diaphragmatic hernia (CDH) found on prenatal ultrasound. He also had dysmorphic craniofacial features and a slight dextrocardia possibly shifted secondary to the CDH. The baby also presented with respiratory distress, bilateral cryptorchidism and bronchopulmonary dysplasia. On echocardiography, he had moderately dilated right ventricle with moderate tricuspid valve insufficiency with tricuspid regurgitation. He had evidence of persistent pulmonary hypertension secondary to pulmonary hypoplasia. Due to his phenotypic abnormalities, the cytogenetic and chromosomal microarray investigations were indicated. Chromosome analysis performed on GTG banded metaphases prepared from cultured lymphocytes revealed a mosaic male karyotype as: 47,XY,+i(12)(p10)[3]/46,XY[47]. Three out of 50 cells had 47 chromosomes with an extra isochromosome 12p resulting in the tetrasomy of 12p. The rest of the 47 cells were normal. The Fluorescence in situ hybridization (FISH) tests were performed by using two chromosome 12 specific FISH probes (Abbott Molecular Inc.), CEP 12 and LSI ETV6/RUNX1 to characterize the extra isochromosome. The FISH analysis revealed the presence of two signal patterns for CEP 12 probe (2 and 3 copies) and ETV6/RUNX1 probes (2 and 4 copies for ETV6 (12p13.2) gene and normal 2 copies for RUNX1 (21q22) gene) confirming the presence of an extra isochromosome 12p. The microarray analysis performed on peripheral blood identified the presence of a 34.7 megabase mosaic triplication (two additional copies) of short arm of chromosome 12 (from 12p13.33 to 12p11.1 region). The cytogenetic, FISH and microarray results confirmed the presence of a mosaic cell population with an abnormal cell line with an extra isochromosome 12p resulting in the tetrasomy of 12p in the peripheral lymphocytes of this newborn. This is consistent with a diagnosis of mosaic Pallister-Killian syndrome. This abnormality occurs in de novo. However, presence of germline mosaicism in one of the parents could not be ruled out.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2338. Pathogenic variants detected by RNA sequencing in Cornelia de Lange syndrome

Authors:

R. Seyama^{1,2}, Y. Uchiyama¹, J. Ceroni³, V. Kim³, I. Furquim³, R. Honjo³, M. Castro³, L. Pires³, H. Aoi², K. Iwama^{1,4}, K. Hamanaka¹, A. Fujita¹, N. Tsuchida¹, E. Koshimizu¹, K. Misawa¹, S. Miyatake¹, T. Mizuguchi¹, S. Makino⁵, A. Itakura², D. Bertola³, C. Kim³, N. Matsumoto¹; ¹Yokohama City Univ., Yokohama, Japan, ²Juntendo Univ., Tokyo, Japan, ³Inst. da Crianca, Faculdade de Med., Univ. de Sao Paulo, Sao Paulo, Brazil, ⁴Yokohama City Univ. Med. Ctr., Yokohama, Japan, ⁵Juntendo Univ. Urayasu Hosp., Urayasu, Japan

Abstract Body:

RNA sequencing is a technique used to determine the sequences of all transcripts derived from target tissues, and it enables the identification of deep-intronic variants that create a cryptic exon by pipelines for detecting aberrant splicing events. However, RNA sequencing analysis of neurodevelopmental disorders is limited by difficulties in obtaining appropriate RNA derived from target neural tissues. Recent studies suggest that transcript isoforms significantly overlap (approximately 60%) between brain tissue and Epstein-Barr virus-transformed lymphoblastoid cell lines (LCLs). Interestingly, 14 cohesion-related genes with variants that cause Cornelia de Lange Syndrome (CdLS), a rare neurodevelopmental disorder with dysmorphic features, are highly expressed in brain and LCLs. Among 66 CdLS families, we previously found either pathogenic single nucleotide variants (SNVs) or copy number variants (CNVs) in 46 families (46/66 = 69.7%), but not in the other 20 families. In this context, we first performed RNA sequencing of LCLs from 22 solved (with pathogenic variants) and 19 unsolved (with no confirmed variants) CdLS cases, and 106 healthy controls in the Genotype-Tissue Expression (GTEx) Biobank. RNA sequencing pipeline was developed using 22 solved CdLS cases and 105 healthy controls with two different methods: short variant analysis (for single nucleotide and indel variants) and aberrant splicing detection analysis. Then, 19 unsolved CdLS cases were subsequently applied to our pipeline, and four pathogenic variants in *NIPBL* (one inframe deletion and three intronic variants) were newly identified. Two of three intronic variants were located at *Alu* elements in deep-intronic regions, creating cryptic exons. As a result, the total diagnostic rate was increased from 69.7% (46/66) to 75.8% (50/66). RNA sequencing with LCLs was useful for identifying hidden variants in exome-negative cases.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2339. Positive Findings in Clinical Mitochondrial Genome Testing by NGS

Authors:

J. Yang¹, C-Y. Kao¹, J. Dong¹, J. Lattier¹, H. DAI², L. Meng², F. Xia², E. Schmitt¹, S. Peacock¹, J. Lattier¹, W. Craigen³, L-J. Wong³, C. Eng², Y. Wang²; ¹Baylor Genetics, Houston, TX, ²Baylor Coll. of Med./Baylor Genetics, Houston, TX, ³Baylor Coll. of Med., Houston, TX

Abstract Body:

Background: Genetic disorders lead to a wide spectrum of human disease. In addition to the nuclear-encoded DNA, maternally inherited mitochondrial DNA plays an important role in human health. The mitochondrial genome is a 16569bp double helix loop and encodes 37 genes. Pathogenic variants on the mitochondrial DNA are associated with human diseases including mitochondria myopathy, maternally inherited diabetes mellitus and deafness (MIDD), Leber's hereditary optic neuropathy (LHON), Leigh syndrome and many others. **Methods:** Baylor Genetics has established a robust platform to sequence the whole mitochondrial genome using Next Generation Sequencing (NGS) technology. Long-range PCR is used to amplify the mitochondrial DNA and exclude nuclear DNA. NGS is applied to sequence the samples in a high throughput manner. NGS data are processed to call single nucleotide variants (SNV) and large deletions in the mitochondrial genome. Variants can be confidently called at heteroplasmic levels as low as 1.5%. We retrospectively analyzed positive findings from cases tested consecutively but without selection based on clinical indications. **Results:** We detected pathogenic or likely pathogenic SNV in 709 cases and large deletions of mitochondrial genomic regions in 600 cases. 27 cases were positive for both SNV mutations and large deletions. The majority of samples were extracted from blood. Other sample types were urine, bone marrow, fibroblasts, amniocytes, buccal swabs, previously extracted DNA and affected tissues, including liver, skeletal muscle and others. The ages of test subjects ranged from prenatal to 89-year-old adults. The pathogenic variant m.3243A>G in the tRNA Leucine gene associated with MELAS syndrome was the most prevalent mutation, which was detected 186 times. It was heteroplasmic in all cases, ranging from 1.5% to 94.9%. **Conclusion:** Mitochondrial genome sequencing is a robust and efficient means of molecular diagnosis for human mitochondrial genome disorders.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2340. Potentially noteworthy intronic variants in the *COL5A1* and *COL5A2* genes in patients with classical Ehlers-Danlos syndrome: The necessity for RNA study and genetic diagnostic approaches

Authors:

W. Heo, J-H. Jang; Samsung Med. Ctr., Seoul, Korea, Republic of

Abstract Body:

The Ehlers-Danlos syndromes (EDS) are clinically and genetically heterogeneous disorders characterized by abnormal collagen production, mainly involving joints, skin, vasculature, and other hollow organs. Among 13 subtypes of EDS classified by the 2017 international classification of the EDS, classical Ehlers-Danlos syndrome (cEDS, OMIM#s 13000, 130010) also known as EDS type I/II has a prevalence of 1 in 20,000-40,000 which is the second most common subtypes of EDS followed by hypermobile EDS. It is an autosomal dominant disorder and 90% of cases result from variants in the *COL5A1* and *COL5A2* genes, which encode the type V collagen, the $\text{pro}\alpha 1$ (V) and $\text{pro}\alpha 2$ (V) chains respectively. Collagen V proteins consist mostly of the triple-helical region containing the repetition of [Gly-Xaa-Yaa] and are heterotrimers (two $\text{pro}\alpha 1$ and one $\text{pro}\alpha 2$). It is known that the in-frame deletion had more severe clinical manifestations than the null variant due to these characteristics of the collagen structure. There were patients without a genetic diagnosis in approximately 10% of cEDS patients. Since functional studies including RNA studies on non-canonical intronic variants have scarcely been conducted, there has room for improving the diagnostic yields of EDS. With the expansion of the test method through massively parallel sequencing, access to deep intronic variants has become easier than before, and since whole-transcriptome analysis is also possible, the technical basis for actively testing splicing alteration has been established. For these reasons, despite the skin biopsy is required to obtain cultured fibroblasts for RNA study of *COL5A1/2*, it is well worth performing. Here, we predicted potential pathogenic variants among non-canonical intronic variants (± 50 nt from closest exon/intron junction without canonical splice-site) in *COL5A1/2* using SpliceAI scores and known variants listed in the Human Gene Mutation Database 2022.1. Our results demonstrate that there was a total of 2,412 intronic variants (1,170 variants of *COL5A1*; 1,242 variants of *COL5A2*) with potentially splicing alteration. We suggest that these variants could be needed in the RNA study of fibroblasts culture when the patient's phenotype is strongly suspected of cEDS. If this method is not available, whole-genome sequencing could be an alternative option for discernment of necessities of the RNA study in limited circumstances.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2341. Rapid whole genome sequencing in the neonatal intensive care unit of Brazilian hospitals.

Authors:

M. Migliavacca; DASA, São Paulo, Brazil

Abstract Body:

The rapid pace of advancement in understanding the molecular basis of genetic diseases has led to a growing appreciation of the impact of genetic diagnosis on provision of medical care. Provision of optimal clinical care for affected children depends on timely ascertainment of the underlying genetic cause, facilitating a shift from empiric treatment to precise management of an identified disorder. Previously, the long turnaround time for genetic test results precluded their real-time application to critical care medicine. Recently, whole genome sequencing (WGS) coupled with a focused phenotype-driven analysis of WGS data, became available as an option for rapid molecular diagnosis. Thus, applicability of WGS to critical care medicine may no longer be hindered by time constraints. Studies have shown that early molecular diagnoses improve outcomes and reduce healthcare costs. Nevertheless, provisioning rapid and affordable genomic testing strategy within a national healthcare service in order to deliver equity of access is challenging. Traditionally, a series of genetic tests are used step wise for the cytogenetic or molecular diagnosis of these diseases. Mainly in low- and middle-income countries, the access to these tests is difficult, they are very expensive and the results are frequently delayed. Because of that precise timely diagnoses are not made and opportunities for prevention, anticipatory guidance, and treatment are missed or delayed. To address these difficulties related to performing multiple genetic tests we will perform rapid WGS on 100 neonatal patients (and their parents) being treated in the neonatal intensive care units (NICU) of Brazilian hospitals (Sabará Hospital Infantil and Hospital de Clínicas de Porto Alegre) with suspected rare Mendelian diseases that fit a pre-established selection criteria. WGS libraries will be sequenced on the NovaSeq 6000 platform at DASA laboratory. The final report will be returned in 15 days and will include the analysis of CNVs, indels, SNVs and a number of selected non-coding variants known to cause rare diseases. Some studies such as the one by Chung CCY et al. (2020) concluded that rapid exome sequencing in a pediatric clinical setting is feasible, has high diagnostic and clinical utility, and reduces healthcare utilization costs. With this project, we hope to shorten the time for accurate diagnosis of neonatal patients with rare Mendelian disease and critical medical conditions. We expect that timely accurate diagnoses will improve the management and treatment of these patients and lower costs to the overall health care system.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2342. Recombinant chromosome with proximal 14q duplication in two sibs with mental retardation and facial dysmorphism due to maternal translocation t(14q21;21p11).

Authors:

S. Tayel^{1,2}, A. Abd Rabuh²; ¹Alexandria Univ., Faculty of Med., Alexandria, Egypt, Egypt, ²Genetics Unit, Alexandria Regional Ctr. for Women's Hlth.and Dev., Alexandria,, Egypt

Abstract Body:

Introduction: Rearrangements involving the proximal segment of the long arm of chromosome 14 (14q) are rare. Clinical phenotypes of duplications 14q vary and genotype-phenotype correlation is difficult due to insufficient data. Material and Methods: Conventional cytogenetic analysis by GTG- banding and Fluorescence in situ Hybridization (FISH) using pre-labeled centromeric and telomeric probes for chromosomes 14 and 21 were performed on peripheral blood metaphase spreads of the different family members of the study. Results: We report two children with severe mental retardation, developmental delay, dysmorphic faces, microcephaly, cleft palate, and fingers and toes anomalies due to proximal duplication 14q (14q10;q21). They had the 47,XY,+mar karyotype. This marker was a small supernumerary chromosome sSMC resembling a small acrocentric one. GTG chromosome banding of the mother (who had 3 early abortions as well) and her sister (who had 4 neonatal deaths) revealed balanced translocation t(14;21) with the karyotype of: 46,XX,t(14;21)(q21.3;p11.2) 46,XX,t(14;21)(14pter→14q21.3::21p11.2→21pter;21qter→21p11.2::14q21.3→14qter). FISH analysis confirmed the cytogenetic finding and delineated the sSMC as a derivative chromosome 14. Conclusions: Carriers of Robertsonian translocations can give rise to unbalanced gametes through adjacent segregation in meiosis I. This family study adds to the specification of the genotype-phenotype correlation of proximal 14q duplications and addresses the importance of the proper complicated genetic counseling of Robertsonian translocation carriers prior to IVF, ICSI and PGD.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2343*. Return of incidental genetic findings to pediatric patients: considerations and opportunities from experiences in genome sequencing

Authors:

K. Bowling^{1,2}, M. Thompson², S. Scollon³, A. Slavotinek⁴, B. Powell⁵, B. Kirmse⁶, L. Hendon⁶, K. Brothers⁷, B. Korf⁸, G. Cooper², J. Grealley⁹, A. Hurst⁸; ¹Washington Univ. Sch. of Med., Saint Louis, MO, ²HudsonAlpha Inst. for Biotechnology, Huntsville, AL, ³Baylor Coll. Med., Houston, TX, ⁴Univ. of California San Francisco, San Francisco, CA, ⁵Univ North Carolina at Chapel Hill, Chapel Hill, NC, ⁶Univ. of Mississippi Med. Ctr., Jackson, MS, ⁷Univ. of Louisville, Louisville, KY, ⁸Univ Alabama at Birmingham, Birmingham, AL, ⁹Albert Einstein Coll. of Med., Bronx, NY

Abstract Body:

The uptake of comprehensive clinical and research genetic testing (e.g. exome/genome sequencing) has introduced unexpected testing results (incidental findings) that have become a major challenge for both testing laboratories and providers. While the American College of Medical Genetics and Genomics has outlined guidelines for laboratory management of clinically actionable secondary findings, debate remains as to whether incidental findings should be returned to patients, especially those representing pediatric populations. The Sequencing Analysis and Diagnostic Yield working group in the Clinical Sequencing Evidence-Generating Research Consortium has collected a cohort of pediatric patients found to harbor a genomic sequencing-detected incidental finding. The incidental variants were not thought to be associated with the indication for testing and were disclosed to patients and families. In total, 24 incidental findings were detected across 23 pediatric patients. These findings span four different research studies/laboratories and demonstrate management heterogeneity across study sites. We summarize specific cases to highlight core considerations that surround detection and return of unexpected findings (uncertainty of disease onset, disease severity, age of onset, clinical actionability, and personal utility), and suggest that interpretation of incidental findings in pediatric patients can be difficult given evolving phenotypes. Further, return of unexpected results can benefit patients and providers, but do present challenges. While there may be considerable benefit to return of incidental genetic findings, these findings can be burdensome to providers and present risk to patients. It is important that laboratories conducting genomic testing establish internal guidelines in anticipation of detection. Heterogeneity in policies related to these findings across clinical labs and research studies underscores the need and potential benefit for shared guidelines. However, future discussion is required to determine whether cohesive guidelines or policy statements are warranted.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2344. Review of Cytogenetic Findings of patients with Turner Syndrome and its variants among Filipinos from 1991 to 2020

Authors:

E. Maceda, C. Calalo, M. Abadingo, E. Salonga, J. Oblefias, C. Padilla; Univ. of the Philippines Manila, Manila, Philippines

Abstract Body:

A review of the karyotyping results of the Cytogenetics Laboratory of the Institute of Human Genetics, National Institutes of Health, University of the Philippines Manila showed that Turner Syndrome accounted for 2.64% of all samples received from 1991-2020. For 30 years, the most common karyotype in Turner syndrome was the classical Turner syndrome or the standard monosomy 45,X. It was noted in 195 patients, which is 37.69 % of all patients diagnosed with Turner syndrome. Mosaicism with a normal female karyotype was noted in 50 patients (9.62%). For the Turner syndrome variants, the most common is isochromosome Xq seen in 125 patients (24%). This is followed by Turner syndrome with marker chromosome in 55 patients (10.58%). Deletion Xp and deletion Xq were noted in 22 patients (4.23%) and 20 patients (3.85%), respectively. From this review, it can be noted that chromosomal analysis or standard karyotyping is a vital and useful diagnostic tool in Turner syndrome. The information obtained from it may be useful in clinical decision-making of families and healthcare providers. Its importance in providing adequate genetic counseling cannot be overemphasized.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2345. Somatic Mutational Landscape of Extracranial Arteriovenous Malformations and Phenotypic Correlations

Authors:

M. Eyries¹, F. EL SISSY¹, M. WASSEF², B. FAUCON³, D. SALVAN³, S. NADAUD⁴, F. COULET¹, H. ADLE-BIASSETTE², F. SOUBRIER⁴, A. BISDORFF⁵; ¹1. Sorbonne Université, Département de génétique, Assistance Publique-Hôpitaux de Paris, Hôpital Pitié-Salpêtrière, PARIS, France, ²2. Dept. of Pathology, Lariboisière Hosp., Assistance Publique-Hôpitaux de Paris, Univ. of Paris, Faculty of Med., PARIS, France, ³3. Dept. of Otorhinolaryngology, Lariboisière Hosp., Assistance Publique-Hôpitaux de Paris, PARIS, France, ⁴4. Sorbonne Université, INSERM UMR_S1166, Unité de recherche sur les maladies cardiovasculaires, le métabolisme et la nutrition, ICAN, PARIS, France, ⁵5. Dept. of Neuroradiology, Lariboisière Hosp., Assistance Publique-Hôpitaux de Paris, PARIS, France

Abstract Body:

Arteriovenous malformations (AVMs) are rare congenital fast-flow vascular anomalies characterized by a direct connection between arteries and veins, with the absence of a capillary network. Most AVMs arise in the brain, but they can occur anywhere in the body, with a predilection for the head and neck and are so-called extracranial AVMs. Somatic genetic variants may be the cause of extracranial AVMs, but few studies have explored these genetic anomalies, and no genotype-phenotype correlations have been identified. This study included twenty-three patients with extracranial AVMs that were confirmed clinically and treated by surgical resection, and for whom frozen tissue samples were available. Targeted next-generation sequencing analysis was performed using a gene panel that included vascular disease-related genes and tumor-related genes. We identified a pathogenic variant in 17 out of 23 samples (73.9%). Pathogenic variants were mainly located in *MAP2K1* (n=7) and *KRAS* (n=6), and more rarely in *BRAF* (n=2) and *RASAI* (n=2). *KRAS* variants were significantly ($p<0.005$) associated with severe extended facial arteriovenous malformations, for which relapse after surgical resection is frequently observed, while *MAP2K1* variants were significantly ($p<0.005$) associated with less severe, limited arteriovenous malformations located on the lips. Our study highlights a high prevalence of pathogenic somatic variants, predominantly in *MAP2K1* and *KRAS*, in extracranial arteriovenous malformations. In addition, our study identifies for the first time a correlation between the genotype, clinical severity and angiographic characteristics of extracranial AVMs. The *RAS/MAPK* variants identified in this study are known to be associated with malignant tumors for which targeted therapies have already been developed. Thus, identification of these somatic variants could lead to new therapeutic options to improve the management of patients with extracranial arteriovenous malformations.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2346. Statistical method for detection of uniparental disomy using SNP microarray or NGS technologies

Authors:

M. Roytman; Bionano Genomics, San Diego, CA

Abstract Body:

Uniparental disomy (UPD), or the inheritance of two paired chromosomes from the same parent, is challenging to detect given that it does not affect chromosome number or structure. However, UPD can cause human disease, due to disturbance of imprinted regions. Knowledge of the parent of origin for the UPD event is important, as maternal and paternal UPD can have significantly different phenotypic expression. Karyotyping and aCGH microarrays cannot detect UPD since there is no copy number change. In contrast, SNP microarrays and sequencing (at a depth sufficient to detect SNP genotypes) can be used to detect presence of UPD with provision of data from at least one parent. Here, we present a statistical approach using a Hidden Markov Model (HMM) to infer stretches of UPD in a patient-derived sample when provided genotypes for at least one parent. This method can detect both heterodisomy, in which both parental homologues are present, as well as isodisomy, in which two copies of one parental homologue are present. In addition, the approach can discriminate between UPD and absence of heterozygosity (AOH) caused by common stretches of DNA inherited from both parents (as in case of closely related parents). The HMM algorithm is built on a model which defines the likelihoods of observing a trio of genotypes, given each possible inheritance mode to include both maternal and paternal inheritance for the isodisomy and heterodisomy types. The HMM can detect both whole-chromosome and segmental UPD patterns, and outputs the parent of origin for UPD events. Results of the algorithm are seamlessly visualized and include color coding of individual SNP probes based on parent of origin, alongside existing copy number and allelic event calls so they can be interpreted in a holistic and whole-genome context.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2347. Strategies to Reduce Inconclusive Genetic Testing Results for Inherited Cardiomyopathies: Perspectives from a Clinical Diagnostic Laboratory

Authors:

L. Bronicki^{1,2}, C. Boswell-Patterson¹, C. Chisholm¹, L. Little¹, M. Ghani³, A. Smith⁴, L. Huang¹, G. Mettler¹, K. Boycott⁴, J. Rutberg¹, J. Richer¹, O. Jarinova¹; ¹CHEO, Ottawa, ON, Canada, ²Univ. of Ottawa, Ottawa, ON, Canada, ³The Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, ⁴Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada

Abstract Body:

Cardiomyopathies harbour a strong genetic component and can substantially increase the risk of sudden death. Cardiomyopathies, including hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM) and arrhythmogenic right ventricular cardiomyopathy (ARVC), are groups of disorders affecting the structure and function of the heart. Advancements in our understanding of the genetic causes of cardiomyopathies are rapidly evolving, with an increasing number of associated genes emerging every year. This, in turn, causes strain on genetic diagnostic laboratories, which are expected to keep pace with advancements in the field while providing cost-effective testing services with the highest possible analytical and clinical validity. At the CHEO Genetics Diagnostic Laboratory (GDL), we have implemented workflows that facilitate keeping our cardiomyopathy next-generation sequencing (NGS) testing relevant and accurate, without severely depleting resources. Two of these recent workflows include biennial revisions of our cardiomyopathy NGS testing menus and implementing current gene- and disease-specific variant interpretation guidelines. Both workflows are derived from internal and published data, and guidelines and recommendations set by the Clinical Genome Resource (ClinGen), an organization comprising of international clinical, laboratory and research experts in genetic diseases. Following these workflows for our pan cardiomyopathy, HCM and ARVC NGS tests reduced the number of genes or total sequence (base pairs) offered from 45 to 30, 19 to 12 and 7 to 7, respectively. We also implemented changes to variant interpretation guidelines, such as decreasing the general population allele frequency thresholds and criteria for interpreting truncating and missense variants in *TTN*. In addition to increasing efficiencies in the laboratory, these changes decreased the number of inconclusive results significantly, including by ~62% for HCM cases. Our data reaffirm that workflows aimed at updating gene testing menus and variant interpretation guidelines lead to improved analytical and clinical validity. These strategies ultimately allow us to continuously provide up-to-date testing services to patients, in a cost-effective and responsible manner.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2348*. The perplexing 22q11.2 duplication syndrome: an important cause of congenital anomalies, medical conditions, cognitive deficits, and behavioral phenotypes - or nothing at all

Authors:

D. McGinn^{1,2}, T. Crowley¹, K. Gaiser^{1,3}, V. Giunta¹, L. Lairson¹, M. Unolt⁴, M. Digilio⁵, B. Emanuel^{6,3}, E. Zackai^{6,3}, N. S. Philip⁷, S. Garcia-Minaur⁸, D. Heine-Suner⁹, A. Bassett¹⁰, A. Swillen¹¹, D. McDonald-McGinn^{6,3}; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Perelman Sch. of Med. of the Univ. of Pennsylvania, Philadelphia, PA, ³Perelman Sch. of Med. of the Univ. of Pennsylvania, Philadelphia, PA, ⁴Sapienza Univ. of Rome, Rome, Italy, ⁵Ospedale Bambino Gesù, Rome, Italy, ⁶Children's Hosp. of Philadelphia, Philadelphia, PA, ⁷Hosp d'Enfants de la Timone, Marseille Cedex 5, France, ⁸Hosp. Univ. rio La Paz, Madrid, Spain, ⁹Hosp. Univ. ri Son Espases, Palma de Mallorca, Spain, ¹⁰Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, ¹¹KU Leuven, Leuven, Belgium

Abstract Body:

Background: The chromosome 22q11.2 duplication syndrome (22q11.2DupS) is expected to be as common as the 22q11.2 deletion syndrome (22q11.2DS) resulting from meiotic non-allelic homologous recombination in patients with *de novo* CNVs. 22q11.2DupS and 22q11.2DS have been identified in 1 in 850 and 1 in 992 unselected fetuses respectively. Despite this prevalence, few 22q Centers have experience with 22q11.2DupS while families are seeking anticipatory guidance. Here we report findings on 295 patients with 22q11.2DupS including etiology (familial v. *de novo*), CNV size, and associated features in probands and affected relatives. **Methods:** Records were retrospectively reviewed under IRB approval on 295 individuals diagnosed between 20 weeks-gestation and 68 years. 211 were members of 133 families followed in Philadelphia, including 3 parents of children with 22q11.2DS. 84 patients were followed at 9 international collaborating sites including Leuven (36), Madrid (9), Mallorca (7), Marseilles (16), Rome (11), and Toronto (5). 53% and 59% of Philadelphia and international patients respectively were male. Mean age at diagnosis (MAAD) was 9.4 years and 10.4 years for Philadelphia and international cohorts respectively. However, MAAD was 4.3 years for probands and 24 years for affected relatives. **Results:** Of the Philadelphia and international cohorts, 78% and 72% were familial respectively. When familial, the diagnosis in the parent was often a surprise as several had advanced degrees. 22 siblings were also ascertained. Most probands were referred for developmental delay or autism: 53% and 12% in Philadelphia and 58% and 11% internationally. Associated medical features overlapped with 22q11.2DS, although frequencies were overall reduced. 55% of Philadelphia and 58% of international patients had a standard LCR22A-LCR22D duplication. **Conclusions:** Probands with 22q11.2DupS have features overlapping with 22q11.2DS including structural anomalies/medical conditions (cardiac, endocrine, ENT, GI, immune, neurologic, palate, and skeletal), neurocognitive and behavioral differences - but the overall frequency is reduced. Familial cases of 22q11.2DupS (72%-78%) are more common than 22q11.2DS (~10%) leading to identification of previously undiagnosed relatives with significant genetic counseling implications. Nested duplications (42%-45%) are frequent compared with 22q11.2DS (~15%). MAAD (9.4-10.4 years) is considerably greater than for 22q11.2DS (3.7 years) often sending families on a protracted diagnostic odyssey. While families look to 22q Centers for guidance, these findings offer important insights, but many discussion points remain.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2349. The power of whole genome sequencing: Identification of an undocumented homozygous partial gene deletion in a patient with a neurological phenotype.

Authors:

G. Fischer¹, F. XU², A. Dennis³, A. Yang⁴; ¹PreventionGenetics, Marshfield, WI, ²Prevention Genetics, Marshfield, WI, ³Oregon Hlth.& Sci. Univ., Portland, OR, ⁴Oregon Hlth.and Sci. Univ., Portland, OR

Abstract Body:

A nineteen-year-old male presenting with seizures, intellectual disability and regression, microcephaly, and various developmental features underwent whole genome sequencing (WGS) after an extensive previous genetic work-up. Previous testing included echocardiogram, metabolic and methylation studies; all collectively unremarkable. Sequence variants reported via epilepsy multi-gene panel, mitochondrial DNA, and whole exome sequencing were subsequently determined to be inherited, reducing the likelihood of their relevance to the patient's phenotype.

WGS trio analysis revealed a 434 base pair homozygous deletion encompassing nine nucleotides of exon 2 and 425 base pairs of intron 2 of *HNMT*. The homozygous deletion was subsequently confirmed via polymerase chain reaction (PCR) and WGS analysis of each parent supported their heterozygous status. A similar deletion has not been reported within copy number variant databases or the literature nor was consanguinity declared or suspected based on patient records and WGS results. *HNMT* encodes a brain-specific histamine N-methyltransferase important in the degradation of histamine. Pathogenic variants in *HNMT* are associated with autosomal recessive intellectual developmental disorder 51 (OMIM: #605238). Pathogenic homozygous sequence variants have been reported in the literature in at least 3 families. However, copy number variants (CNVs) have not been reported to date, and we classify this as a variant of uncertain significance. Antihistaminergics and histamine diet restriction have ameliorated some clinical features in affected patients, suggesting a potential treatment option.

This case study provides evidence for the power of WGS to identify novel structural variants not detected by other technologies that may benefit patients in a diagnostic odyssey. We will also discuss the strengths and limitations of different technologies to detect such CNVs.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2350. The utility of whole exome sequencing in atypical cases of heterogeneous neurological disorders

Authors:

Z. Ali¹, J. Klar², M. Jameel³, a. fatima⁴, S. Baig⁵, N. Dahl⁶; ¹Ctr. for Biotechnology & Microbiol., Univ. of Swat, Pakistan, Mingora, Pakistan, ²IGP, Uppsala, Sverige, Sweden, ³Dept. of Biological and BioMed. Sci., Aga Khan Univ., Karachi 74000, Pakistan, Karachi, Pakistan, ⁴Aga Khan Univ., Karachi, Karachi, Pakistan, ⁵Pakistan Sci. Fndn., Constitution Avenue Islamabad, Pakistan, Islamabad, Pakistan, ⁶Uppsala Univ, Uppsala, Sweden

Abstract Body:

Autosomal recessive spastic ataxia is a neurological disorder characterized by cerebellar ataxia, spasticity and peripheral neuropathy. It is an inherited disease caused by homozygous mutations in different genes involved in development of cerebellum. In this study, we investigated two consanguineous families consisting of eight affected individuals manifesting peripheral neuropathy, spastic ataxia, intellectual disability (ID), aggressive behavior and epilepsy. Using whole exome sequencing, we identified two novel truncating mutations in *SACS* gene [c.2656C > T, p.(Gln886*) and c.4756_4760delAATCA, p.(Asn1586Tyrfs*3)], thus diagnosing it as autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS). Brain Magnetic Resonance Imaging (MRI) performed on probands from each family showed classical cerebellar and pontine changes, typically associated with ARSACS. Besides these, multiple supratentorial abnormalities were also observed, that are likely contributing factors to the cognitive symptoms. ID, behavioral abnormalities and epilepsy have been reported in few cases of ARSACS but are not a part of the classical triad of symptoms that includes cerebellar ataxia, spasticity and peripheral neuropathy. Our study shows that non-motor symptoms such as cognitive decline, ID and behavioral abnormalities may be important and even dominating clinical features in atypical cases with *SACS* mutations. This call for attention as ID is a disabling condition that further complicates the state, ability and care planning in this group of patients. Our combined findings further the knowledge about the phenotypic spectrum and genetic variability associated with the *SACS* gene that will contribute towards its improve diagnostics and this study also demonstrated the usefulness of WES as a diagnostic tool in atypical cases or in case of missing clinical information.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2351. Towards dried blood spot-based RNAseq in routine diagnosis of genetic disease

Authors:

C. Beetz, R. Al-Ali, M. Radefeldt, S. Lemke, P. Bauer; Centogene, Rostock, Germany

Abstract Body:

Introduction A high proportion of genetic diagnostic reports remains negative or indeterminate. Transcriptomic data have the potential to (partially) close this gap, but RNA analysis faces specific challenges in routine diagnostic setting. We set out to develop an RNAseq protocol that uses dried blood spots (DBSs), i.e. a sample type that is both easy and cheap to collect, ship and store. **Methods** Several protocols were tested to optimize filter card type, method of RNA extraction, depletion of unwanted RNA species, library preparation and bioinformatics pipeline. **Results** Our optimized protocol applies a proprietary filter card (CentoCard®). 100 ng total RNA are extracted from two blood spots (corresponding to 100 µl blood). Up to 14 days after sampling, RINs are between 3 and 6. Hemoglobin RNA depletion and polyA-tail capture were superior to alternative methods for mRNA enrichment. Sequencing generates 8-16 Million uniquely mapped RNA-seq reads. Gene body plots show acceptable underrepresentation of 5' ends of mRNAs. Gene expression patterns from DBS highly correlate with patterns obtained from fresh blood. Variants and splicing abnormalities were successfully detected across a wide range of genes. **Conclusions** Analyzing RNA from DBS has great potential for rare disease diagnostics. Combined with the primary DNA-based observations, higher diagnostic yield and less diagnostic uncertainty can be expected. Our protocol will facilitate routine implementation of RNAseq as a valuable add on to standard genetic diagnostics.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2352. Transposable element insertions in 1000 Swedish individuals

Authors:

J. Eisfeldt¹, K. Bilgrav¹, D. Nilsson¹, H. Thonberg¹, E. Tham¹, A. Ameer², A. Lindstrand³; ¹Karolinska Inst.t, Solna, Sweden, ²Uppsala Univ, Uppsala, Sweden, ³Karolinska Inst.t, Solna, Sweden

Abstract Body:

The majority of rare diseases are genetic, and regardless of advanced high-throughput genomics-based investigations, around 60 % of patients remain undiagnosed. A major factor limiting our ability to identify disease-causing alterations is a poor understanding of the morbid and normal human genome. A major genomic contributor of which function and distribution remain largely unstudied are the transposable elements (TEs), which constitute 50 % of our genome. In order to resolve this knowledge gap and increase the diagnostic yield of rare disease patients, we characterized TEs in 1000 Swedish individuals from the SweGen dataset using an in-house TE identification pipeline with a median of 1748 novel TEs detected. The pipeline was utilized to call novel TEs in 2504 individuals from the 1000 Genomes Project (1KGP) and a database with resulting TEs was created. Filtering the SweGen data using the TE-database effectively reduced the median to 18 (< 1 %) TEs remaining, proving that most insertions are common across populations. Next, we aligned rare and common TEs to genomic elements and found them to be significantly enriched in intronic regions ($p < 0.001$) and depleted in exonic regions ($p < 0.001$). Finally, we applied our TE identification pipeline on two clinical cases where disease causing TEs were suspected and could verify the presence of pathogenic TE insertions in both. Altogether we demonstrate the importance of TE detection and highlight possible clinical implications in rare disease diagnostics.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2353. Two new Tunisian cases of achondroplasia/hypochondroplasia : phenotype-genotype correlation.

Authors:

B. Abdelmoula, N. Bouayed Abdelmoula; Med. Univ. of Sfax, Sfax, Tunisia

Abstract Body:

Background: Achondroplasia and hypochondroplasia are among the most common skeletal dysplasia and short-limbed dwarfing conditions. The two autosomal dominant entities are similar, but the features of achondroplasia tend to be more severe. This study aimed to report two novel Tunisian cases of achondroplasia/hypochondroplasia for whom clinical, radiological and genetic assessment were conducted. **Material and methods:** The first patient was a female teenager with a severe short stature and a subtle body disproportion. The second patient was a female child with macrocephaly (+2 DS), external hydrocephalus and a failure to thrive (-2 SD). The two patients had specific facial dysmorphism of achondroplasia/hypochondroplasia. DNA was isolated from blood samples and genetic assessment of FGFR3 was conducted using 1/targeted sequencing for the first patient to confirm the presence of the 1620C>G mutation of FGFR3, common in hypochondroplasia, and 2/ PCR-RFLP for the second patient to confirm the presence of the typical 1138G>A mutation of FGFR3, common in achondroplasia. **Results:** Despite the absence of respectively Asn540Lys and Gly380Arg mutations in our two cases and based on clinical and radiological characteristics, the diagnosis of hypochondroplasia associated to severe dysplastic pulmonic valvular stenosis was suggested for the first patient, whereas the diagnosis of achondroplasia was suggested for the second patient. **Conclusion:** Most cases of achondroplasia/hypochondroplasia are due to a principal de novo mutation in the fibroblast growth factor receptor 3 (FGFR3 located at 4p16.3) which is the Gly380Arg mutation revealed in 98% of affected patients by achondroplasia. FGFR3 mutations c.1620C>A or c.1620C>G (p.Asn540Lys) appear to be frequent in children with hypochondroplasia with a more severe phenotype. Mutations of FGFR3, which is a negative regulator of linear bone growth by decreasing chondrocyte proliferation and differentiation in the growth plate, activate the receptor resulting in the gain-of-function with inadequate growth of the affected tissues. The most relevant features to the diagnosis are the shortening of the arms, macrocephaly, short fingers, trident configuration and hypermobility of the hips and knees as well as specific facial dysmorphism. When the diagnosis is uncertain after careful clinical and radiological assessment, molecular testing may be suggested. Nevertheless, targeted mutation analysis is not sufficient in our environment.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2354. Ultra-low DNA input with long-read sequencing identified complex chromosomal rearrangements involving *NIPBL* in a Cornelia de Lange syndrome patient

Authors:

N. Jiang¹, E. Spiegel¹, C. Lambert², P. Baybayan², A. Wenger², K. Anyane-Yeboah¹, R. Wapner¹; ¹Columbia Univ. Med. Ctr., New York, NY, ²Pacific BioSci.s, Menlo Park, CA

Abstract Body:

Background: Cornelia de Lange syndrome (CdLS) is a multisystem disorder characterized by distinctive facial features, growth restriction, upper limb malformations and mild to severe intellectual disability. Over the years, many genes have been identified to be associated with CdLS in an autosomal dominant or X-linked manner. Among them, *NIPBL* is the most commonly mutated CdLS gene. Approximately, disease causing variants can be detected in ~80% of the CdLS patients including mosaicism which leaves many CdLS cases genetically unexplained due to limitations of current diagnostic technologies. Here we describe a CdLS neonate who went through a long genetically unexplained medical mystery since the prenatal stage.

Study Description: The patient was first seen in the MFM clinic of our institute. Fetal ultrasound showed multiple abnormalities. Postnatally, He had clinical features that were compatible with CdLS, including synophrys, thin upper lip, micromelia, hirsutism. His presentation was severe and he passed away in two days. A variety of genetics testings have been done on this patient since the prenatal stage, including Prenatal FISH, karyotype, chromosomal microarray(trio) and WES(trio). However, the results were either negative or inconclusive. No pathogenetic variants or even VOUSes were identified in the known CdLS genes. Interestingly, multiple de novo deletions were seen on chromosome 5p with sizes ranging from 32kb to 2.05Mb, including a de novo 32 kb deletion in intron 1 of *NIPBL*. We hypothesized there were complex chromosomal rearrangements involving 5p which disrupted *NIPBL*.

With a very limited amount of fetal DNA left, we then performed whole genome HiFi sequencing using the PacBio ultra-low DNA input protocol. With only 30 ng of DNA input, we achieved excellent sequencing performance with 20.5x coverage and 9.9 kb average insert size. Multiple complex structural variants were identified involving chromosomal region 5p15.2p13.1. *NIPBL* was disrupted by multiple inversions and translocations, including an exon 1 inversion.

Conclusion: We hereby report a unique CdLS case with de novo complex chromosome 5p rearrangements that disrupted *NIPBL*. These disease-causing variants were not identified initially due to the limitations of the genetic sequencing options currently offered in the clinical lab. With the PacBio ultra-low DNA input long-read sequencing, we were able to identify clinically relevant structural variants with very limited amount of DNA. This case demonstrates the potential utility of long-read sequencing in routine genetic testing to increase diagnostic rate and shorten the diagnostic odyssey for patients and their families.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2355. Unexpected CNV anomalies using exome or genome-wide approach for complex phenotypes: report of a Regional Genetic Center experience

Authors:

V. Plaiasu, D. Ozunu, G. Motei, M. Ivan, L. Ghita, B. Pascu, G. Bar; INSMC Alessandrescu-Rusescu, Bucharest, Romania

Abstract Body:

Background Copy number variation (CNV) is a common source of genomic variation and an important genetic cause of disease. Copy number variation disorders arise from the dosage imbalance of one or more genes, resulting from deletions, duplications or other genomic rearrangements (translocations and inversions) that lead to the loss or gain of genetic material. Karyotype and microarray analyses have served as gold standards in molecular diagnostics for CNVs. The identification of CNVs in exome/genome data can improve diagnostic yield for a broad genetic disorder spectrum ranging from complex neurodegenerative, neuropsychiatric, sensory disorders, intellectual disability to congenital abnormalities. **Material and methods** The patients were recruited from the Regional Centre of Medical Genetics in the Paediatrics Clinics of INSMC Alessandrescu-Rusescu Bucharest. We studied 6 patients (5 girls and 1 boy, 1 year-17 years) who were seen in a 3-year period (July 2019 - September 2021) with diverse phenotypic alterations: neurodevelopmental disorders, metabolic disorders, craniofacial anomalies for which exome or genome sequencing was performed in a diagnostic setting due to the suspicion of monogenic condition. Three exome, respectively three genome sequencing data were analyzed for genomic findings. **Results** Different CNVs anomalies were identified involving chromosomes 1, 3, 10, 15 and X and they were validated by Chromosomal microarray analysis. 1 gain CNV and 5 loss CNVs were found, with CNVs ranging in size from 18,121 bp to 11,2 Mb, which resulted in 6 conclusive diagnoses based on a pathogenic CNV. CNVs were considered clinically relevant if they contained one or more disease genes for which the phenotype described in the literature corresponded with the patient's phenotype. One patient presented with double genetic pathology: a CNV anomaly combined with a monogenic disorder. The patients' CNV results have overlap with 3q27.3 microdeletion syndrome, 10q24.31q24.32 microduplication syndrome, large one copy loss on the long arm of chromosome 1 including 68 genes, encompassing the entire *LHX4* gene, 1q44 microdeletion syndrome combined with autosomal recessive Congenital disorder of glycosylation, type Im, heterozygous exon 1 and 2 deletion of the *CHD2* gene, a large intronic hemizygous 1-copy loss of chromosome X, affecting partially the *DMD* gene. **Conclusion** With the introduction of NGS (Next Generation Sequencing) technologies, it is now possible to detect both SNVs (Single nucleotide variants) and CNVs using an exome- or genome-wide approach with a single test and to increase in the diagnostic yield without additional testing of rare conditions.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2356. Unmasking of a chromothripsis event using the integrated approach of Chromosomal Microarray Analysis (CMA) and Optical Genome Mapping (OGM)

Authors:

S. Loddo¹, S. Di Tommaso¹, G. Astrea², S. Genovese¹, V. Alesi¹, D. Lopergolo², A. Morgia¹, R. Milone², A. Novelli¹; ¹Bambino Gesù Childrens' Hosp., Rome, Italy, ²Fondazione Stella Maris, Pisa, Italy

Abstract Body:

The phenomenon of chromothripsis is defined as a "chromosomal catastrophe", characterized by the shattering of one or several chromosomes up to hundreds of fragments, restored in random order and orientation, leading to complex cryptic rearrangement. We describe the case of a 3-year-old girl presenting with global developmental delay, axial hypotonia, dysmorphisms, hypertrichosis, short stature, renal anomalies and familial hyperCKemia with myopathic features. Clinical exome analysis revealed no pathogenetic variants in genes associated with the reported clinical condition. CMA (Chromosomal Microarray Analysis) was then performed, showing a 2p21 microduplication, of about 317 Kb, and two non-contiguous microdeletions at 7q36.1 and 7q36.3 regions, respectively of 1.9 Mb and 643 Kb in length. All the microrearrangements arose de novo. We used the Optical Genome Mapping (OGM), a new technology able to provide information on structural chromosomal rearrangements at high resolution, in order to verify the possible correlation between the microrearrangements detected by CMA. The analysis revealed a cryptic event of chromothripsis with six chromosomal breaks on the long arm of a chromosome 7, resulting in loss of 7q36.1 and 7q36.3 regions, corresponding to CNVs (Copy Number Variations) previously identified, and a random assembly of the rest of genomic fragments. The 2p21 microduplication seems to be occurred independently and plays a limited role. We hypothesize that the complex clinical features of our patient is the results not only of gene loss, but also of positional effects and/or disruptions of coding sequences related to the chromothripsis phenomenon. The integrated use of high resolution cytogenomic techniques has proven to be the key approach for a better diagnostic defining and for the understanding of the molecular mechanism at the base of disease.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2357. Unusual X-chromosome composition in mother and daughter with different karyotype constitutions

Authors:

M. Melaragno, A. Di Battista, M. Moysés-Oliveira; Univ.e Federal de São Paulo, São Paulo, Brazil

Abstract Body:

X-chromosome rearrangements are usually associated with distinct X inactivation patterns, which can influence in patient's phenotype. We studied two patients, mother and daughter, carriers of the same altered X-chromosome with a 10,6 Mb Xp deletion and a 32,4 Mb Xq duplication, but present as a supernumerary chromosome in the daughter. The mother presented intellectual disability, short stature, facial and body asymmetry, and hypothyroidism. She had five abortions and a daughter who showed normal stature, mild intellectual disability, and severe visual impairment, as well as an aggressive head cancer, which led to her death at the age 26 years. The mother showed a cytogenomic result as: 46,X,der(X)(qter→q25::p22.2→qter).arr Xp22.33p22.2(60,701_10,626,284)×1,Xq25q28(122,793,915_155,208,244)×3, and the daughter: 47,XX,der(X)(qter→q25::p22.2→qter)mat. HUMARA and EdU incorporation assay revealed an extremely preferential X inactivation pattern of the abnormal X-chromosome in the mother (100:1) and in the daughter (91:9), who probably had casual inactivation of both normal X-chromosomes. Their phenotypic differences are probably due to genic dosage imbalance in regions that escape X inactivation. In the mother, the partial X-chromosome monosomy resulted in some features of Turner syndrome, while the marked asymmetry (with also renal agenesis and ovarian absence at the right) could not be explained. These severe phenotypes were not transmitted to her daughter, who had 47 chromosomes including two normal X chromosomes and the abnormal X-chromosome. The daughter had a cytogenetically and phenotypically normal son, a finding that can be comparable with literature data that indicates fertility in XXX females. Financial support: FAPESP (2019-21644-0).

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2358*. Updated clinical practice guidelines for managing children and adults with 22q11.2 deletion syndrome

Authors:

A. Bassett^{1,2,3}, S. Óskarsdóttir⁴, T. Crowley⁵, J. C. Loo², E. Boot⁶, **D. McDonald-McGinn**⁵; ¹Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, ²The Dalglish Family 22q Clinic, Univ. Hlth.Network, Toronto, ON, Canada, ³Dept. of Psychiatry, Univ. of Toronto, Toronto, ON, Canada, ⁴Queen Silvia Children's Hosp. and Dept. of Pediatrics, Sahlgrenska Academy at Univ. of Gothenburg, Gothenburg, Sweden, ⁵Children's Hosp. of Philadelphia, Philadelphia, PA, ⁶Advisium's Heeren Loo Zorggroep, Amersfoort, Netherlands

Abstract Body:

Background: Clinical practice guidelines for managing individuals with 22q11.2 deletion syndrome (22q11.2DS) were first published in 2011 and 2015 as part of an international collaborative effort. During the last decade knowledge has advanced with new research broadening our understanding of the many aspects of this multi-system condition over the lifespan. The aim of this investigation was to systematically review the literature and provide updated recommendations to facilitate optimal care while transcending nationalities, health care system differences, and subspecialty biases. **Methods:** The 22q11.2 Society recruited international 22q11.2DS experts to revise the clinical guidelines via a systematic literature review, abstracting relevant data and references, and preparing consensus recommendations, following the Preferred Reporting Items for Systematic reviews and Meta-Analysis protocol (PRISMA). In addition, leaders from eight 22q11.2DS patient advocacy organizations, based in seven countries on three continents and representing 7624 families, completed a REDCap survey, indicating top priorities for these guidelines including relevant subspecialty topics, best methods for knowledge transfer, and barriers to care. **Results:** 6018 publications were initially identified. After review, 2448 remained, 954 related to children/adolescents, 303 to adults, and 591 with uncertain/both pediatric and adult relevance. Recommendations were incorporated into pediatric and adult consensus documents. Given the evidence was limited for all studies (almost exclusively level III or IV), individual recommendations were not formally graded but provided as statements of good practice. Patient advocate survey results supported updated guidelines towards improving awareness, access, genetic testing, and genetic counseling, while suggesting knowledge transfer be shareable, portable, and available on social media. **Conclusions:** These guidelines present the best international consensus recommendations currently available during childhood and adolescence, with a major focus on changing issues across developmental stages. They include practical suggestions regarding evaluation, surveillance, and management of 22q11.2DS-associated physical, cognitive, behavioral and psychiatric co-morbidities, inclusive of genetic counseling and psychosocial issues. There is guidance regarding recommended investigations by age and at diagnosis, as well as suggested “do’s and don’ts”. Like the initial publications, these guidelines will continue to require updating as new information becomes available.

Molecular and Cytogenetic Diagnostics Posters - Thursday

PB2359. Utilization of the Age-based Semi Quantitative Metric (ASQM) to evaluate gene-condition pairs for inclusion in Early Check's newborn sequencing panel.

Authors:

H. Cope¹, L. Milko², C. M. Powell², N. deJong², J. Hunter¹, A. K. Foreman², B. Powell², S. Shone³, E. Jalazo², J. Berg², H. Peay¹; ¹RTI Intl., Research Triangle Park, NC, ²Univ North Carolina at Chapel Hill, Chapel Hill, NC, ³North Carolina State Lab. of Publ. Hlth., Raleigh, NC

Abstract Body:

While genomic sequencing as a primary method of screening remains on the horizon for state-run newborn screening (NBS) programs, increasing numbers of research studies and commercial laboratories offer new parents supplementary screening via sequencing panels. However, methods used to select genes for these panels are not explicit, and the included genes vary widely. Here we describe our utilization of the previously published Age-based Semi Quantitative Metric (ASQM) to select gene-condition pairs for inclusion in a newborn sequencing panel for Early Check, a voluntary NBS study available to ~120,000 babies who receive NBS in North Carolina annually. The ASQM was previously used to score >800 gene-condition pairs enriched for pediatric onset on five criteria of clinical actionability on a scale of 0-3: **severity** and **likelihood** of the manifested condition, **efficacy** and **acceptability** of the intervention, and **knowledge** about the condition and intervention. A multidisciplinary Early Check Gene Panel Working Group (WG) used the scores (prioritizing scores ≥ 13) to identify an initial set of clinically diverse gene-condition pairs (n=26) to assess for inclusion in the panel. Age of action (initiation of the intervention) was set at ≤ 2 years to ensure identification of babies with conditions actionable neonatally or in infancy. The WG developed curated evidence summaries and a review process culminating in a group discussion and vote to include on/exclude from the panel. Of the 26 initial gene-condition pairs assessed, 23 were approved for inclusion. The WG identified additional important factors during the review process that impacted decisions for some conditions. For example, certain genes and conditions with variable ages of action but well-studied genotype-phenotype correlations (e.g., *RET*: Multiple endocrine neoplasia IIB), were included if specific early-action pathogenic variants (e.g., p.M918T) could be identified for return. Other genes and conditions were excluded because they accounted for an exceedingly small proportion of the associated phenotype (e.g., *ZAP70*: Immunodeficiency) or lacked adequate evidence to support genotyping in the absence of a clinical phenotype (e.g., *TTPA*: Ataxia with isolated vitamin E deficiency). The ASQM provides a method for ongoing evaluation of gene-condition pairs for inclusion in newborn sequencing panels that aligns with current NBS criteria. The Early Check process provides a reproducible methodology for selecting gene-condition pairs for NBS sequencing pilot studies that emulates a future integration into public health.

Molecular and Cytogenetic Diagnostics Posters - Wednesday

PB2360. Value of DNA testing in the diagnosis of Sickle Cell Anemia in Childhood in an environment with a high prevalence of other causes of anemia.

Authors:

G. Mbayabo Gloire¹, P. Lumbala Kabuyi¹, M. Ngole¹, A. Lumaka², K. Devriendt³, P. Lukusa Tshilobo¹, C. Van Geet⁴; ¹Univ. of Kinshasa/KU Leuven, Kinshasa, Congo, Democratic Republic of the, ²Univ. of Kinshasa/Univ. of Liège, Kinshasa, Congo, Democratic Republic of the, ³Univ Hosp Leuven, Leuven, Belgium, ⁴KU Leuven, Leuven, Belgium

Abstract Body:

Background: Sickle cell anemia (SCA) is the most common genetic disease worldwide caused by a single mutation in the gene *HBB*. DNA testing can help clarifying the diagnosis when Hb electrophoresis is inconclusive. We evaluated the usefulness and feasibility of DNA-based diagnosis of SCA in rural Central Africa. **Methods:** This is a cross-sectional study conducted from November 2016 to end October 2017 in the Hôpital Saint Luc de Kisantu, located 120 km from Kinshasa. This hospital offers the management of SCA patients, mainly identified using the Sickling test (Emmel test) combined with clinical features. We included patients aged 6 months to 18 years locally diagnosed as SCA and we collected clinical and hematological data. All patients were offered Hb electrophoresis and DNA testing at the Center for Human Genetics of the University of Kinshasa. **Results:** This study included 160 patients. Hemoglobin capillary electrophoresis suggested that 136 (85%) were homozygote SS, 13 (8.1%) were heterozygote (AS) and 11 (6.9%) were homozygote normal (AA). DNA testing confirmed these electrophoresis findings, with the exception of four patients, two AS in electrophoresis were found SS due to recent transfusion, and two SS in electrophoresis were found AS because they have compound heterozygous form S/ β^0 -thalassemia. The diagnosis of SCA was therefore wrongly ascertained with Emmel test in 15% of patients. **Conclusion:** This study reveals a high proportion of false positive SCA diagnoses in a rural environment in Central Africa, and underlines the importance of DNA testing in clarifying the diagnosis of SCA in the case of doubtful Hb electrophoresis.

Epigenetics Posters - Thursday

PB2361. A Bayesian interaction model to estimate cell-type-specific Methylation Quantitative Trait Loci (meQTLs) incorporating priors from flow-sorted sequencing data

Authors:

Y. Cheng¹, B. Cai¹, H. Li¹, A. Justice², K. Xu³, H. Zhao⁴; ¹Yale Sch. of Publ. Hlth., New Haven, CT, ²Yale Sch. of Med., New Haven, CT, ³Yale Sch. Med., New Haven, CT, ⁴Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Abstract Body:

Methylation Quantitative Trait Loci (meQTLs) identification can shed lights on the complex interactions among genetic variants, DNA methylation (DNAm) patterns and complex human traits. The limited collection of large-scale, cell-type-specific (CTS) methylation profile hinders the extension of this framework from “bulk level” to “cell type level”. Although several methods have been developed to estimate cell-type-specific differentially methylated CpG sites with phenotypes, the field of cell-type-specific genetic effects on DNAm (CTS-meQTL) is less explored. In this presentation, we introduce a Bayesian interaction model to infer CTS meQTLs from bulk methylation data, and our model allows the incorporation of flow-sorted sequencing data of CTS methylome from a relatively small number of samples. By integrating bulk data and flow-sorted sequencing data, we show in simulations that our model improves the estimation of CTS genetic effects when compared to other methods that could provide sample-level CTS methylation estimates and then be applied to identify CTS meQTL as a comparison for our model. We applied our method to identify CTS-meQTL on cocaine use in a Veterans Aging Cohort Study Biomarker Cohort (N = 811). We were able to identify meQTLs on cocaine use related DNAm in granulocytes, lymphocytes, and monocytes. For example, the genetic effects for the CpG site cg03082779 (mapped to *ZMYND8*) were discovered in monocytes: nine meQTLs for this CpG were significant in monocytes ($p < 0.001$) while none of them were significant in granulocytes or lymphocytes. In contrast, for the CpG site cg00339441 (mapped to *RBM20*), we identified 17 meQTLs in lymphocytes while none of them were identified in the other two cell types. These findings highlight the importance of identifying genetic influences on methylation at cellular level. Our general framework will also facilitate combining CTS methylome for a small number of samples with the bulk data to enhance signal detections.

Epigenetics Posters - Wednesday

PB2363. A long short-term memory autoencoder approach for detecting CRISPR genomic edits.

Authors:

P. Hinson, M. Farris, P. Texter, T. Patterson III, E. Mace, K. Gemp, S. Russell; The Mitre Corp., McLean, VA

Abstract Body:

Clustered regularly interspaced short palindromic repeat (CRISPR) technology has revolutionized the way molecular genomic edits are implemented. As the accessibility of CRISPR genome-editing technology continues to expand in capability and application, there is an increased need to accurately identify all genomic modifications induced by CRISPR genomic edits, even those that may range beyond the modified nucleotides. Long short-term memory (LSTM) autoencoder machine learning techniques are proficient for anomaly detection within sequential data, like those observed within the methylation pattern of epigenome sequences. In this paper, we take an unsupervised learning approach using a LSTM autoencoder to examine anomalies within the sequential methylation patterns of CRISPR-edited mouse epigenomes, generated by whole genome bisulfite sequencing. Using 300 unedited mouse epigenomes from various tissues for training within a LSTM autoencoder, we developed a basis to reconstruct expected sequential biological noise across the mouse epigenome. The model was evaluated on 24 CRISPR-edited mouse epigenomes. The deep learning model includes an encoder that progressively decreases the number of LSTM units per layer and a decoder that progressively increases in units per layer until dimensionality of the layer matches the input dimensionality. These steps are performed using dropout within each LSTM layer to combat overfitting. To evaluate the effectiveness of our model in accurately predicting CRISPR edits, we developed an automated system that scans through all the sequences in a given genome and flags potential edits based on the reconstruction error for each of the sequences. Our findings demonstrate that CRISPR edit detection is possible by using CRISPR-edited epigenome samples for prediction and flagging CRISPR edits when reconstruction error surpasses an error threshold.

Epigenetics Posters - Thursday

PB2364. A methylation-based COVID-19 classification model to predict severe disease in vaccinated individuals

Authors:

G. Harrison¹, W. Zhou¹, M. Preethi Boorgula², M. Campbell², S. Chavan², B. Peterson¹, B. Barnes³, R. PORECHA³, R. Mathias⁴, A. Taye³, I. Yang², C. Gignoux², A. Monte², K. Barnes¹; ¹Tempus Labs, Chicago, IL, ²Univ. of Colorado, Denver, CO, ³Illumina, Inc., La Jolla, CA, ⁴Johns Hopkins Univ, Baltimore, MD

Abstract Body:

The development of vaccinations, antivirals, and other interventions, has precipitated a transition from the acute phase of the SARS-COV-2 pandemic to an endemic infection cycle. In this phase, mitigating the worst outcomes of an infection with SARS-COV-2, and managing burdens on hospital systems, is reliant on our ability to identify persons at high risk of developing severe COVID-19 disease and administering treatments efficiently. Epigenomic pattern alterations in response to SARS-COV-2 infection are evident in circulating white blood cells. Early in the pandemic, we designed a machine learning platform that leveraged methylation risk scores (MRS) derived from differentially methylated signatures in infected and uninfected individuals (Konigsberg et al 2021 *Comm. Med.*). This approach yielded highly predictive scores, measured as a classification-threshold-invariant (AUC), for both presence of SARS-CoV-2 infection (AUC=93.6%) and as a measure for COVID-19 disease severity (AUC=79.1%-84.4%). The goal of this work was to determine the prediction efficacy of a previously developed sparse regression based MRS model towards infection status and disease severity. The population in consideration included vaccinated individuals who visited the Emergency Department after March 2021. To accomplish this, we profiled peripheral blood samples from 304 additional patients (211 cases, 93 controls) on the customized Infinium MethylationEPIC array. These patients included a majority who had received at least one dose of the COVID-19 vaccine at the time of infection. Disease severity was assessed by hospitalization status, ICU admittance, administration of ventilator and death. This information was extracted from patient electronic health records. We assessed the efficacy of our previously developed sparse regression based MRS model in predicting disease severity in vaccinated individuals. This was achieved by comparing infection status and severity AUCs between the aforementioned cohort and an unvaccinated cohort from the original study. Our approach used a machine learning approach predicated on MRS to predict COVID-19 severity and demonstrated its utility in vaccinated and unvaccinated populations. Our study provides insights into how vaccination affects methylation status thereby offering protection from severe immune reactions due to deadly infections. In summary, our model previously designed and implemented on biospecimens from patients with SARS-CoV-2 infection prior to vaccination provides a powerful tool for clinicians to evaluate a patient's propensity to develop severe COVID-19 disease regardless of vaccination status.

Epigenetics Posters - Wednesday

PB2365. A simple workflow for methyl DNA analysis in FFPE-preserved Alzheimer and other neurodegenerative brain samples

Authors:

S. Jackson; Thermo Fisher Scientific, South San Francisco, CA

Abstract Body:

Neurodegenerative diseases, including Alzheimer and Parkinson diseases, affect millions of people and their families worldwide. Knowing the etiology, inheritance patterns and sensitivity factors are critical to understanding the pathology of these diseases. Changes in expression levels mediated by changes in DNA methylation patterns contribute to the pathologies. Several genes have been shown to be differentially methylated in these diseases. In this study, we demonstrate a workflow for analyzing methyl DNA at specific loci. Genomic DNA was extracted from FFPE-preserved samples and bisulfite converted. Primers specific for bisulfite-treated and unconverted sequences were used to amplify specific loci (ANK1 and HOX3). Resulting amplicons were cycle sequenced and analyzed on SeqStudio Flex instruments. This workflow provides researchers with a method for understanding changes in epigenetic patterns in neurodegenerative diseases.

Epigenetics Posters - Thursday

PB2366. A single cell chromatin accessibility and transcriptome atlas of the human heart improves the identification of risk variants and genes of Atrial Fibrillation

Authors:

A. Selewa¹, K. Luo¹, M. Wasney², L. Smith¹, H. Eckart², I. Moskowitz¹, A. Basu², X. He¹, S. Pott²; ¹Dept. of Human Genetics, Univ. of Chicago, Chicago, IL, ²Dept. of Med., Univ. of Chicago, Chicago, IL

Abstract Body:

Most genetic variants associated with complex traits from GWAS are located in noncoding regions with no clear functions, making it difficult to identify causal variants and understand their mechanisms. Single-cell genomics promises to close these gaps, by de-convoluting complex tissues into constituent cell types, and annotating functional elements in each cell type. In this work, we combined single cell RNA-seq and ATAC-seq to profile the human heart. We detected 8 major cell types and identified 349,000 open chromatin regions (OCRs), half of which are cell-type specific.

We used this dataset to study the genetics of Atrial Fibrillation (AF). AF is the most common cardiac arrhythmia and affects 33 million people world-wide. Among ~120 AF-associated loci, the causal variants have been identified and characterized only in a small number of loci. Using the single-cell data, we found that AF risk variants were >10-fold enriched in OCRs of cardiomyocytes (CMs) but not enriched in other cell types. Using a Bayesian fine-mapping procedure that leverages functional information, we identified 54 variants at posterior inclusion probability (PIP) > 0.5 (*i.e.* >50% probability of being causal), increasing the number of such variants by ~40% compared to fine-mapping without functional information. These SNPs show a high proportion of overlapping heart enhancers, and transcription factor (TF) targets. Several high confidence SNPs were experimentally validated.

We developed a novel computational procedure, Mapgen, that combines fine-mapping results with information linking SNPs to genes, to identify causal genes. This procedure identified 45 high-confidence risk genes. Using AF-related gene pathways to assess the “plausibility” of prioritized genes, we found that Mapgen performs much better than existing methods for gene prioritization, including nearest genes, Activity-by-Contact (ABC) scores and expression QTLs. Our results support the importance of gene regulatory networks centered on TFs, including *TBX5*, *GATA4*, *NKX2-5*, and *PITX2*, in the AF genetics. Together, our study provides a comprehensive map of AF risk variants and genes, and demonstrates the power of combining single-cell genomics and advanced computational techniques to reveal the genetic basis of complex traits.

Epigenetics Posters - Wednesday

PB2367. A single-cell atlas of 3D genomic structure and DNA methylation in human subcutaneous adipose tissue.

Authors:

Z. Chen^{1,2}, K. Abuhanna³, M. Alvarez³, Y. Zhang³, O. Avram^{4,1,2}, K. H. Pietiläinen⁵, E. Halperin^{1,2,3,4}, C. Luo³, P. Pajukanta^{3,6}; ¹Dept. of Computer Sci., Univ. of California, Los Angeles, Los Angeles, CA, ²Dept. of Computational Med., Univ. of California, Los Angeles, Los Angeles, CA, ³Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ⁴Dept. of Anesthesiology and Perioperative Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ⁵Obesity Res. Unit, Res. Program for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, ⁶Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract Body:

Obesity is a serious, fast-growing global health problem, predisposing to multiple cardiometabolic disorders, such as type 2 diabetes and non-alcoholic fatty liver disease. In obesity, subcutaneous adipose tissue forms the key fat depot that has to expand to accommodate the extra fat. In some obese individuals, this task and other important adipose tissue functions are hampered by the low-grade inflammation induced by obesity, thus driving the obesity co-morbidities. Exploring how obesity impacts the cell-type composition of subcutaneous adipose tissue could shed light on how proinflammatory changes and biological mechanisms lead to obesity-related cardiometabolic disorders. Studying genomic cytosine methylation allows for further understanding of gene regulation at the nucleotide level, particularly for cell-type specificity. Chromatin conformation analysis, in turn, elucidates gene regulation at the chromosomal level by delving into the three-dimensional structure of the genome. Recent advancements in sequencing technologies like single-nucleus methyl-3C sequencing (sn-m3C-seq) permit us to simultaneously capture the methylation and chromatin conformation profiles within the same single cell. We employed sn-m3C-seq on more than 6,600 human nuclei isolated from subcutaneous adipose tissue biopsies from five obese female patients undergoing liposuction, with a BMI of >30. Utilizing both methylation and chromatin conformation data, sn-m3C-seq robustly discerns seven major cell types: adipocytes, endothelial cells, mast cells, myeloid cells, perivascular cells, stromal cells, and T cells. Moreover, cell type clustering demonstrates high consistency when performed independently under each modality. We further reconstruct cell-type-specific chromatin conformation maps and present the first-ever single-cell atlas of 3D genome structure and DNA methylation in human subcutaneous adipose tissue. These data comprise an extensive resource for exploring the decomposition of bulk adipose tissue methylation data as well as cell-type specific differential DNA methylation and chromatin conformation in obese individuals.

Epigenetics Posters - Thursday

PB2368. Activation of *Xist* by an evolutionarily conserved function of KDM5C demethylase

Authors:

M. Samanta¹, S. Gayen², C. Harris¹, E. Maclary³, Y. Murata Nakamura¹, R. Malcore⁴, R. Porter⁵, P. Garay¹, C. Vallianatos¹, P. Samollow⁶, S. Iwase⁷, S. Kalantry⁸; ¹Univ. of Michigan, Ann Arbor, MI, ²Indian Inst. of Sci., Bangalore, India, ³Univ. of Utah, Salt Lake city, UT, ⁴Univ. of Michigan Ann Arbor, Ann Arbor, MI, ⁵Univ MICHIGAN, Ann Arbor, MI, ⁶Texas A&M Univ., College Station, TX, ⁷The Univ. of Michigan, Ann Arbor, MI, ⁸Univ. of Michigan Med. Sch., Ann Arbor, MI

Abstract Body:

XX female and *XY* male therian mammals equalize X-linked gene expression through the mitotically-stable transcriptional inactivation of one of the two X chromosomes in female somatic cells. Here, we describe an essential function of the X-linked homolog of an ancestral X-Y gene pair, *Kdm5c-Kdm5d*, in the expression of *Xist* lncRNA, which is required for stable X-inactivation. Ablation of *Kdm5c* function in females results in a significant reduction in *Xist* RNA expression. *Kdm5c* encodes a demethylase that enhances *Xist* expression by converting histone H3K4me2/3 modifications into H3K4me1. Ectopic expression of mouse and human *KDM5C*, but not the Y-linked homolog *KDM5D*, induces *Xist* in male mouse embryonic stem cells (mESCs). Similarly, marsupial (opossum) *Kdm5c* but not *Kdm5d* also upregulates *Xist* in male mESCs, despite marsupials lacking *Xist*, suggesting that the KDM5C function that activates *Xist* in eutherians is strongly conserved and predates the divergence of eutherian and metatherian mammals. In support, prototherian (platypus) *Kdm5c* also induces *Xist* in male mESCs. Together, our data suggest that eutherian mammals co-opted the ancestral demethylase KDM5C during sex chromosome evolution to upregulate *Xist* for the female specific induction of X-inactivation.

Epigenetics Posters - Wednesday

PB2369. Additive effects of stress and alcohol exposure on accelerated epigenetic aging in Alcohol Use Disorder

Authors:

J. Jung¹, D. McCartney², J. Wagner¹, J. Yoo¹, A. Bell¹, L. Mavromatis¹, D. Rosoff¹, C. Hodgkinson³, H. Sun³, M. Schwandt³, N. Diazgranados³, A. Smith⁴, V. Michopoulos⁴, A. Powers⁴, J. Stevens⁴, B. Bradley⁴, N. Fani⁴, R. Walker², A. Campbell², D. Porteous², A. McIntosh², S. Horvath⁵, R. Marioni², K. Evans², D. Goldman³, F. Lohoff¹; ¹Natl. Inst. on Alcohol Abuse and Alcoholism, Bethesda, MD, ²Univ. of Edinburgh, Edinburgh, United Kingdom, ³Natl. Inst. on Alcohol Abuse and Alcoholism, Rockville, MD, ⁴Emory Univ. Sch. of Med., Atlanta, GA, ⁵Univ California los Angeles, Los Angeles, CA

Abstract Body:

Background: Stress contributes to premature aging and susceptibility to alcohol use disorder (AUD) and AUD itself is a factor in premature aging; however, the interrelationships of stress, AUD and premature aging are poorly understood. We constructed a composite score of stress (CSS) from thirteen stress-related outcomes in a discovery cohort of 317 individuals with AUD and controls. We then developed a novel methylation score of stress (MS Stress) as a proxy of CSS comprising 211 CpGs selected by a penalized regression model. The effects of MS Stress on health outcomes and epigenetic aging were assessed in a sample of 615 AUD patients and controls using epigenetic clocks and DNAm telomere length (DNAmTL). Statistical analysis with an additive model using MS Stress and a methylation score for alcohol consumption (MS Alcohol) were conducted. Results were replicated in two independent cohorts (Generation Scotland GS n=7028 and the Grady Trauma Project GTP n=795). CSS and MS Stress were strongly associated with heavy alcohol consumption, trauma experience, epigenetic age acceleration (EAA) and shortened DNAmTL in AUD. Together, MS Stress and MS Alcohol additively showed strong stepwise increases in EAA. Replication analyses showed robust association between MS Stress and EAA in the GS and GTP cohort. A methylation-derived score tracking stress exposure is associated with various stress-related phenotypes and EAA. Stress and alcohol have additive effects on aging, offering new insights into the pathophysiology of premature aging in AUD, and potentially, other aspects of gene dysregulation in this disorder.

Epigenetics Posters - Thursday

PB2370*. Alcohol use-disorder associated DNA methylation differences in human nucleus accumbens and dorsolateral pre-frontal cortex

Authors:

J. White¹, C. Willis¹, S. Han², R. Tao², A. Deep-Soboslay², R. D. Mayfield³, B. T. Webb¹, E. O. Johnson¹, J. E. Kleinman², L. J. Bierut⁴, D. B. Hancock¹; ¹RTI Intl., Raleigh, NC, ²Lieber Inst. for Brain Dev., Baltimore, MD, ³The Univ. of Texas at Austin, Austin, TX, ⁴Washington Univ Sch Med, St Louis, MO

Abstract Body:

Alcohol use disorder (AUD), characterized by compulsive alcohol seeking despite adverse social, occupational, or health consequences, is a leading cause of preventable death worldwide. AUD has a strong genetic component: 50-60% heritability, with many loci identified in increasingly large genome-wide association studies. However, our understanding of the regulation of genes associated with AUD and their tissue specificity is limited. Previous studies have identified differential DNA methylation (DNAm), an epigenetic mechanism of gene regulation, associated with alcohol consumption and AUD in peripheral blood as well as postmortem brain tissue. However, recent work has suggested that there is little overlap in significant findings from different sample types, implying tissue specificity but also potentially due to different analytical strategies and small sample sizes, especially in studies of postmortem brain tissue. Here, we report a large epigenome-wide association study (EWAS) of AUD, focusing on two key tissues in the addiction cycle, nucleus accumbens (NAc) and dorsolateral prefrontal cortex (DLPFC). Illumina HumanMethylation EPIC array data from 122 decedents (61 cases with 2+ AUD symptoms, 61 controls with no AUD symptoms) were analyzed using robust linear regression, with adjustment for major depression diagnosis, age at death, sex, current smoking, DNAm-derived principal components, assay plate, and row position on plate. Results were corrected for inflation using the empirical null distribution estimation method implemented in *bacon*. At a false discovery rate (FDR) < 0.05, we identified 84 CpGs (annotated to 71 genes) significantly associated with AUD status for NAc and 55 CpGs (47 genes) for DLPFC. Three genes important to stress response and neuronal cell repair reached FDR significance in both NAc and DLPFC in our study: *PHF10* (novel), required for the renewal and proliferation of neural stem cells, *C11orf73* (novel), involved in protein transport following heat stress, and *RREB1* (previously implicated), a transcription factor implicated in axon death. Six CpGs (4 in NAc, 2 in DLPFC) were within 500 kb of results previously reported in EWAS of AUD in other postmortem brain tissues - ventral striatum, caudate nucleus, and BA9 and BA10 of the PFC. Additionally, 46 CpGs (27 in NAc, 19 in DLPFC) were annotated to genes previously implicated in EWAS of alcohol-related traits in blood. Taken together, our results highlight a subset of CpGs that may represent alcohol-related associations shared across blood and brain as well as a broader set of CpGs that may represent associations shared among and specific to brain regions important in addiction.

Epigenetics Posters - Wednesday

PB2371. Alterations of regulatory regions in T-cell Prolymphocytic Leukemia.

Authors:

H. Yan, S. Tian, H. Jin-Lee, H. Zhang, P. Hampel, E. Klee, W. Ding; Mayo Clinic, Rochester, MN

Abstract Body:

T cell prolymphocytic leukemia (T-PLL) is a rare disease showing a rapid clinical course with a median survival of <1 year. T-PLL patients respond poorly to conventional chemotherapy and demonstrate inevitable relapse after immunotherapy due to acquired resistance. Prevalence of structural variants, most notably *inv(14)(q11q32)*, *t(14;14)(q11;q32)* and *t(X;14)(q28;q11)*, has been identified. Recurrent somatic mutations were also identified in genes encoding chromatin regulators and in those associated with the JAK-STAT signaling pathway. Alterations of regulatory regions have been implicated in disease susceptibility and progression, in particular for distal regulatory elements such as enhancers. Given a lack of genome-wide epigenomics studies, the extent of epigenetic changes has not been well defined in T-PLL. We used histone H3 lysine 27 acetylation (H3K27ac) ChIP-seq to identify active enhancers and micrococcal nuclease digestion of linker DNA and sequencing (MNase-seq) to identify open chromatin regions in patients (n=6) and age-matched healthy individuals (n=3). Samples were collected with written consent and approval from the institutional review board at Mayo Clinic. Paired-end reads were mapped to the human reference sequence hg19 using BWA, and peaks were identified by MACS2 and filtered by blacklisted regions. Unsupervised clustering for both H3K27ac peaks and open chromatin regions revealed two distinct clusters separating T-PLL cases from normal, indicating a genome-wide change of regulatory regions in T-PLL. Specifically, we revealed a strong trend of losing H3K27ac-marked active enhancers in T-PLL, which was coupled with the down-regulation of nearby genes based on RNA-seq data. These genes are enriched in the immune system and adaptive immune response pathway. Similarly, we identified over 2,500 regions that lost open chromatin regions in T-PLL, but only 376 regions that gained open chromatin regions. A large portion of enhancers that lost H3K27ac in T-PLL became primed enhancers still carrying H3K4me1, while the gained active enhancers were often associated with an increase in chromatin accessibility. Together, our analyses provide further insights into the epigenetic alterations in T-PLL.

Epigenetics Posters - Thursday

PB2372*. Altered and allele-specific open chromatin landscape reveals epigenetic and genetic regulators of innate immunity in COVID-19

Authors:

Z. Zhang^{1,2,3,4}, B. Zhang^{1,2}, V. A. Koeken^{5,2,1}, S. Kumar^{1,2}, M. Aillaud⁶, H-C. Tsay^{1,2}, Z. Liu^{1,2}, K. R. M. Anke^{1,2,7,8,9}, C. Soon^{2,7}, I. Odak¹⁰, B. Bošnjak¹⁰, A. Vlot¹¹, Deutsche COVID-19 OMICS Initiative (DeCOI), M. A. Swertz^{3,4}, U. Ohler¹¹, R. Geffers¹², T. Illig^{13,14}, A-E. Saliba¹⁵, L. E. Sander^{16,17}, R. Förster^{8,9,10}, C-J. Xu^{1,2,5,18}, M. Comberg^{1,2,7,8,9}, L. N. Schulte^{6,17}, Y. Li^{1,2,5,9}; ¹Dept. of Computational Biology for Individualised Infection Med., Ctr. for Individualised Infection Med., Hannover, Germany, ²TWINCORE, a joint venture between the Helmholtz-Ctr. for Infection Res. (HZI) and the Hannover Med. Sch. (MHH), Hannover, Germany, ³Genomics Coordination Ctr., Univ. of Groningen and Univ. Med. Ctr. Groningen, Groningen, Netherlands, ⁴Dept. of Genetics, Univ. of Groningen and Univ. Med. Ctr. Groningen, Groningen, Netherlands, ⁵Dept. of Internal Med. and Radboud Ctr. for Infectious Diseases, Nijmegen, Netherlands, ⁶Inst. for Lung Res., Philipps Univ., Marburg, Germany, ⁷Dept. of Gastroenterology, Hepatology and Endocrinology, Hannover Med. Sch., Hannover, Germany, ⁸German Ctr. for Infection Res. (Deutsches Zentrum für Infektionsforschung DZIF), Partner Site Hannover-Braunschweig, Hannover, Germany, ⁹Cluster of Excellence Resolving Infection Susceptibility (RESIST; EXC 2155), Hannover Med. Sch., Hannover, Germany, ¹⁰Inst. of Immunology, Hannover Med. Sch., Hannover, Germany, ¹¹Berlin Inst. for Med. Systems Biology (BIMSB), Max Delbrück Ctr. for Molecular Med. in the Helmholtz Association (MDC), Berlin, Germany, ¹²Genome Analytics, Helmholtz-Ctr. for Infection Res. (HZI), Braunschweig, Germany, ¹³German Ctr. for Lung Res. (DZL), NA, Germany, ¹⁴Hannover Unified Biobank, Hannover Med. Sch., Hannover, Germany, ¹⁵Helmholtz Inst. for RNA-based Infection Res. (HIRI), Helmholtz-Ctr. for Infection Res. (HZI), Würzburg, Germany, ¹⁶Charité - Univ.smedizin Berlin, Dept. of Infectious Diseases and Respiratory Med., Charité, Univ.smedizin Berlin, Berlin, Germany, ¹⁷German Ctr. for Lung Res. (DZL), Na, Germany, ¹⁸Dept. of Gastroenterology, Hepatology and Endocrinology, Hannover Med. Sch., Hannover, Netherlands

Abstract Body:

While SARS-CoV-2 infection causes mild respiratory disease in the majority of individuals, a small group of patients develop severe COVID-19. Dysfunctional innate immune responses have been identified to contribute to differences in COVID-19 severity, but the key regulators are still unknown. Here, we present an integrative single-cell epigenetics, transcriptomics, and genetics analysis of peripheral blood mononuclear cells from hospitalized and convalescent COVID-19 patients. In classical monocytes, we identified 41.3% of significantly up-regulated genes in hospitalized COVID-19 patients potentially induced by differential chromatin accessibility. Sub-clustering and motif-enrichment analyses of monocytes reveal disease condition-specific regulation by transcription factors, such as C/EBPs and SPI1, and their targets, including a long-noncoding RNA *LUCATI*, which further regulates interferon responses and is associated with the need for oxygen supply of COVID-19 patients. The interaction between C/EBPs and *LUCATI* was validated through loss-of-function experiments. Finally, we investigated genetic risk variants that exhibit allele-specific open chromatin (ASoC) in promoters/enhancers of COVID-19 patients. Integrating our data with publicly available expression quantitative trait loci and chromosomal interactions indicates that ASoC SNP rs6800484-C is associated with lower expression of *CCR2*, which may contribute to higher viral loads in lungs and higher risk of COVID-19 hospitalization. Altogether, our study highlights the diverse genetic and epigenetic regulators that contribute to the innate immune responses of different COVID-19 patients.

Epigenetics Posters - Wednesday

PB2373. Analysis of methylation QTLs in breast cancer characterises the influence of germline variation on the abnormal cancer methylome

Authors:

R. Hannah, D. Sproul; Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract Body:

Introduction:

Population genetics has unravelled an important relationship between epigenetic modifications and transcriptional activity. One example is the observation that gene activity associates with DNA methylation patterns. Abnormal changes to the methylome are a pathological feature of many cellular disease and a hallmark of cancer. While significant changes in the DNA methylome of cancer are evident, the mechanisms underpinning these alterations are still unclear. Investigating the mechanisms and directionality of the interplay between DNA methylation and gene regulation may therefore reveal important insights into tumorigenesis. Previous studies have established the effects of genetic variation on the methylome of normal tissue. However, a similar characterisation of the genetic effects on the cancer methylome is missing from literature.

Methods:

Current models suggest that methylation changes may occur in an allele-dependent manner whereby genetic variation may associate with both local and global changes in the methylome. Methylation QTL (methQTL) analyses allow for the implication of genetic loci in regulating CpG methylation changes. Additionally, comparisons of functionally annotated methQTLs between tissues can help characterise their overall methylome identity. I compared methQTLs between normal and breast tissue samples to understand how local associations change during tumorigenesis. Focusing on methylation and genotype data from the cancer genome atlas (TCGA), I developed a robust pipeline for identifying methQTLs in cancer cells. I evaluated established processing steps for these datasets (including normalisation approaches, quality control filtering and addressing population structure bias in the genotype data) and consequently justified an optimal analysis model for the cohort.

Results:

I performed a methQTL analysis using germline genotype data and matched tumour methylation data in 333 subjects after robust data processing and filtering. A preliminary analysis using non-imputed genotype data confirms that germline variation does associate with changes in the breast cancer methylome.

Epigenetics Posters - Thursday

PB2374. *ASXL3* links chromatin biology to neurodevelopment disorders.

Authors:

E. Peirent¹, Y-C. Tsan², C. Ryan³, S. Bielas^{1,2,4,3}; ¹NeuroSci. Graduate Program, Univ. of Michigan Med. Sch., Ann Arbor, MI, ²Dept. of Human Genetics, Univ. of Michigan Med. Sch., Ann Arbor, MI, ³Cell and Molecular Biology Program, Univ. of Michigan Med. Sch., Ann Arbor, MI, ⁴Dept. of Pediatrics, Univ. of Michigan Med. Sch., Ann Arbor, MI

Abstract Body:

Human genetics studies of neurodevelopmental disorders highlight chromatin's importance in corticogenesis, with pathogenic variants enriched in networks linked to chromatin regulation. Dynamic regulation of histone modifications is especially critical for the transcriptional plasticity required during this cellular differentiation. One such modification is mono-ubiquitination of histone H2A (H2AUb1), a conserved, traditionally repressive histone mark that is reversed by the polycomb repressive deubiquitinase (PR-DUB) complex. We identified *de novo* dominant truncating variants in *ASXL3*, a key component of PR-DUB, as the genetic basis of both Bainbridge Ropers Syndrome (BRS) and autism spectrum disorder (ASD), characterized by failure to thrive, global developmental delay, feeding problems, hypotonia, profound speech deficits, and intellectual disability. We identified dysregulation of H2AUb1 as a key molecular pathology in primary cells derived from individuals with BRS. To investigate this *ASXL3*-dependent neuropathology in early corticogenesis, we generated 3D human cerebral organoids that histologically and molecularly recapitulate the context-dependent features of *in vivo* cortical development in a reproducible manner. In organoid models from CRISPR-edited and patient-derived iPSC lines, we observe *ASXL3*-dependent defects in the differentiation of pluripotent cells to the full spectrum of mature cortical neuron subtypes. We then utilized transcriptomic and epigenomic techniques to probe the role of *ASXL3*-dependent H2AUb1 deubiquitination in regulating transcriptional profiles critical to NPC fate decisions during corticogenesis. Together, our functional investigation of BRS- and ASD-associated genetic variants provides molecular insights towards uncovering the elusive role of H2AUb1 in neural development.

Epigenetics Posters - Wednesday

PB2375. Atypical Prader-Willi and Angelman syndrome deletion: Importance of parent of origin detection.

Authors:

N. Al-Sweel, B. Richardson, E. Palen, A. Baxter, M. Martin; Bionano Genomics, San Diego, CA

Abstract Body:

Prader-Willi (PWS) and Angelman syndrome (AS) are imprinting disorders caused by genetic defects of the 15q11.2q13 region that are molecularly distinguished based on whether the defect exists on the maternally- or paternally-derived chromosome. The majority of PWS and AS cases are caused by deletions of the paternal or maternal 15q11.2q13 region, respectively. Depending on a person's age and symptoms, clinical correlation can aid in the distinction of these two syndromes; however, determining the parent of origin is ultimately necessary to molecularly confirm the diagnosis and provide families and providers with the information needed for appropriate medical management. Traditionally, methylation analysis is the gold standard follow-up testing when a deletion of the PWS/AS region is identified by CNV analysis. However, methylation testing is not always informative, particularly in instances of atypical deletions of this region. Here we report an atypical 15q11.2 deletion of the PWS/AS critical region, identified on CMA, in a subject with global developmental delays, possible aspiration, overeating, hypotonia, and behavioral and sensory challenges. There is noted clinical overlap with both PWS and AS making clinical correlation challenging. The 791 kilobase deletion included *UBE3A*, but due to a gap in probe coverage, it is unclear if it extends to *SNURF-SNRPN*, which includes the imprinting center. To date, seven individuals with smaller atypical deletions of the PWS/AS region on the paternally-derived chromosome 15 have been reported with features of PWS. Because the imprinting center was undisturbed and methylation testing was normal, other methodologies to determine the parent of origin of these atypical deletions were necessary. Commercial testing options and interpretation software with the ability to identify and report on parent of origin for atypical deletions are not widely available due to lack of clinical validation for this purpose and insufficient informative genotyping markers within the region of interest. This case highlights the importance of parent of origin testing in cases of atypical AS/PWS deletions, and the utility and ease of use of the NxClinical software to accurately classify genetic conditions that exhibit a parent of origin effect, particularly in instances when traditional methylation analysis is uninformative.

Epigenetics Posters - Thursday

PB2376*. Cardiac development in mammalian models lacking the PR-DUB component ASXL3.

Authors:

S. Regan¹, B. McGrath¹, Y-C. Tsan¹, A. Srivastava², S. Bielas¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Sanjay Gandhi PostGraduate Inst. of Med. Sci., Lucknow, India

Abstract Body:

ASXL3 (*Additional-sex combs like 3*) is the genetic basis of Bainbridge Ropers Syndrome (BRS). *ASXL3* is a component of the polycomb repressive deubiquitination (PR-DUB) complex, that deubiquitinates histone H2A mono-ubiquitination of (H2AUb1), highlighting a critical role for H2AUb1 in the pathogenic mechanisms of neurodevelopmental disorders. *De novo* dominant *ASXL3* truncating variants result in elevated H2AUb1 and altered transcriptional regulation. To understand epigenomic alterations and early development defects caused by truncating variants of *Asxl3*, we used CRISPR/Cas9-mediated genome editing to generate clinically relevant truncating variants in a novel mouse model of BRS. Constitutive loss of *Asxl3* results in highly penetrant developmental heart defects and perinatal lethality. *Asxl3*^{-/-} mice display severe right ventricle cardiac hypoplasia. An analysis of structural heart defects between the *Asxl1*, *Asxl2* and *Asxl3* null mouse models implicate nonredundant functions during heart development despite their interchangeable incorporation into the PR-DUB complex. This has implications for the three distinct developmental disorders associated with these genes. To understand the role of *Asxl3* in cardiac differentiation we have generated cardiac-fated tissue differentiated from *ASXL3*^{-/-} hESCs. Comparing the developmental mechanisms and single cell transcriptional changes revealed similar ECM pathology in heart phenotypes with species specific differential expression. These findings underscore the importance of *ASXL3* and family members in polycomb transcriptional repression during development.

Epigenetics Posters - Wednesday

PB2377. Catalysis of histone H3K27me3 by Polycomb-independent mechanisms.

Authors:

I. Mercado-Hernandez¹, K. Kasliwal², C. Harris¹, S. Kalantry¹; ¹Univ. of Michigan Med. Sch., Ann Arbor, MI, ²Weill Cornell Med., New York, NY

Abstract Body:

Histone marks are post translational modifications on histone tails that are associated with distinct transcriptional states. Histone H3 lysine 27 trimethylation (H3K27me3) is a facultative chromatin mark that often marks silenced genes. H3K27me3 is believed to be deposited solely by the Polycomb Repressive Complex 2 (PRC2). Insights into PRC2 and H3K27me3 functions have occurred through the study of X-chromosome inactivation. X-inactivation results in the silencing of genes on one of the two X-chromosomes in female cells to equalize X-linked gene expression between female and male mammals. The silencing of X-linked genes in X-inactivation occurs in significant part through chromatin modifications, including H3K27me3. PRC2 proteins and H3K27me3 are enriched on the inactive X-chromosome. We investigated PRC2 function in X-inactivation by ablating core PRC2 subunits in mouse extraembryonic endoderm (XEN) stem cells, which are a model of imprinted X-inactivation of the paternal X-chromosome. Unexpectedly, we found that loss of the PRC2 core proteins had a minimal impact on gene expression and X-inactivation. Upon examining H3K27me3 in XEN cells lacking PRC2 function, we found that H3K27me3 was not abrogated, thus controverting the existing models of H3K27me3 catalysis solely by PRC2. We therefore sought to identify non-PRC2 catalysts of H3K27me3.

Epigenetics Posters - Wednesday

PB2378. Characterization of nucleosome positioning and DNA methylation signatures using prostate cancer cells from diverse ethnic groups using NOME-EM-seq.

Authors:

L. Gonzalez-Smith, L. Yuri, E. Nelson-Moore, S. Rhie; Univ. of Southern California, Los Angeles, CA

Abstract Body:

Prostate cancer is the second most prevalent malignancy for men worldwide with over 1.4 million new cases and 400 thousand fatalities every year. However, this is not uniform amongst patients with different racial and ethnic backgrounds. For example, men of African ancestry have higher incidence of prostate cancer and more aggressive prostate cancer cases with more than double the mortality rate compared to men of European ancestry. Therefore, it is crucial to characterize tumor samples from different ethnic and racial backgrounds to better understand prostate cancer subgroups. Most cancer studies are centered around genetic variants, but more recent research has shown that epigenetic alterations, such as DNA methylation, can play a role in prostate carcinogenesis. Here, we have characterized normal prostate and prostate cancer cells from diverse ethnic groups using Nucleosome Occupancy Methylome Sequencing (NOME-seq), an epigenetic method that can profile endogenous methylation and nucleosome positioning simultaneously. NOME-seq traditionally utilizes bisulfite (NOME-bisulfite-seq) as its means of deaminating unmethylated cytosines converting them into uracils, allowing for differentiation from methylated cytosines upon downstream sequencing. Although successful, bisulfite is a harsh chemical that induces DNA damage and fragmentation which can decrease mappability in downstream bioinformatic analyses. Novel enzymatic alternatives (NOME-EM-seq) utilizes oxidizing enzyme TET2 and mRNA editing enzyme APOBEC2, as substitutes for bisulfite treatment. By performing NOME-bisulfite-seq and NOME-EM-seq methods in prostate cells and tissue samples from diverse ethnic groups, we characterized the epigenomes of prostate cells, identifying differentially nucleosome-positioned and methylated regions between normal prostate and prostate cancer cells and among ethnic groups. This study will help understand the underpinnings of prostate cancer specific epigenomic alterations and ethnicity-related differences.

Epigenetics Posters - Thursday

PB2379. Characterization Of Pten Expression Cis-Regulatory Elements Involved In Cowden Disease

Authors:

T. Matis^{1,2,3}, M. Lalanne^{2,3}, I. Lamrissi-Garcia^{2,3}, I. Moranvillier^{2,3}, E. Darbo^{2,3}, S. Dabernat^{2,3}, M. Longy^{1,3}, N. Sévenet^{1,2,3}; ¹Inst. Bergonié, Bordeaux, France, ²Université de Bordeaux, Bordeaux, France, ³BoRdeaux Inst. of Oncology, Bordeaux, France

Abstract Body:

PTEN Hamartoma Tumor Syndrome is a hereditary cancer predisposition syndrome characterized by the appearance of benign and/or malignant tumors affecting multiple organs and is associated with heterogeneous clinical traits, resulting from a germline pathogenic variant (GPV) in *PTEN*. *PTEN*'s GPVs, are also found in a wide phenotypic spectrum ranging from macrocephalus-autism syndrome to certain juvenile polyposis. The extension of the phenotypic spectrum was accompanied in the clinical laboratory by an increase in *PTEN* germline analysis indication. As a result, the identification of *PTEN*'s GPV led to defining a PTEN Hamartoma Tumor Syndrome (PHTS). In the last decade, the number of patients with a suspicious PHTS phenotype without any identified *PTEN*'s GPV variant has increased and is estimated at 20%, raising the question of missing heritability. The recent identification of *PTEN* structural modifications through Alu insertion in the coding region prompts us to continue exploring the allelic heterogeneity by hypothesizing that cis-regulatory region alterations may alter *PTEN* expression. *PTEN* expression is ubiquitous and controlled through a poorly defined promoter and unknown enhancers. Enhancers can target promoters by chromatin approximation in the 3D structure known to be organized in topologically associated domains (TAD). We propose that differences in *PTEN* genotype-phenotype can be explained by a loss of expression due to a TAD disruption or alterations of cis-regulatory noncoding regions. Using ENCODE data, we identified 4 putative enhancers located in 3' in a 1Mb-TAD. Our aim is to verify and confirm their role in *PTEN*'s TAD by Hi-C analysis as well as define their location by Tiled-C analysis to obtain an accurate definition of the *PTEN* cis-regulation region. and assess whether *PTEN*'s TAD may be altered using 8 cell lines, carrier of an increasing *PTEN* alteration state.

Epigenetics Posters - Thursday

PB2381. ChromGene: Gene-Based Modeling of Epigenomic Data

Authors:

A. Jaroszewicz, J. Ernst; Univ. of California Los Angeles, Los Angeles, CA

Abstract Body:

Various computational approaches have been developed to annotate epigenomes on a per-position basis by modeling combinatorial and spatial patterns within epigenomic data. However, such annotations are less suitable for gene-based analyses, in which a single annotation for each gene is desired. To address this, we developed ChromGene, which annotates genes based on the combinatorial and spatial patterns of multiple epigenomic marks across the gene body and flanking regions. Specifically, ChromGene models epigenomics maps using a mixture of hidden Markov models learned *de novo*, and is implemented on top of ChromHMM. Based on a ChromGene model with 12 mixture components, we generated annotations for the human protein-coding genes for over 100 cell and tissue types. We characterized the different mixture components based on their chromatin marks and relationship to external data. We found that gene expression levels varied substantially between some ChromGene annotations, but also that some ChromGene annotations with similar expression levels were associated with distinct sets of chromatin marks, showing that ChromGene captures biological information beyond gene expression. We identified specific ChromGene annotations enriched for genes with high probability of being loss of function intolerant, and found substantial variation among gene sets with similar expression distributions, further highlighting the additional information ChromGene captures relative to gene expression. We compared ChromGene to baselines that averaged epigenomic mark presence throughout the gene body or only considered annotations at transcription start sites. We found ChromGene annotations were more predictive of gene expression, and they yielded more significant enrichments for both cancer associated gene sets and Gene Ontology gene sets. Additionally, ChromGene annotations were less likely to be simply reflecting gene length than gene averaging annotations. We also identified specific ChromGene gene annotations that were relatively cell type specific, and others that were more constitutive across cell types. We expect that ChromGene and generated annotations will be a useful resource for gene-based epigenomic analyses.

Epigenetics Posters - Wednesday

PB2382. Clinical Epigenomic Testing in Canada: Discovery and Clinical Assessment of Episignatures

Authors:

H. McConkey¹, J. Kerkhof², M. Levy³, R. Relator³, K. Rooney³, A. Foroutan¹, S. Haghshenas¹, B. Sadikovic⁴; ¹Western Univ., London, ON, Canada, ²London Hlth.Sci. Ctr., London, ON, Canada, ³London Hlth.Sci. Ctr., London, ON, Canada, ⁴LHSC, London, ON, Canada

Abstract Body:

Introduction: Neurodevelopmental disorders (NDDs) often present with overlapping phenotypic characteristics, making a clinical diagnosis difficult. First-tier genetic testing includes microarray gene sequencing based on the phenotypic presentation. Approximately 75% of patients do not receive a diagnosis from this testing and must undergo reflex testing, which includes large gene panels, whole exome sequencing or whole genome sequencing. This testing can take months or years and is costly, and patients can be left with no variants detected or a variant of uncertain significance. DNA methylation is another mechanism that can impact gene expression. Expanding number of NDDs exhibit unique DNA methylation patterns in peripheral blood, called episignatures, which can be used as diagnostic biomarkers. EpiSign is a test that can determine if a patient's methylation pattern matches an episignature specific to a NDD. This study aims to assess the use of EpiSign in clinical setting, to expand the number of detectable episignatures, and assess the biological significance of genomic methylation changes in NDDs. **Methods:** EpiSign analysis is performed to assess genome wide DNA methylation using Illumina Infinium methylation arrays, then compares the detected profiles to known NDD episignatures. EpiSign-CAN is a Canadian national study designed to assess clinical utility and health system impact of EpiSign first-tier and reflex test setting. Second study objective is to expand the clinical utility of the EpiSign test through discovery of additional epigenetic signatures. **Results:** EpiSign includes more than 120 episignatures. Episignatures can be protein complex, gene, sub-gene, protein domain and even single nucleotide-level specificity. The EpiSign-CAN study is underway and enrolling patients in both the first-tier screening phase of their journey and in the reflex testing stage with 19% episignature positive results (June 2022). **Discussion:** The use of EpiSign in Canadian genetics clinics provides an additional strategy for physicians to assess patients with ambiguous clinical presentation or genetic findings. Additionally, it has the potential to impact healthcare resource allocation and provide a more cost-effective approach for the diagnosis of rare disease.

Epigenetics Posters - Thursday

PB2383. Conserved GATA4/5/6-mediated gene regulation in cardiac development and disease

Authors:

M. Wilson^{1,2}, M. Song^{1,2}, A. Prentice^{1,2}, X. Yuan^{3,1}, H. Hou¹, E. Erkut^{1,2}, C. Carlisle^{1,2}, S. Monis^{1,2}, E. Duncan¹, C. Somerville¹, R. Kim¹, R. Jobling⁴, I. Scott^{1,2}; ¹Hosp. for Sick Children, Toronto, ON, Canada, ²Univ. of Toronto, Toronto, ON, Canada, ³DFKZ, Heidelberg, Germany, ⁴The Hosp. for Sick Children, Toronto, ON, Canada

Abstract Body:

Early cardiac development is orchestrated by a core set of deeply conserved cardiac transcription factors (TFs) and the *cis*-regulatory elements (CREs) they bind. Mutations in these TFs, or the CREs that regulate them, can lead to congenital heart disease (CHD). However, we still have an incomplete knowledge of the CREs required for normal heart development. To identify cardiac CREs essential for heart development, we performed a comparative zebrafish-human epigenomic study focussing on the GATA4/5/6 family of TFs. GATA4/5/6 function at or near the top of the cardiac regulatory network hierarchy in animals and are among the earliest TFs expressed in the cardiac mesoderm. We first profiled the open chromatin landscape of *gata5*-expressing pre-cardiac cells in mid-gastrulation zebrafish embryos. Using a *Gata5/6* knockdown, we identified 1,470 differentially accessible regions (DARs) with reduced accessibility upon *Gata5/6* loss. Indicative of a direct role of *Gata5/6* in establishing chromatin accessibility, these closed DARs showed a strong enrichment for GATA motifs. We identified 47 mesendodermal-specific GATA-dependent DARs as accessible regions conserved between zebrafish and human, which we termed GATA-dependent accessible conserved non-coding elements (GaCNEs). 17 out of 18 GaCNEs tested so far displayed cardiac activity in transgenic zebrafish embryos. Supporting their functional conservation, three GaCNEs were identified as being GATA4 targets in human cardiomyocytes and accessible in cardiac progenitors. Zebrafish deletions of GaCNE1, which forms long range interactions with zebrafish *hand2* (130 kb) and human *HAND2* (460 kb), resulted in reduced *hand2* expression and laterality defects, including cardiac patterning defects. We found the remaining two GaCNEs contain hits for ultra-rare human variants in patients with unsolved congenital heart disease. One such region near *TBX20* (GaCNE20) interacts with the *TBX20* promoter in the human mesoderm. Both the zebrafish (zGaCNE20) and orthologous human (hGaCNE20) sequences can spatiotemporally recapitulate endogenous zebrafish *tbx20* expression. Initial work modelling the ultra-rare variants seen in hGaCNE20 in zebrafish enhancer reporter assays revealed that the mutant sequence drives ectopic skeletal muscle expression. Collectively, we identified zebrafish-human conserved GATA-dependent cardiac CREs, which likely contribute to vertebrate cardiac development. Given the proximity of many conserved accessible non-coding elements to dosage sensitive trans-factors, human genetic variation impacting these regions warrants further functional study.

Epigenetics Posters - Wednesday

PB2384*. Cord blood DNA methylation alterations potentially mediate associations between adverse pregnancy outcomes and childhood blood pressure

Authors:

J. Hu^{1,2}, J. Li^{1,2}, X. Hong³, G. Wang³, F. Hu^{1,2}, X. Wang³, L. Liang¹; ¹Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ²Brigham and Women's Hosp., Boston, MA, ³Johns Hopkins Univ. Bloomberg Sch. of Publ. Hlth., Baltimore, MD

Abstract Body:

Background: High child blood pressure (BP) contributes to higher risk cardiovascular diseases in adulthood. Adverse pregnancy outcomes (APOs; e.g., gestational hypertension, pre-eclampsia, gestational diabetes, preterm birth, and low birth weight) have been associated with child BP. Cord blood DNA methylation (DNAm) alterations have been associated with APOs and child BP. However, it remains unknown whether cord blood DNAm mediates the associations between APOs and child BP.

Methods and results: This study included 855 mother-newborn pairs from the Boston Birth Cohort. Genome-wide DNAm was profiled on cord blood samples using Illumina MethylationEPIC BeadChip; 721,395 DNAm sites on autosomes and the X chromosome were eligible for analysis after quality control steps. Repeated measures of child systolic (SBP) and diastolic (DBP) BP from age 3 to 10 years (5,123 observations) were used. SBP and DBP percentiles were calculated based on the 2017 American Academy of Pediatrics Clinical Practice Guideline. **First**, using linear mixed-effects model, we observed 2,875 DNAm sites significantly associated with repeated measures of DBP percentiles (FDR<0.05) after adjusting for age at BP measurement, newborn sex, maternal characteristics (age at delivery, pre-pregnancy body mass index [BMI], self-reported race, education, and smoking) and cord blood cell compositions. No DNAm sites showed significant associations with SBP percentiles. **Second**, we used elastic net with 10-fold cross validation and identified DNAm signatures for BP percentiles for each age from 3 to 10 years. The identified DNAm signatures were predictive of SBP and DBP percentiles at age 3 and 4 years ($p \leq 0.01$). **Third**, we observed significant sex ($p < 1.2 \times 10^{-6}$) and race (maternal self-reported Black vs. others; $p < 0.04$) differences in DNAm signatures for SBP and DBP percentiles at age 3 and 4 years; the DNAm signatures were associated with pre-pregnancy BMI ($\beta = 0.05$; $p = 0.03$), severe pre-eclampsia ($\beta = 1.32$; $p = 0.02$), gestational diabetes ($\beta = 2.24$; $p = 0.04$), smoking during pregnancy ($\beta = -1.32$; $p = 0.03$), and gestational age at delivery ($\beta = -0.19$; $p = 0.03$). **Finally**, in mediation analysis, DNAm signatures mediated 14-26% of associations of severe pre-eclampsia, gestational diabetes, smoking during pregnancy, and gestational age at delivery with child BP percentiles at age 3 and 4 years, with marginal significance ($p = 0.06-0.08$).

Conclusions: Cord blood DNAm signatures were associated with APOs and child BP, and might mediate associations between APOs and child BP at younger ages. These findings indicate that DNAm changes may underline the link between APOs and offspring future cardiovascular risk.

Epigenetics Posters - Thursday

PB2385. Could the difference in DNA methylation status explain phenotypic variability in patients with 5p- syndrome?

Authors:

V. Almeida¹, S. Chehimi², G. Carvalho¹, Y. Gasparini³, A. Nascimento⁴, L. Liro¹, B. Wolff⁴, L. Kulikowski⁵; ¹Faculdade de Med. da Univ.e de São Paulo, São Paulo, Brazil, ²Faculdade de Med. da USP, Sao Paulo, Brazil, ³Univ. of São Paulo, São Paulo, Brazil, ⁴Faculdade de Med. na Univ.e de São Paulo, São Paulo, Brazil, ⁵Univ.e de Sao Paulo, São Paulo, Brazil

Abstract Body:

Cri Du Chat Syndrome or 5p- Syndrome (OMIM #123450) is characterized by a genomic loss in the short arm of chromosome 5 and by variable clinical manifestations, that include high-pitched cry in newborns. The phenotypic variability in this syndrome may not be limited only to variations in gene structure - such as deletions, duplications, inversions, insertions and translocations - as DNA methylation mechanisms, which occurs mainly in the “CpG Islands”, are also possible. Therefore, we studied the DNA methylation of the remaining allele of region of breakpoint at 5p- in fifteen patients. DNA samples from fifteen patients with 5p- delimited by genomic array HumanCyto850K BeadChip were evaluated using the Infinium MethylationEPIC BeadChip platform. The bioinformatics analysis was performed in R programming language. We noticed a significant difference in methylation status in the remaining allele at the breakpoint of region of 5p- in the patients. We infer that this difference in methylation status may explain the phenotypic differences in patients with this syndrome.

Epigenetics Posters - Wednesday

PB2386. Critical epigenomics association between methylation and clinical phenotype

Authors:

L. Kulikowski¹, G. Carvalho¹, L. Vieira², Y. Gasparini³, V. Almeida², A. Nascimento^{4,1}, C. Kim⁵; ¹Univ.e de Sao Paulo, São Paulo, Brazil, ²Faculdade de Med. da Univ.e de São Paulo, São Paulo, Brazil, ³Univ. of São Paulo, São Paulo, Brazil, ^{4,5}Inst. da Criança, Sao Paulo, Brazil

Abstract Body:

Introduction: Copy number variations (CNVs) are DNA fragments deleted or duplicated in relation to a reference genome. Some genome variants, classified as variants with uncertain significance (VUS), do not present a safe conclusion with clinical phenotypes, which makes the diagnostic conclusion difficult. We know that hypermethylation of gene promoter regions often leads to transcriptional silencing, in addition to DNA methylation changes in gene and/or intergenic regions can play a critical role in genomic regulation and stability. With epigenomic investigation, we can gain a more comprehensive understanding of the modulation in gene expression associated with genomic imbalance in different diseases and in different genomic variations. **Materials and Methods:** In order to better understand the impact of methylation status associated with genomic imbalances, in clinical phenotypes, our study evaluated, using the Infinium MethylationEPIC BeadChip platform, 10 DNA samples from patients with intellectual disability and dysmorphic features with CNVs classified as VUS without a definitive clinical diagnosis, in addition to six control samples from participants without clinical conditions. **Results:** We were able to identify significant differences when comparisons of methylation status were performed individually. The ontological analyzes of methylation results suggest a phenotypic impact consistent with the clinical presentation of the patients, indicating a possible association between the phenotype and the DNA methylation status. **Conclusion:** Genomic methylation, associated with genomic structural imbalances, can play a critical role in clinical features and can be used as an important marker for patients with uncertain clinical diagnosis.

Epigenetics Posters - Thursday

PB2387. Cross-species and tissue imputation of species-level DNA methylation samples.

Authors:

E. Maciejewski¹, S. Horvath^{1,2}, J. Ernst¹; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²Altos Labs, San Diego, CA

Abstract Body:

DNA methylation is an epigenetic marker that is widely profiled in humans, but for many other mammals there is more limited data available. The ability to accurately profile DNA methylation across mammalian species at loci highly conserved at the sequence level was recently vastly enhanced by the development of the mammalian methylation array. However, biological samples from certain tissues in some species, including human, can be difficult to obtain. This motivates the development of methods to accurately impute DNA methylation samples representing species and tissue type combinations that were not experimentally profiled, but other tissues were profiled in the species of interest and the tissue of interest was profiled in other species. However, existing methods for DNA methylation imputation do not take advantage of said cross-species DNA methylation information. To address this limitation, we designed a method to leverage cross-species DNA methylation to predict mean DNA methylation levels for combinations of tissues and species for which such data was not previously available. The method uses a neural network, specifically a Conditional Variational Autoencoder (CVAE) with optimal hyperparameters selected via grid search, to generate a prediction of DNA methylation values for unobserved combinations of species and tissue types in its input compendium. We simulate different training scenarios by applying this method to various subsets of 13,245 samples generated by the Mammalian Methylation consortium representing 746 species-tissue combinations from 59 tissues. We show that the method when applied in cross-validation correlates well with held-out ground truth values and outperforms multiple baselines. We demonstrate that the method yields high-quality imputed samples even with limited other tissue types in a target species available. We also show that the predictions effectively preserve species and tissue relationships among samples. After performing cross-validation analyses, we use all available data to impute mean methylation samples for species-tissue combinations that have not previously been experimentally profiled. We expect the method and imputed data resource we have developed will be useful for DNA methylation analyses of species and tissue-level characteristics across mammalian species.

Epigenetics Posters - Wednesday

PB2388*. Cross-tissue patterns of DNA hypomethylation reveal genetically distinct histories of cell development

Authors:

T. Scott¹, T. Hansen², E. McArthur¹, E. Dorans¹, E. Hodges^{1,2}; ¹Vanderbilt Genetics Inst., Vanderbilt Univ., Nashville, TN, ²Dept. of Biochemistry, Vanderbilt Univ., Nashville, TN

Abstract Body:

DNA methylation (DNAm) is essential for balanced multi-lineage cellular differentiation, but exactly how DNAm drives cellular phenotypes is unclear. While >80% of CpG sites are stably methylated, tens of thousands of discrete CpG loci form hypomethylated regions (HMRs). Because they lack DNAm, HMRs are considered transcriptionally permissive, but not all HMRs participate in genome regulation. Contrary to long-held views about the role of DNAm in controlling promoter function, most promoter HMRs are invariable across cell-types regardless of transcriptional status. In contrast, a subset of HMRs is cell-type specific and enriched for tissue specific enhancers. Recently, we showed that, while enhancer HMRs correlate with chromatin accessibility and other indicators of permissive chromatin, their temporal dynamics are distinct from chromatin—HMRs can persist long after accessibility is lost. These studies suggest that establishment of HMRs is an important step in enforcing cell identity, and that patterns of HMR establishment are more complex than previously appreciated.

To understand their functional significance, we systematically dissected HMR patterns across diverse cell types and developmental timepoints, including human embryonic stem cells, fetal spine and heart, adult hematopoietic lineage, liver, and adrenal tissues. Unsupervised clustering of 102,390 distinct HMRs revealed that levels of HMR “sharedness” across cell-types reflects a developmental hierarchy supported by enrichment of stage-specific transcription factors and gene ontologies. Using a pseudo-time course of embryonic to adult hematopoietic development, we further show that a majority of HMRs observed in differentiated cells (~70-75%) are established at early developmental stages and accumulate as development progresses. HMRs that arise during differentiation frequently (~35%) establish near existing HMRs (≤ 6 kb away), leading to the formation of HMR clusters associated with stronger enhancer activity. Using SNP-based partitioned heritability from GWAS summary statistics of diverse traits and clinical lab values, we discovered that genetic contribution to trait heritability is enriched within HMRs. Moreover, heritability of cell-relevant traits increases with both increasing developmental specificity and HMR clustering, supporting the role for distinct HMR subsets in regulating normal cell function. Altogether, our findings reveal that HMRs can predict cellular phenotype by providing genetically distinct historical records of a cell’s journey through development, and this attribute distinguishes DNAm from other epigenomic features.

Epigenetics Posters - Thursday

PB2389. Defining the gene regulatory roles of non-coding variants in the pathogenesis of autism.

Authors:

E. Sosa, J. Greally; Albert Einstein Coll. of Med., Bronx, NY

Abstract Body:

Autism spectrum disorder (ASD) is a neurodevelopmental disorder of variable severity characterized by restricted interests/behaviors and difficulties with social interactions. Advanced genomic sequencing approaches applied to large autism cohorts have successfully identified pathogenic variants in genes, partially explaining the genetic underpinnings of the phenotype. However, diagnostic sequencing using these insights only yields a diagnosis in a limited proportion of cases ($\approx 17\%$), leaving most families without answers. While it is increasingly clear that variants in the non-coding majority of the human genome are mediating human diseases, we currently lack the tools to predict pathogenicity of variants outside coding sequences. We hypothesize that non-coding variants fulfill pathogenic roles in genetically unresolved cases of autism by disrupting the transcriptional regulatory landscape of neuronal cells, resulting in disturbed synaptic function and excitatory-inhibitory balance.

Our goal in this project is to understand how de novo variants (DNVs) in non-coding DNA cause autism. We have generated preliminary data from whole genome sequencing (WGS) data that show an enrichment of DNVs in cis-regulatory sequences from brain cell types. Using genomic editing in induced pluripotent stem cells (iPSCs) followed by in vitro differentiation to GABAergic neurons, we will test high-priority candidate pathogenic DNVs for molecular genomic and cellular phenotypes, not only compared with unedited cells, but also cell lines engineered with known pathogenic coding sequence variants.

We have generated preliminary data combining WGS from $\approx 11,000$ individuals from the MSSNG cohort (Autism Speaks) with ATAC-seq data from 54 cell types showing the locations of cis-regulatory loci. We show that DNVs in children with autism are significantly enriched at the cis-regulatory open chromatin regions (OCRs) in both glial cells and certain neuronal types, and that these OCRs with DNVs are located near genes implicated in autism. Through permutation analyses we have determined that the burden of DNVs in the cis-regulatory loci of brain cells is significantly enriched. This finding was reconfirmed in the Simons Simplex Collection (SSC) cohort (1,902 individuals), and validated using the TOPMed cohort, a negative control group not selected for neurodevelopmental disorders. Finally, we will introduce prioritized DNVs into iPSCs to understand their contributions to neuronal morphology. This approach will provide a powerful tool for identifying and modelling the effects of non-coding variants, enhancing our ability to diagnose and understand autism.

Epigenetics Posters - Wednesday

PB2390*. Defining the genetic background of immunoglobulin G galactosylation in humans

Authors:

A. Frkatovic¹, A. Mijakovac², K. Miscec², O. Polašek³, K. Fischer⁴, M. Beekman⁵, M. Wuhler⁵, M. Schulze^{6,7}, C. Wittenbecher^{6,7}, C. Gieger⁸, T. Spector⁹, A. Köttgen^{10,11}, C. Hayward¹², J. Wilson^{12,13}, F. Vuckovic¹, J. Kristic¹, V. Zoldos², L. Klaric¹², G. Lauc^{1,14}; ¹Genos, Zagreb, Croatia, ²Faculty of Sci., Univ. of Zagreb, Zagreb, Croatia, ³Sch. of Med., Univ. of Split, Split, Croatia, ⁴Inst. of Mathematics and Statistics, Univ. of Tartu, Tartu, Estonia, ⁵Leiden Univ. Med. Ctr., Leiden, Netherlands, ⁶German Inst. of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany, ⁷German Ctr. for Diabetes Res. (DZD), Neuherberg, Germany, ⁸Helmholtz Zentrum Munchen, Neuherberg, Germany, ⁹King's Coll. London, London, United Kingdom, ¹⁰Faculty of Med. and Med. Ctr., Univ. of Freiburg, Freiburg, Germany, ¹¹Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ¹²Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, ¹³Usher Inst., Univ. of Edinburgh, Edinburgh, United Kingdom, ¹⁴Faculty of Pharmacy and Biochemistry, Univ. of Zagreb, Zagreb, Croatia

Abstract Body:

Glycosylation is the key posttranslational modification of IgG that affects its stability and function. However, genetic mechanisms behind this essential structural modification are not well known. Previous genome-wide association studies (GWAS) cumulatively identified 29 genomic regions that associate with IgG glycosylation, but only a few of these associations were functionally validated. One of the key functional aspects of IgG glycome composition is galactosylation, the presence of galactose units in the glycan structure. In this study, we performed GWAS of IgG galactosylation phenotypes (agalactosylation, monogalactosylation and digalactosylation) in seven cohorts of European descent with a total of 13,705 individuals. We identified 16 genome-wide significant genomic loci at $p \leq 2.5 \times 10^{-8}$ and replicated twelve ($p \leq 0.05/16$ loci = 0.0031) in 7,775 individuals. Gene prioritization efforts including pleiotropy with gene expression in whole blood, gene-based association testing and functional annotation of associated SNPs, resulted in a total of 37 credible genes with a putative role in IgG galactosylation. We then used recently developed transient expression system HEK293FS for IgG production to up- and downregulate expression of seven new candidate genes, including *KIF3C*, *NFKB1*, *MANBA*, *SLC38A10*, *TNFRSF13B*, *EEF1A1* and *HIVEP2*, followed by release of glycans from secreted IgG and quantification using ultra-performance liquid chromatography. The effect on IgG galactosylation was observed upon upregulation of three genes: *EEF1A1*, *MANBA*, and *TNFRSF13B*. These findings further indicate that IgG galactosylation is indeed regulated by a complex network of genes that were previously not known to be involved in IgG glycosylation. However, the mechanism behind their involvement in the IgG galactosylation pathway remains to be elucidated.

Epigenetics Posters - Thursday

PB2391. Differential newborn DNA methylation among individuals with complex congenital heart defects and pediatric lymphoma.

Authors:

M. Richard¹, W. Yang², P. Sok¹, M. Li³, S. Carmichael⁴, J. von Behren⁵, P. Reynolds⁵, P. Fisher², T. Collins², C. Hobbs⁶, P. Lupo¹, G. Shaw⁷; ¹Baylor Coll. of Med., Houston, TX, ²Stanford Univ., Palo Alto, CA, ³Indiana Univ. at Bloomington, Bloomington, IN, ⁴Stanford Univ. Sch. of Med., Stanford, CA, ⁵Univ. of California San Francisco, San Francisco, CA, ⁶Rady Children's Inst. for Genomic Med., San Diego, CA, ⁷Stanford Univ., P, CA

Abstract Body:

Background: Children with birth defects are at increased risk for pediatric cancer. Data from the California Birth Defects Monitoring Program and the California Cancer Registry indicated individuals with complex congenital heart defects (CHDs) were at 8-fold greater risk for pediatric lymphoma. We sought to characterize molecular changes related to complex CHD and subsequent pediatric lymphoma development (“CHD-lymphoma”) through newborn DNA methylation patterns. **Methods:** From > 3 million live births (1988-2004) in the linkage cohort, we obtained newborn dried bloodspots from eight children with CHD-lymphoma through the California Newborn Screening Program. We performed case-control epigenome-wide association analyses (EWAS) to identify differential methylation (CpGs) using two comparison groups: 1) 45 boys without birth defects or cancer, and 2) 46 individuals with complex CHD. Analyses of the Illumina EPIC array were adjusted for technical variation and CpGs at $P < 6.9 \times 10^{-8}$ were considered statistically significant. **Results:** After correction for multiple testing, individuals with CHD-lymphoma had differential newborn methylation at 177 CpG sites compared to unaffected individuals and two CpG sites compared to individuals with complex CHD. *PPF1A1* cg25574765 was hypomethylated in CHD-lymphoma cases (mean beta=0.04) relative to both unaffected individuals (mean beta=0.93, $P=1.5 \times 10^{-12}$) and individuals with complex CHD (mean beta=0.95, $P=3.8 \times 10^{-8}$). *PPF1A1* encodes a ubiquitously expressed liprin protein that is conserved in evolution and resides in one of the most commonly amplified regions in many cancers (11q13). Further, cg25574765 is a proposed differentially methylated marker of pre-eclampsia, a maternal risk factor for CHDs that has not been fully evaluated for lymphoma risk in offspring. Additionally, *EXD3* cg13408086 was hypomethylated in CHD-lymphoma cases (mean beta=0.04) relative to unaffected individuals (mean beta=0.95, $P=1.0 \times 10^{-17}$) and individuals with complex CHD (mean beta=0.95, $P=3.8 \times 10^{-7}$). *EXD3* encodes an exonuclease protein associated with atrial septal defects. Annotation of significant CpG sites in the EWAS Catalog revealed methylation patterns previously associated with prenatal smoking, autoimmune traits, pollution, and markers of maternal health present in newborn DNA. **Conclusions:** Using multiple approaches, we identified associations between molecular changes present in the genome at birth and risk of pediatric lymphoma among those with CHD. Our findings also highlight novel environmental and perinatal exposures that may underlie methylation changes in CHD predisposing to lymphoma.

Epigenetics Posters - Wednesday

PB2392. Differentially methylated probes and regions associated with the Healthy Eating Index.

Authors:

K. Yuan, R. M. Lucia, W-L. Huang, D. Forman, D. Goodman, A. Ziogas, H. L. Park, T. M. Norden-Krichmar; Univ. of California, Irvine, Irvine, CA

Abstract Body:

Background: Diet quality has been linked with various health outcomes but may be difficult to measure in epidemiologic studies. DNA methylation markers from blood samples have been utilized to reveal associations with a variety of environmental exposures. However, the association between DNA methylation markers and measures of diet quality remains unclear. This epigenome-wide association study (EWAS) was conducted to identify Differentially Methylated Probes (DMPs) and Differentially Methylated Regions (DMRs) associated with diet quality scores according to the Healthy Eating Index (HEI). **Methods:** 199 participants from the Markers for Environmental Exposure (MEE) Study, comprised of postmenopausal women residing in Southern California, were included in the study. HEI scores were calculated from each participant's responses to two diet quality survey instruments: the Automated Self-Administered 24-hour Dietary Assessment Tool (ASA24) and the Diet History Questionnaire II (DHQII). Peripheral blood DNA methylation was measured using the Illumina Infinium MethylationEPIC Chip. The DMPs and DMRs associated with the HEI scores were identified using linear models with a false discovery rate p-value of 0.05, using a resampling method to improve the stability of results. **Results:** HEI scores calculated using the ASA24 versus DHQII were moderately correlated (Pearson correlation 0.557). We identified 6375 and 4257 significant DMPs associated with the ASA24 and DHQII HEI scores, respectively, 486 of which were associated with both scores. Sixteen DMRs were associated with the ASA24 HEI scores, and 17 with the DHQII HEI scores; none were associated with both scores. Among the top DMRs associated with ASA24 HEI scores was the gene *EBF3*, which is found in pathways related to differentiation of white and brown adipocytes. The top DMRs associated with DHQII HEI scores were the genes *FOXP2* (a transcription factor), and *PREX1* (a gene which has been associated with gastric cancer). **Conclusions:** Our results suggest that dietary intake is a complex exposure, potentially affecting many methylation sites in the epigenome. We also found that there was little overlap between results that used different dietary recall instruments to calculate the HEI scores, reflecting differences in the information collected by the two approaches. These preliminary results suggest that DNA methylation could be used as a biomarker for measuring diet quality, but further research is needed to identify markers that are robust to different methodological approaches.

Epigenetics Posters - Thursday

PB2393. Digitally untangling the 3D genome

Authors:

G. Fudenberg; Univ. of Southern California, Los Angeles, CA

Abstract Body:

Maps of 3D genome folding reveal that mammalian interphase genomes are segmented into a series of self-associating regions. Genetic variants disrupting this folding have been associated with disease phenotypes. Experimental work points to two major factors underlying this mode of interphase genome folding: cohesin, a Structural Maintenance of Chromosomes (SMC) complex that actively compacts chromosomes, and CTCF, an 11 zinc finger protein that binds specific DNA sequence motifs and delineates domain boundaries. Still, despite this knowledge, it remains challenging to predict how a given DNA variant may affect genome folding. Depending on the locus, folding can either be sensitive to single nucleotide changes, or resilient even to large-scale structural variants. The finite throughput for experimental perturbations and large potential consequences of disrupted folding call for computational solutions to both quickly and accurately predict the impacts of DNA variants on genome folding. Convolutional neural networks (CNNs) are powerful computational methods for modeling genomic data. Starting from DNA sequences as inputs, CNNs now enable state-of-the-art predictions for transcription factor binding, DNA accessibility and transcription. Moreover, rather than requiring exhaustive engineering of all predictive features by hand, CNNs can instead learn the relevant features from the data. New analysis methods make DNA sequence features learned by these models increasingly interpretable. Here I present the latest sequence-to-folding models of 3D genome organization. I will illustrate the strength of these approaches by drawing applications that include: (i) rapid DNA sequence mutagenesis, (ii) predicting the impacts of disease and engineered genetic variants, (iii) predicting genome folding across species. Finally, I will discuss an emerging computational toolkit for analyzing large sparse genome folding datasets that were crucial to the development of sequence-to-folding models for the 3D genome.

Epigenetics Posters - Wednesday

PB2394. Disparities in epigenetic modifications in *NRF1* and *FTO* genes in children.

Authors:

P. Patel, V. Selvaraju, X. Wang, R. Jeganathan, G. Thangiah; Auburn Univ., Auburn, AL

Abstract Body:

Despite vigorous efforts by health experts, childhood obesity has affected the health of millions of children worldwide. Obesity prevalence rates differ significantly by race and ethnicity, with African Americans being 50% more likely to be obese than non-Hispanic whites. Identifying factors influencing different populations is critical to successfully reducing obesity-related health disparities. A remarkable breakthrough has been seen in the field of epigenetics in the past few years. Epigenetic alterations, including DNA methylation, histone modification, and mRNA modifications are increasingly being implicated in the development of obesity. The expressions of various obesity-related genes are affected due to such epigenetic modifications. Here we used real-time quantitative PCR-based multiplex MethyLight technology to assess the DNA methylation percentage of the genes *NRF1* and *FTO* from the saliva of children aged 6-10 years. *ALU* was used as a reference gene, and the Percent Methylation Rate (PMR) was calculated for each sample. Results showed that European American children had a significant increase in PMR of *NRF1* and *FTO* in overweight/obese participants compared to normal weight. In contrast, there was no significant increase in overweight/obese children's PMR of *NRF1* and *FTO* than the normal weight in African American children. After adjusting for maternal education and annual family income by regression analysis, the PMR of *NRF1* and *FTO* was substantially linked with BMI z-score in European Americans, but not in African American children. The results also indicate that in the African American population, not only the obesity status but also other factors such as different health disparities could be playing a role in causing increased DNA methylation as we did not see a significant association between obesity markers and DNA methylation of *NRF1* and *FTO* gene in them. These findings contribute to a race-specific link between *NRF1* and *FTO* gene methylation and childhood obesity.

Epigenetics Posters - Thursday

PB2395. Dissecting mechanisms underlying expression divergence of human duplicated genes

Authors:

M. Dennis¹, C. Shew¹, G. Kaya²; ¹Univ. of California, Davis, Davis, CA, ²Univ. of California, Davis, Davis, CA

Abstract Body:

A growing body of evidence links genes within human-specific segmental duplications (HSDs) to traits and diseases unique to our species. Despite being nearly identical by sequence (>98.5%), HSD genes display paralog-specific expression patterns across human cell lines and primary tissues, suggesting that they have functionally diverged. We identified candidate *cis*-regulatory elements (cCREs) within HSDs using ChIP-seq data from human lymphoblastoid cell lines (LCLs) and adult/fetal cortex and showed that paralogous enhancers can be differentially active *in vitro*. We hypothesize that changes to the expression of recently duplicated and rearranged genes can be modeled with a more complete picture of their regulatory environment. To assay duplicated cCRE activity in high throughput, we performed a massively parallel reporter assay (MPRA) to quantify *cis*-regulatory activity of 5,701 HSD sequences in human LCLs and SH-SY5Y neuroblastoma cells to identify differentially active sequences. From 5,064 pairwise comparisons of human cCREs against chimpanzee orthologs in LCLs, we identified 371 differentially active sequences, of which 218 (59%) belonged to human derived paralogs and 211 (57%) showed higher human activity. Of the five top-scoring sequences, three reside within genes with known roles in human neurodevelopment: *ARHGAP11A*, *HYDIN*, *SRGAP2*. Additionally, 304 of 2,952 human derived sequences were differentially active with respect to the ancestral human locus. Preliminary analysis of Hi-C in human and chimpanzee LCLs, using a modified approach that quantifies multimapping Hi-C data across duplicated regions, identified human-specific contacts associated with the insertion of *DUSP22B* on chromosome 16, as well as paralog-specific chromatin loops at the *NCF1* promoters. Both genes mediate autoimmune response. Together, these assays will identify differences in the connectivity and activity of a comprehensive set of regulatory elements associated with HSDs and other structural variants in humans. This work represents the first comprehensive analysis of *cis* regulatory mechanisms contributing to divergent duplicated gene expression. Results will also offer potential insights into neurological and immunological traits unique to our species.

Epigenetics Posters - Wednesday

PB2396*. Dissecting the molecular machinery and sequence of histone to protamine transition during spermiogenesis.

Authors:

M. Rabbani, L. Moritz, S. Hammoud; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Spermiogenesis is the process by which post meiotic haploid germ cells undergo global chromatin remodeling to transition from a histone-based chromatin packaging to an alternative packaging system, driven primarily by sperm-specific nuclear proteins called protamines. Successful completion of this step is critical to producing mature sperm and maintaining male fertility. Despite being one of the most remarkable chromatin remodeling events known to occur during cellular differentiation, little is known about the chromatin remodeling events or the molecular machinery involved in this process. To explore this question, we tagged the endogenous protamine 1 and 2 loci with V5-tag and performed immunoprecipitation followed by mass spectrometry (IP-MS) to identify interactors. The ISWI chromatin remodeling family members, notably, SNF2H and BAZ1A, were among the most highly enriched candidates of protamine 1 interactors. The ISWI complex is a canonical nucleosome remodeling complex, that uses ATP hydrolysis to regulate nucleosome positioning and DNA accessibility. The identification of the ISWI complex was unexpected since it was presumed that histone removal and protamine deposition are temporally uncoupled events. Specifically, it was thought that histones are first replaced by transition proteins and subsequently by protamines. Given our MS data, we carefully followed the dynamics of these proteins during spermiogenesis and found that indeed protamine 1 deposition precedes transition protein and protamine 2 incorporations. These findings suggest that histone to protamine exchange may not occur as previously outlined, but rather protamine 1 may directly replace histones, and the reliance on classic histone machinery is therefore plausible. To confirm the plausibility of these interactions, we co-stained the testis cross sections with anti-V5 and SNF2H and show that the two proteins colocalize in vivo. In vitro reciprocal IPs with SNF2H confirm its specific interaction with protamine 1 but not transition proteins. Consistent with these molecular findings, germ cell knockouts of the ISWI complex display defects in histone to protamine exchange. Based on our observations, we propose a model where the ISWI complex may directly interact with protamine1 and facilitate the removal of histones and allow the deposition of protamine 1.

Epigenetics Posters - Thursday

PB2397. Distinct signatures in malignant PEComa and leiomyosarcoma identified by integrative RNA-seq and H3K27ac ChIP-seq analysis

Authors:

K. Giannikou^{1,2,3}, A-A. Azim², E. Adib², I-M. Schaefer⁴, N. Kesten³, L. Taing³, Z. Zhu², J. L. Hornick⁴, M. S. Hirsch⁴, H. W. Long³, A. J. Wagner⁵, M. L. Hemming⁵, D. J. Kwiatkowski²; ¹Div. of Hematology and Oncology, Cancer and Blood Disease Inst., Children's Hosp. Los Angeles, Los Angeles, CA, ²Cancer Genetics Lab., Div. of Pulmonary and Critical Care Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, ³Ctr. for Functional Cancer Epigenetics, Dana-Farber Cancer Inst., Harvard Med. Sch., Boston, MA, ⁴Dept. of Pathology, Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, ⁵Ctr. for Sarcoma and Bone Oncology, Dept. of Med. Oncology, Dana-Farber Cancer Inst., Boston, MA

Abstract Body:

Background: Malignant perivascular epithelioid cell tumor (PEComa) and leiomyosarcoma (LMS) are two sarcoma tumor subtypes with overlapping morphologic and immunophenotypic features which can make their diagnostic distinction challenging. **Aim:** To characterize the transcriptional and epigenetic landscape of PEComa and LMS to identify distinguishing features. **Methods:** We performed whole transcriptome RNA-sequencing on 19 PEComas and compared their gene expression profile to 259 sarcomas from The Cancer Genome Atlas (TCGA) including 104 LMS. ChIP-sequencing for H3K27ac, a histone modification associated with activation of nearby genes/open chromatin, was conducted on 9 malignant PEComas and 12 LMS and were compared with publicly available data from 5 other sarcoma subtypes (GIST; chordoma; osteosarcoma; undifferentiated pleomorphic sarcoma; rhabdomyosarcoma). **Results:** Genome-wide epigenetic and transcriptional analyses revealed overlapping patterns between PEComa and LMS, which were distinct from other sarcomas. However, we also identified a set of highly expressed and epigenetically distinct transcripts which may represent diagnostic biomarkers: e.g., *DAPL1*, *MLANA*, *SULT1C2*, *GPR143*, and *CHI3L1* for PEComa; and *MYOCD*, *WDFC2*, *DES*, *MYH11*, and *CNN1* for LMS; each of which showed >17x fold higher expression for each tumor entity by DESeq2 (FDR<0.0001). Gene Set Enrichment Analyses (GSEA) demonstrated enrichment in the KEGG Lysosome pathway for PEComa (FDR=0.11), whereas myogenesis and smooth muscle contraction pathways were enriched in LMS (FDR=0.09). Integrative transcriptomic and epigenetic analyses revealed a unique set of master core transcription factors for each tumor type including among others MYOCD for LMS; MITF for PEComa, which require further functional investigation. Twelve selected genes including new as well as known and standard diagnostic markers (e.g., *DAPL1*, *MLANA*, *GPR143*, *PNL2*, *CHI3L1*, *DES*, *MYH11*, *ER*, *CD68*, *PU.1*, *pS6* and *CNN1*) were validated by immunohistochemistry (IHC) in multiple sections from PEComa and LMS (n=26 tumors/marker). The combination of three melanocytic markers (HMB45, *MLANA*, *PNL2*) and *pS6* can distinguish LMS from PEComas (p<0.0001). IHC for *CD68* and *PU.1* macrophage markers did not show any difference regarding the degree of immune infiltration in PEComa vs. LMS. **Conclusions:** PEComa and LMS present distinct expression and epigenetic features, however there are cases which overlap each tumor type. Given the success of mTOR inhibitors in the treatment of *TSC1/TSC2* mutant PEComa, genetic analysis of LMS for *TSC1/TSC2* mutations should be considered for all LMS patients to facilitate diagnosis.

Epigenetics Posters - Wednesday

PB2398. Divergent age-related methylation patterns in long and short-lived mammals

Authors:

A. Haghani; Altos Labs, San Diego, CA

Abstract Body:

Age-related changes to cytosine methylation have been extensively characterized across the mammalian family. Some cytosines that are conserved across mammals exhibit age-related methylation changes that are so consistent that they were used successfully to develop cross-species age predictors. In a similar vein, the methylation state of some conserved cytosines correlates extremely well with species lifespan. Surprisingly, little to no commonality is found between these two sets of cytosines even though the relationship between aging and lifespan is, by most measures, linked. We ventured to address this conundrum by first identifying age-related cytosines whose methylation change in opposite directions between short and long-lived species. We hypothesized that age-related CpGs that are also associated with species lifespan would tap into biological processes that simultaneously impact aging and lifespan. To this end, we analyzed age-related cytosine methylation patterns in 82 mammalian species. For each CpG, we correlated the intra-species age correlation with maximum lifespan across mammalian species. We refer to this correlation of correlations as “*Lifespan Uber Correlation (LUC)*”. This approach is unique in incorporating age and species lifespan in a single analysis. We identified 629 CpGs with opposing methylation aging patterns in long and short-lived species. Many of these are found to be near *BCL11B*, *NPTN*, and *HOXC4* loci. Methylation and transcription analyses of *BCL11B* knockout mice indicate that this gene partially regulates the methylation state of LUC CpGs. We developed DNAm age estimators (epigenetic clocks) based on LUC CpGs. These LUC clocks exhibited expected behavior in benchmark aging interventions such as caloric restriction, growth hormone receptor knockout and high-fat diet. Furthermore, we found that *BCL11B* knockout mice increased the epigenetic age of their striatum. Overall, we present a bioinformatics approach that identified CpGs and their associated genes implicated in both aging and lifespan. These cytosines lend themselves to developing highly accurate epigenetic clocks that are sensitive to perturbations that impact both age and lifespan.

Epigenetics Posters - Thursday

PB2399. DNA methylation changes among Ethiopian women diagnosed for cervical cancer.

Authors:

B. Kumbi Jufara¹, Y. Gebrehiot², D. Seifu³, D. Beyene², A. Lorincz⁴; ¹Ethiopian Police Univ., Addis Ababa, Ethiopia, ²Addis Ababa Univ., Addis Ababa, Ethiopia, ³Univ. of Global Hlth.Equity, Kigali, Rwanda, ⁴Queen Mary Univ. of London, London, United Kingdom

Abstract Body:

Infection by high risk human papillomavirus (hrHPV) is the major risk factor for cervical cancer with almost all cases being infected. HPV infection is a very common sexually transmitted infection that majority of sexually active women acquire. Most of the infected women however clear the infection spontaneously in short time while it persists and causes cervical cancer in only small portion of the infected women. Various factors are known to determine the outcome of the infection but with scarce information on the mechanisms. DNA methylation changes in both human genes and HPV genes are among the biological events associated with cervical cancer progression. This study was aimed at determining prevalence of hrHPV, socio-demographic risk factors for cervical cancer and epigenetic changes associated with cervical cancer and evaluate their potential as diagnostic markers. The study was conducted as an observational case-control study. The human *EPB41L3* gene promoter region and HPV L1 and L2 regions were PCR amplified from bisulphite converted DNA. The PCR amplicons were then pyrosequenced and proportion of converted cytosine is measured and means of the targeted CpG sites methylation were recorded. The DNA methylation assays were evaluated and compared for their performance using Receiver Operating Characteristics (ROC) curve analysis. HPV16 was the most prevalent virus constituting 84% of all hrHPV positive cases and 33.3% of hrHPV positive controls. HPV45, HPV18 and HPV31 were detected in 17.7%, 5.2% and 3.8% respectively of the hrHPV positive cases. Level of methylation in both human and hrHPV DNA was found to be higher in higher grade lesions than in low grade lesions (CIN1) and normal cervical cells. Methylation assays, both *EPB41L3* promoter methylation and S5 score discriminated normal and CIN1 from CIN3 or worse lesions with sensitivity and specificity of greater than 95%. In conclusion, higher parity and earlier age at first sexual intercourse are among the factors that put women at higher risk of cervical cancer in addition to hrHPV infection. HPV16 is the most prevalent (69.8%) hrHPV type followed by HPV45 (14.6%) in Ethiopian women with cervical lesions. Methylation levels of the human *EPB41L3* promoter region and HPV L1 and L2 regions are potential biomarker to improve precision of diagnosing the cancer and targeting for therapy. EBP41L3 methylation alone discriminated normal and CIN1 cells from CIN3 or worse lesions with 95% sensitivity and 96% specificity while S5 detected with 96% sensitivity and 95% specificity.

Epigenetics Posters - Wednesday

PB2400. DNA methylation changes associate with measured glomerular filtration rate in an American Indian cohort with type 2 diabetes.

Authors:

S. Day¹, R. Hanson¹, K. Susztak², R. Nelson¹, C. Bogardus¹, L. Baier¹; ¹NIDDK, Phoenix, AZ, ²Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Diabetes is a leading cause of progressive kidney disease characterized by a decline in the glomerular filtration rate (GFR), but the molecular mechanisms underlying this decline are not completely understood. DNA methylation is an epigenetic mechanism that can regulate gene expression. The goal of this project was to identify DNA methylation changes associated with declining GFR in Southwest American Indians (SWAI). In 108 SWAI with type 2 diabetes (T2D), blood for DNA methylation and urine for GFR analyses were collected at 3 exams that were spaced 5 years apart (10 years total). Methylation was measured using the Infinium HumanMethylation450 array and GFR was directly measured by the urinary clearance of iothalamate. Normalization and analysis of variance (ANOVA) was performed using Partek Genomics Suite. To maximize the power of this limited sample size, association analysis was assessed between methylation and concomitantly measured GFR using all time points. One of our top methylation probes associated with GFR was cg19693031, in the 3'untranslated region of Thioredoxin Interacting Protein (*TXNIP*). This probe was also strongly associated with hyperglycemia as measured by HbA1c ($r = -0.55$; $P=1 \times 10^{-17}$), and the effect on GFR was explained by adjustment for HbA1c, a major risk factor for kidney disease progression. A functional effect of the methylation site captured by cg19693031 on *TXNIP* expression has been previously established, as well as its association with HbA1c and risk of T2D in other ethnic groups. The present results show that this methylation finding is also observed in SWAI. In analysis of association with GFR adjusted for HbA1c, no single probe achieved epigenome-wide significance (3.0×10^{-8}). Therefore, we conducted a pathway analysis using DAVID on a set of 1,540 genes (2,263 probes) that contained methylation sites associated with GFR at nominal significance ($P \leq 0.01$). This analysis identified endocytosis to be the most enriched pathway; loss of endocytosis function is associated with chronic kidney disease. The top methylation probe associated with GFR was cg12590521 in Calcium/Calmodulin Dependent Protein Kinase Kinase 2 (*CAMKK2*; $r = 0.32$; $P=3 \times 10^{-6}$). This site has not been previously implicated in diabetic kidney disease, but falls within a predicted enhancer region. Our plans for this ongoing project include functional assessment of this DNA methylation site (cg12590521) in *CAMKK2*.

Epigenetics Posters - Thursday

PB2401. DNA methylation epi-signature and biological age in attention deficit hyperactivity disorder patients

Authors:

G. Carvalho¹, T. Costa², A. Nascimento¹, B. Wolff¹, J. Damasceno¹, L. Vieira¹, V. Almeida¹, Y. Gasparini², C. Mello³, M. Muszkat³, L. Kulikowski¹; ¹Faculdade de Med. da Univ.e de São Paulo, São Paulo, Brazil, ²Univ.e de Sao Paulo, Sao Paulo, Brazil, ³Departamento de Psicobiologia da Univ.e Federal de São Paulo, São Paulo, Brazil

Abstract Body:

Attention Deficit/Hyperactivity Disorder (ADHD) is a common behavioral syndrome that begins in childhood and affects 3.4 % of children worldwide. Due to its etiological complexity, there are no consistent biomarkers for ADHD, however the high heritability that the disorder presents indicates a influence of genetic/epigenetic factors. The main epigenetic mechanisms is DNA methylation, a process with an important role in gene expression. Methylation status can be used as a biomarker for clinical conditions, such as cancer and neuropsychiatric diseases with specific epi-signatures, and it is used as a marker of biological senescence by several studies to determine biological age (DNAmAge). Thus, our study sought to identify epi-signatures biomarkers in 29 children clinically diagnosed with ADHD. Our results suggest that the biological response in patients with ADHD is not sufficient to determine an epi-signature. However, the results highlight the interaction of energy metabolism and oxidative stress pathways in ADHD patients detected by differential methylation patterns. We were able to identify a marginal association between the DNAmAge and ADHD. Thus, we present new methylation biomarkers in ADHD patients and propose that further multiethnic studies with larger cohorts should be done to demonstrate a definitive association between ADHD and these methylation biomarkers.

Epigenetics Posters - Wednesday

PB2402. DNA Methylation Episignature of Valproate Embryopathy.

Authors:

S. Haghshenas¹, J. Reilly¹, M. A. Levy¹, **R. Relator**¹, H. McConkey¹, J. Kerkhof¹, P. Edery², G. Lesca², C. Coubes³, M. Willems³, M. Barat-Houari⁴, Q. Sabbagh³, D. Genevieve^{3,5}, A. Putoux², B. Sadikovic^{1,6}; ¹London Hlth.Sci. Ctr., London, ON, Canada, ²Hospices Civils de Lyon, Service de Génétique, Groupement Hosp.ier Est, Bron, France and Université Claude Bernard Lyon 1, Lyon, France, ³Ctr. for Rare Disease Malformative Syndromes, Genetic Clinic Unit, CHU Montpellier, Montpellier, France, ⁴Laboratoire de Génétique des Maladies Rares et Autoinflammatoires, CHU Montpellier, Montpellier, France, ⁵Montpellier Univ., Inserm U1183, Montpellier, France, ⁶Western Univ., London, ON, Canada

Abstract Body:

Valproate is an effective antiepileptic drug and a mood-stabilizer with generally mild side effects. However, it is also recognized as a teratogen, since its consumption during pregnancy has been associated with increased susceptibility of the offspring to congenital anomalies and neurodevelopmental defects, referred to as valproate embryopathy or fetal valproate syndrome (FVS). Diagnosis of FVS can be difficult as there are currently no reliable molecular biomarkers for FVS. Peripheral blood DNA methylation patterns, known as episignatures have been established as stable and accurate diagnostic biomarkers for an increasing number of genetic neurodevelopmental disorders. However, it is currently not known if similar episignatures exist in patients affected by neurodevelopmental disorders that are caused by non-genetic factors, and in particular teratogenic exposures. By assessing a cohort of patients affected by FVS we demonstrate the existence of a highly accurate, sensitive and specific episignature for FVS. We developed a binary classification mode enabling an accurate molecular diagnosis of patients with FVS. We also describe the genome-wide changes in DNA methylation in FVS and compare the changes relative to above 120 other genetic neurodevelopmental syndromes with known DNA methylation episignatures. This expands the rapidly growing list of disorders with a known diagnostic episignatures and demonstrates diagnostic utility of EpiSign analysis beyond genetic syndromes.

Epigenetics Posters - Thursday

PB2403. DNA methylation in twins discordant for van der Woude syndrome

Authors:

A. Petrin¹, X. J. Xie¹, E. Zeng¹, D. Moretti-Ferreira², M. Marazita³, J. Murray⁴, L. Moreno¹; ¹Univ. of Iowa Coll. of Dentistry and Dental Clinics, Iowa City, IA, ²São Paulo State Univ. - UNESP, Botucatu, Brazil, ³Univ Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA, ⁴Univ. of Iowa Carver Coll. of Med., Iowa City, IA

Abstract Body:

Van der Woude Syndrome (VWS) is an autosomal dominant disorder responsible for 2% of all syndromic orofacial clefts (OFCs). Mutations in *IRF6* and *GRHL3* account for 70% and 5% of VWS cases, respectively. VWS presents with lip pits and/or cleft lip, cleft lip/palate, or cleft palate, with markedly variable expression. The phenotypic variation observed in monozygotic twins suggests epigenetic influence on the phenotype. We report the genome-wide DNA methylation (DNAm) profiling of one pair of monozygotic twins discordant for VWS. The affected twin has a missense mutation in *IRF6* that is absent in the unaffected co-twin. Our goal is to explore epigenetic contributions to this phenotypic discordance such as DNAm differences, which can alter the expression of additional craniofacial genes, especially genes of the *IRF6* pathway. Genome-wide DNAm profiles were generated using Illumina's Infinium Methylation EPIC BeadChip. We used ChAMP/RnBeads R packages to obtain beta values, which were used for calculation of the absolute DNAm difference. Moreover, we used the beta values of the 13 unaffected individuals as additional control. Cell type composition was estimated with EpiDISH followed by a logistic regression to correct for cell type heterogeneity. Finally, we performed gene ontology (GO) and enrichment analysis using GREAT. We found 19,196 differentially methylated positions (DMPs). 78 genes that contained DMPs when the Affected > Unaffected (affected more methylated) also contained binding sites for *IRF6*. GO analysis showed 170 genes enriched for the DMPs when Affected > Unaffected, with *TP63* ($p = 7.82E^{-12}$) among the top hits. Unaffected > Affected showed 91 genes enriched, with *TNF* ($p = 8.69E^{-09}$) and *PAX7* ($p = 2.82E^{-03}$) among the top hits. Several relevant human phenotypes associated with clefting had their causal genes enriched in the dataset including epidermal, ectodermal, and epithelial disorders, and craniofacial anomalies like craniofacial dysostosis and broad philtrum. Expression of *IRF6* requires normal function of *TP63*. Our data shows higher DNAm levels in *TP63* and many *IRF6* targets in the affected twin in addition to the causal *IRF6* mutation. The change in DNAm can add to the disruption of the biological regulatory loop that controls epithelial proliferation and differentiation, therefore contributing to phenotypic variability, often seen among VWS cases. This exploratory approach paves the way to further studies on epigenetic contributions to the phenotypic variability. Epigenetic changes can act as modifiers to causal mutations, contributing to phenotypic variability and penetrance of craniofacial Mendelian disorders like VWS.

Epigenetics Posters - Wednesday

PB2404. DNA methylation profile in multiple system atrophy and progressive supranuclear palsy

Authors:

P. Reho¹, S. Solaiman¹, T. G. Beach², G. E. Serrano², K. L. Newell³, R. Chia¹, H. Houlden⁴, J. C. Troncoso⁵, O. Pletnikova⁶, H. P. Easwaran⁷, S. W. Scholz¹; ¹NIH, Bethesda, MD, ²Banner Sun Hlth.Res. Inst., Sun City, AZ, ³Indiana Univ. Sch. of Med., Indianapolis, IN, ⁴Univ. Coll. London, London, United Kingdom, ⁵Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ⁶Jacobs Sch. of Med. and BioMed. Sci., Buffalo, NY, ⁷The Johns Hopkins Sch. of Med., The Sidney Kimmel Comprehensive Cancer Center, Baltimore, MD

Abstract Body:

Multiple system atrophy (MSA) and progressive supranuclear palsy (PSP) belong to a heterogeneous group of neurodegenerative disorders called atypical parkinsonism syndromes. Despite the discoveries of genetic risk factors, the causes and epigenetic changes underlying these understudied conditions are poorly understood. We performed an epigenome-wide association study on cerebellar DNA samples from 65 MSA and 95 PSP patients, and 193 neurologically healthy controls. The mean age of each group of subjects was 68.8 years, 73.3 years, and 72.8 years respectively. All samples were genotyped using the Illumina Infinium MethylationEPIC and Infinium Neuro Consortium (v1-1) arrays. After stringent quality control checks according to the gold-standard of the field, we performed a DNA methylation profiling focusing on, genome-wide significance, enrichment pathways analysis, differentially methylated regions analysis, and known risk genes for MSA and PSP (*GBA*, *SNCA*, *COQ2* and *MAPT*). We detected 175 differentially methylated probes (adjusted P value < 0.05) in MSA, of which 152 (87 %) were hypermethylated and 23 (13 %) were hypomethylated in cases compared to controls. Among PSP patients, we identified 108 differentially methylated probes, of which 49 (45 %) were hypermethylated and 59 (55 %) were hypomethylated. Of note, MSA and PSP patients shared five differentially methylated probes in *EXOSC4*, *RIMBP2*, *FIBCD1* and *C10orf129* genes. We found no evidence for differential methylation signatures at known risk genes. Here we will describe differentially methylated regions and disease-associated pathways. Our preliminary data highlight an important epigenetic modulation in MSA and PSP that could potentially lead to the identification of new molecular targets for therapeutic development.

Epigenetics Posters - Thursday

PB2405. DNA methylation signatures in Arab and Middle Eastern families with neurodevelopmental disorders provide insight to disease etiology.

Authors:

A. Nour^{1,2}, R. Mathew¹, I. Diboun¹, M. Hashmi¹, M. Ghorbani¹, N. Yousri³, K. Fakhro^{4,5,2}, N. Hajj², Y. Mokrab^{1,5}; ¹Sidra Med., Doha, Qatar, ²Hamad Bin Khalifa Univ., Doha, Qatar, ³Weill Cornell Med. - Qatar, Doha, Qatar, ⁴Sidra Med., Doha, Qatar, Qatar, ⁵Weill Cornell Med.-Qatar, Doha, Qatar

Abstract Body:

Epigenetic signatures (EpiSigns) have been identified in more than 60 neurodevelopmental disorders (NDDs) and helped reclassify variants of uncertain significance, aiding in the diagnosis of many idiopathic disorders. Most EpiSigns were studied in European and American subjects, while little is known about them in Arab and Middle Eastern ancestries. Here, we used the Epic Illumina array (850,000 CpG) to explore DNA methylation levels amongst 10 families, recruited in Qatar, with disease patients with NDDs (mostly intellectual disability and epilepsy). Each family consisted of unaffected parents and a sibling and an undiagnosed proband. Next, we investigated whether published NDD-associated epigenetic signatures explain our emerging data. For this, we trained an Orthogonal Partial Least Squares classifier on epigenetic data from 89 children with intellectual disability and 18 controls from a previously published study (PMID: 26003415). The classifier detected one significant component explaining 60% variation in the phenotype due to risk of NDD and was able to discriminate one of our probands from its control parents. CpGs assigned very important (VIP) scores by the classifier corresponded to three sites: (1) *chr7:117171038* from the *Cystic Fibrosis Transmembrane Conductance Regulator* gene (*CFTR*) involved in the regulation of neuronal excitability, (2) *chr3:93780377* in the 5' untranslated region (5'UTR) of *Dihydrofolate Reductase* gene (*DHFR1L1*), an enzyme that produces tetrahydrofolate which regulates DNA methylation and (3) *chr17:73056681* within *Potassium Channel Tetramerization Domain Containing 2* gene (*KCTD2*) which has a well-established neuromodulatory function (all chromosomal positions based on GrCh37). This study sheds light on the epigenetics of an understudied population and leads to valuable results that can help hasten the clinical diagnostic journey for NDD patients.

Epigenetics Posters - Wednesday

PB2406. Dysregulation of serum and exosomal miR-7-1-5p and miR-223-3p in Parkinson's disease

Authors:

L. Citterio¹, R. Mancuso¹, S. Agostini², M. Meloni¹, M. Clerici^{1,3}; ¹IRCCS Don Carlo Gnocchi Fndn., Milan, Italy, ²IRCCS Don Gnocchi Fndn., Milano, Italy, ³Univ. of Milan, Milan, Italy

Abstract Body:

The etiology of Parkinson's disease (PD) is poorly understood mainly because of the interaction between genetic and environmental factors that appears to play a fundamental role. In this context, it is essential to investigate the presence of possible biomarkers potentially useful both for prognostic and diagnostic purposes and to understand the mechanisms that regulate the etiology of PD. Over the past decade, several studies reported dysregulated microRNAs (miRNAs) expression in neurodegenerative disorders, included PD. Using ddPCR, we investigated the potential differences of expression of miR-7-1-5p, miR-499, miR-223-3p and miR-223-5p, miRNAs involved in α -Synuclein pathway and inflammation, extracted from serum of 94 subjects (45 PD patients and 49 age- and sex-matched HC). Furthermore, we evaluated the expression of these four miRNAs also from serum-isolated exosomes, extracellular vesicles able to carry small molecules, including miRNAs, from a body district to another even passing through the blood brain barrier. Serum miR-7-1-5p expression resulted to be significantly increased in PD (35.54 copies/ng; 5.82 - 89.61 copies/ng) compared to HC (0.00 copies/ng; 0.00 - 30.97 copies/ng; $p=0.0007$). Notably, it also showed a positive correlation with α -Synuclein in the PD group ($p=0.05$), while its expression in exosomes showed no statistically relevant differences. miR-223-3p showed as well a statistically increased expression in the serum of PD subjects (7843.02 copies/ng; 3761.14 - 26972.84 copies/ng) compared to HC (2593.12 copies/ng; 502.96 - 10989.89 copies/ng; $p=0.0006$) and, remarkably, it resulted to be more expressed also in exosomes from PD (4715.67 copies/ng; 2832.13 - 16175.03 copies/ng) compared to exosomes from HC (1698.53 copies/ng; 463.20 - 5547.85 copies/ng; $p=0.0002$). Furthermore, we also found that this miRNA positively correlates with the age of the enrolled population ($p=0.04$) and, in the PD group, with the equivalent daily dose of levodopa (LEDD, mg/mL), the precursor of dopamine neurotransmitters and to date one of the most effective treatments for this disorder, both in serum ($p=0.0008$) and in exosomes ($p=0.006$). On the other hand, miR-499-3p and miR-223-5p showed no statistical differences between the two groups and resulted to be undetectable in exosomes. Our results suggest that both miR-7-1-5p and miR-223-3p are able to distinguish between patients and healthy subjects and in particular, due to the correlation with α -Synuclein, that serum miR-7-1-5p has the potential to be a useful and no-invasive biomarker in Parkinson's disease.

Epigenetics Posters - Thursday

PB2407. Early Embryos Employ a Unique Mechanism of CENP-A Equalization during the First Cell Cycle

Authors:

C. Tower¹, G. Manske¹, K. Jorgensen¹, S. Chakraborty¹, B. Ma², M. Abo-Elenin³, S. Schon¹, B. Black⁴, K. Schindler³, X. Chen⁵, S. S. Hammoud¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Johns Hopkins Univ., Baltimore, MD, ³Rutgers Univ., Piscataway, NJ, ⁴Univ. of Pennsylvania, Philadelphia, PA, ⁵Johns Hopkins, Baltimore, MD

Abstract Body:

CENP-A is a centromere-specific histone H3 variant that is responsible for defining centromeric chromatin in most eukaryotic species. Previous investigations studying CENP-A inheritance in somatic cells have shown that existing CENP-A nucleosomes direct the deposition of new CENP-A nucleosomes during late telophase of mitosis. This mechanism results in quantitative maintenance of CENP-A levels and preserves the genomic location of centromeres across mitotic divisions. In addition to somatic inheritance, CENP-A nucleosomes are also known to be inherited intergenerationally. Work from our lab and others has established that CENP-A nucleosomes are inherited from both the maternal and paternal germline into the totipotent embryo. Additionally, our lab has recently uncovered that the paternal genome contributes lower levels of CENP-A to the early embryo, which poses a problem for faithful chromosome segregation during the first cell division. This suggests that early embryos might employ specific mechanisms during the first cell cycle to equalize CENP-A between homologous chromosomes. To investigate the mechanisms regulating CENP-A dynamics in early embryos, our lab tagged the endogenous Cenpa locus in mice and performed various in vitro fertilization experiments to monitor the distribution of CENP-A across different pronuclear stages in the zygote in response to various treatments. From these experiments, we observed that, unlike in somatic cells, deposition of CENP-A into the paternal pronucleus is linked to the start of S phase (PN3), which occurs after the protamine to histone exchange. Moreover, we found that a large portion of these newly deposited CENP-A nucleosomes are originally derived from maternally inherited, chromatin-bound CENP-A. Although this redistribution of CENP-A nucleosomes was linked to the start of S phase, it is not dependent on DNA replication. Additional experiments also demonstrated that zygotic transcription and translation were not necessary for the redistribution of maternal CENP-A into the paternal pronucleus to occur. Lastly, dissemination of maternal CENP-A to the paternal pronucleus does not appear to be predominantly regulated by CDK1 or PLK1, which were previously identified as mitotic CENP-A licensing factors in somatic cells. Future studies in our lab will focus on identifying the key factors regulating CENP-A dynamics in early embryos.

Epigenetics Posters - Wednesday

PB2408*. Epigenetic Age Acceleration is Suggestively Associated with Stroke Precursor Phenotypes

Authors:

N. Dueker¹, H. Zhao², C. Dong³, D. Cabral⁴, R. L. Sacco⁴, S. Blanton³, L. Wang⁴, T. Rundek³; ¹Univ. of Miami Miller Sch. of Med., Miami, FL, ²Yale Univ. Sch. of Publ. Hlth., New Haven, CT, ³Univ Miami, Miami, FL, ⁴Univ. of Miami, Miami, FL

Abstract Body:

Epigenetic clocks have been developed using DNA methylation levels at select CpG sites to predict chronological age (Hannum and Horvath clocks) and mortality (GrimAge and PhenoAge clocks). Epigenetic age acceleration (AgeAccel), which occurs when a person's epigenetic age is older than their chronological age (CA), has been associated with adverse health outcomes including cardiovascular diseases and obesity. Studies investigating the association between epigenetic aging and stroke precursor phenotypes, however, are limited. We previously assembled a collection of 61 extended families (n=799 people) from the Dominican Republic on whom stroke precursor phenotypes and CpG methylation data are available. We utilized this data to determine if epigenetic AgeAccel was associated with total carotid intima-media thickness (cIMT), carotid bifurcation thickness (BIF) and left ventricular mass (LVM). CpG methylation in blood was assayed using the Illumina Infinium Human MethylationEPIC BeadChip. Using these data, we estimated participants' epigenetic age using four epigenetic clocks; Horvath, Hannum, GrimAge and PhenoAge. AgeAccel measures for each of these epigenetic clocks were estimated by regressing resulting epigenetic ages onto CA. We then performed linear mixed model analyses regressing each AgeAccel measure onto LVM, cIMT and BIF. Adjustment was made for CA, sex, ancestry principal components and cell composition proportions as fixed effects and family was included as a random effect. Mean CA in our sample was 46.3 years and mean AgeAccel was -0.33 years for Horvath, 0.05 years for Hannum, 1.45 years for GrimAge and 7.38 years for PhenoAge epigenetic clocks. We found that Horvath AgeAccel is associated with increased BIF (beta=0.02, p=0.01). We also observed the following suggestive associations (0.05 < p < 0.1): 1). Horvath AgeAccel and increased cIMT and LVM, 2). Hannum AgeAccel and increased LVM, and 3). PhenoAgeAccel and increased BIF. GrimAgeAccel was not associated with any precursor phenotype (p > 0.1). A previous study in adults (mean CA = 57y) found epigenetic AgeAccel estimated using the Horvath and Hannum epigenetic clocks to be significantly associated with increased cIMT. Taken together, these results suggest that epigenetic AgeAccel, as measured by the Horvath epigenetic clock, is associated with stroke precursor phenotypes in older individuals. Additional analyses investigating the effects of CA and sex on the association between epigenetic AgeAccel and stroke precursor phenotypes are in progress.

Epigenetics Posters - Wednesday

PB2409. Epigenetic profiling of isolated blood cells reveals highly cell-type specific smoking signatures and links to disease risk

Authors:

X. Wang, M. Campbell, H-Y. Cho, S. Martos, D. Bell; Natl. Inst. of Environmental Hlth.Sci., Research Triangle Park, NC

Abstract Body:

Tobacco smoking leads to many human health problems and diseases. We hypothesize alteration of DNA methylation may underpin some of the pathogenesis of smoking-associated complex diseases. However, knowledge of how smoking-driven epigenetic effects in specific blood cell-types may link with disease risk is unknown. In this study, we isolated DNA from 6 major leukocyte subtypes (including CD14+ monocytes, CD15+ granulocytes, CD19+ B-cells, CD4+ T cells, CD8+ T cells, CD56+ natural killer cells) and whole blood from healthy adult smokers (n=64) and nonsmokers (n=71), and generated DNA methylation profiles with Illumina 450k and EPIC arrays. We identified 238 cell-type specific, smoking-associated CpGs (or smoking-CpGs) at genome-wide ($p < 1.2E-7$) from 6 major leukocytes. The numbers of smoking-CpGs varied from only 5 CpGs in CD8+ T cells to 111 CpGs in CD19+ B-cells, although B-cells and CD8+ T cells are both less than 10% of total leukocytes in normal whole blood samples. We found unique smoking effects in each cell-type, some of which were not apparent in whole blood because the whole blood effect is a sum of the collective change in individual cell-types. Using cell-type deconvolution analysis we observed that naïve B-cell proportions were reduced among smokers. We observed distinctive cell-type specific enrichment of canonical pathways, in particular, canonical pathways related to immune response and transcriptional regulation via transcription, were highly enriched among B-cell smoking-CpGs. Compared with published EWAS results, the cell-type specific smoking-CpGs were overrepresented among 8 EWAS CpG sets, most significantly, the “lung function” EWAS CpG set, which overlaps 35 smoking-CpGs (normalized enrichment score = 3.48; $p = 8.8E-88$). Integrating with large-scale public mQTLs and GWAS datasets, we found that 52 smoking-CpGs had mQTL SNPs that were either GWAS SNPs or in complete LD with GWAS SNPs, which associated with 4 categories of human phenotypes including lung function, disease risk, blood traits, and other traits. Tracing the cell type origin of these smoking-CpGs, we found lung function traits mainly linked to myeloid cells; but disease risk and blood traits were linked to both B-cell and myeloid cells. Examining functional connections between smoking-CpGs and gene expression, we found 8 smoking-CpGs associated in cis with mRNA levels at $p < 1.0E-5$, and 44 smoking-CpGs had 341 mQTL SNPs that were also cis-eQTL SNPs (10 kb to TSS). This comprehensive analysis of smoking-associated DNA methylation changes in human blood cell-types provides a path to better understanding of potential mechanisms in smoking-associated diseases.

Epigenetics Posters - Thursday

PB2410. Epigenetic regulation of Wnt signaling pathway associated with age-related mobility loss.

Authors:

E. Quillen, B. Frye, J. Negrey, T. Register, C. Shively; Wake Forest Sch. of Med., Winston Salem, NC

Abstract Body:

Non-human primate models (NHPs), including vervets (*C. pygerythrus*), show age-related decline in musculoskeletal mass and physical function similar to what is seen in humans with a decline starting around age 20 (60 human-equivalent years) and a broad distribution of gait speed within each age group. To identify epigenetic variants associated with lean muscle mass or gait speed in a cohort of 30 female vervets aged 8-28 years (~25-90 years in humans), we investigated methylation levels at 107,490 genome-wide loci in vastus lateralis biopsies collected during routine health checks. We performed differential methylation analysis using the R package *minfi* and fit generalized linear mixed models to account for underlying genetic relatedness. Two methylation loci are associated with variation in DXA-derived muscle mass after age adjustment at a false discovery rate (FDR) < 0.05. One - *LOC103231767* - is known to be differentially expressed in white blood cells with age based on RNAseq data from the vervets, but its role is unknown. Decreased muscle mass was also associated with increased methylation of *CD82*, which would be expected to decrease gene expression. In mice, knock-out of *CD82* results in loss of muscle mass and - critical for older adults - decreased activity of satellite cells necessary for myofiber repair. This may be due to the role of *CD82* in regulating autophagy in response to intracellular stressors via the Wnt signaling pathway. Additionally, we identified 15 methylation loci associated with gait speed, a reliable indicator of overall physical function in vervets as in humans. Many of these genes are implicated in the disordered aging of both bone and muscle, reflecting the known pleiotropy at some of these genes highlights the overlap of these physically integrated tissues. For example, *SOX2* and *WNT2* are both well-known regulators of Wnt signaling in osteoblasts, but the Wnt signaling pathway is also essential for the self-renewal of muscle. *LBX2* is a highly conserved gene regulating myofibril formation via the Wnt pathway. Master regulator of Wnt and other signaling pathways, *HOXC10* has also been linked to bone, muscle, and ligament formation and regeneration. We note that these associations in the HOX gene family overlap several of our previous findings in cartilage and bone methylation changes associated with bone shape and osteoarthritis in baboons. Additional genes with methylation changes in muscle associated with gait speed have only previously been associated with bone mass or fracture risk including *HAND122*, *CALCR23*, *IRX224*. In total, these findings point to the involvement of the Wnt signalling pathway in age-related decline in mobility.

Epigenetics Posters - Wednesday

PB2411. Epigenetic signatures of asthma in nasal epithelium from African ancestry populations from the CAAPA consortium

Authors:

M. Boorgula¹, M. Campbell¹, B. Szczesny², K. Kammers², I. Ruczinski², S. Chavan¹, C. Arehart¹, C. Cox¹, R. K. Johnson¹, I. R. Konigsberg¹, E. Thompson³, A. Morin³, C. G. McKennan⁴, C. Figueiredo⁵, C. N. Rotimi⁶, R. C. Landis⁷, H. Watson⁷, N. N. Hansel², I. V. Yang¹, C. O. Olopade⁸, C. Ober³, A. H. Liu⁹, CAAPA Consortium, E. Kenny¹⁰, K. C. Barnes¹, R. A. Mathias², M. A. Taub²; ¹Univ. of Colorado, Aurora, CO, ²Johns Hopkins Sch. of Med., Baltimore, MD, ³Univ. of Chicago, Chicago, IL, ⁴Univ. of Pittsburg, Pittsburg, PA, ⁵Univ. of Bahia, Salvador, Brazil, ⁶Natl. Human Genome Ctr., Howard Univ. Coll. of Med., Washington DC, DC, ⁷The Univ. of the West Indies, Queen Elizabeth Hosp., Bridgetown, St Michael, Barbados, ⁸Univ. of Chicago Med., Chicago, IL, ⁹Children's Hosp., Aurora, CO, ¹⁰Icahn Sch. of Med. at Mt Sinai, New York, NY

Abstract Body:

Asthma is a complex disease with striking disparities across racial and ethnic groups. A main goal of the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) is to understand the etiology of asthma through epigenetic mechanisms in the nasal epithelium focusing on populations with asthma health disparities. We measured DNA methylation in nasal epithelial cells from subjects collected across four CAAPA sites (Baltimore, Chicago, Denver, and Washington DC) using the Illumina EPIC array. Sample and probe level QC was performed using the minfi Bioconductor package. Cases were defined as asthmatics with asthma severity index indicating active asthma (N=149) and controls were non-asthmatics (N=182). We ran a differential methylation analysis adjusting for age, sex, plate, site, 2 ancestry principal components estimated from MEGA chip genotypes and 12 latent factors estimated using the CorrConf R package to correct for unwanted variation. CpGs with a q-value <0.05 were considered differentially methylated (DMCs). We also performed a stratified analysis in 195 adults (age ≥ 18, 83 cases and 112 controls) and 136 pediatric subjects (66 cases and 70 controls) to identify epigenetic signature differences between adult and pediatric asthmatics. CpGs were mapped to the nearest gene as well as target genes identified using an external promoter-capture dataset. We found 192 DMCs in the full sample set. Genes near the top DMCs (q<0.001) included *ZBTB16*, *CCL26*, *FKBP5*, *ATP6V0d2* and *CDHR3*, which are primarily associated with T cell development, chemokine binding, and responsiveness to glucocorticoid receptors, respectively, and are known susceptibility loci for asthma. *CDHR3* is an epithelial cell receptor for RV-C and is associated with asthma exacerbations/hospitalizations in the first 5 years of life. The DMC mapping to the *ZBTB16* gene was also among the 182 DMCs in pediatric subjects (q=0.03). Genes near other DMCs in pediatric subjects include *RAB37*, *SOX5* and *ACOT7*, all involved in immune related pathways and previously associated with asthma. The DMC near *ZBTB16* is in the promoter of *C11orf71*, and using CAAPA-generated RNA-Seq data from the same individuals, we observed a negative association in a multivariate linear regression relating methylation levels at this DMC to gene expression levels of *C11orf71* in all samples (p=0.0012), indicating potential regulation of this asthma-associated gene by methylation. A full integrative analysis of methylation and gene expression data is ongoing. The epigenetic associations we identify may reveal underlying mechanisms of asthma, including some contributing specifically to pediatric disease.

Epigenetics Posters - Thursday

PB2412*. Epigenome-wide association meta-analysis of DNA methylation with lifetime cannabis use

Authors:

F. Fang¹, B. Quach¹, K. G. Lawence², J. van Dongen³, J. A. Marks¹, S. Lundgren⁴, M. Lin⁵, V. V. Odintsova³, R. Costeira⁶, Z. Xu², L. Zhou¹, M. Mandal¹, J. Vink⁷, L. J. Bierut⁸, M. Ollikainen⁴, J. Taylor², J. Bell⁶, J. Kaprio⁴, D. Boomsma³, K. Xu⁹, D. P. Sandler², D. Hancock¹, E. Johnson¹⁰; ¹RTI Intl., Research Triangle Park, NC, ²NIH/NIEHS, Research Triangle Park, NC, ³VU Univ. Amsterdam, Amsterdam, Netherlands, ⁴Univ Helsinki, Helsinki, Finland, ⁵Yale Univ., Tenafly, NJ, ⁶King's Coll. London, London, United Kingdom, ⁷VU Univ, Amsterdam, Netherlands, ⁸Washington Univ Sch Med, St Louis, MO, ⁹Yale Sch. Med., New Haven, CT, ¹⁰RTI Intl., Duluth, MN

Abstract Body:

Cannabis is a commonly used drug. Understanding its links to adverse health outcomes would be enhanced by the availability of a reliable biomarker of cannabis exposure. DNA methylation can serve as a reliable biomarker of substance exposures, as previously shown for cigarette smoking. We conducted an epigenome-wide association study (EWAS) of peripheral blood-based CpG methylation and lifetime cannabis use (ever vs. never) in a meta-analysis of 7 European and the United States cohorts, including 9,436 participants (7,795 European- and 1,641 African-ancestry). Our EWAS meta-analysis revealed four CpG sites significantly associated with lifetime cannabis use at a false discovery rate of 0.05 ($p < 5.85e-7$): cg22572071 near gene *ADGRF1*, cg15280358 in *ADAM12*, cg00813162 in *ACTN1*, and cg01101459 near *LINC01132*. None of the four cannabis use-associated CpGs has been reported at genome-wide significance in EWAS for any substance exposure, including cigarette smoking, which is an important comorbid factor for cannabis use. Although our EWAS model accounted for cigarette smoking as a covariate, we further investigated the associations between CpG methylation and lifetime cannabis use in the subset of participants who never smoked cigarettes (N=3,861). All four top CpGs from the overall analysis remained associated with cannabis use in never smokers ($p < 0.05$). Additionally, the EWAS meta-analysis for cannabis use in never smokers identified another epigenome-wide significant CpG, cg14237301 annotated to *APOBR*, which was implicated in a prior genome-wide association study of lifetime cannabis use ($p = 7.56e-9$).

To examine whether a peripheral blood-based biomarker is indicative of lifetime cannabis use, we trained a multi-CpG classifier with penalized regression models (LASSO) on DNA methylation data from our single largest cohort, the Sister Study (N=2,073). We compared predictive models trained from the significant CpG sets obtained in EWAS using models with and without cigarette smoking as a covariate, with different p-value cutoffs. Based on the area under the receiver operating characteristic (AUROC) metric, predictors trained on the CpGs sets from the EWAS model without cigarette smoking as a covariate performed better. The best predictor, consisting of 64 CpGs, produced an AUROC=0.67 ($p = 9.57e-10$) in an independent dataset from the Sister Study (N=517). Thus, our preliminary results in the Sister Study demonstrated that the methylation signature can be used to predict lifetime cannabis use with reasonable accuracy.

Epigenetics Posters - Wednesday

PB2413. Epigenome-wide association of DNA methylation markers for dilated cardiomyopathy in left ventricular heart tissues.

Authors:

K. Tan¹, W. Tan², E. Wong³, M. Morley⁴, G. K. Singhera⁵, P. J. Hanson⁵, T. Cappola⁴, K. Margulies⁴, R. Foo⁶, M. Loh¹, F. Tai¹; ¹Lee Kong Chian Sch. of Med., Nanyang Technological Univ., Singapore, Singapore, ²Stanford Cardiovascular Inst., Stanford Univ., Palo Alto, CA, ³Genome Inst. of Singapore, Agency for Sci., Technology and Res., Singapore, Singapore, ⁴Penn Cardiovascular Inst., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ⁵Ctr. for Heart Lung Innovation, Univ. of British Columbia, Vancouver, BC, Canada, ⁶Yong Loo Lin Sch. of Med., Natl. Univ. of Singapore (NUS), Singapore, Singapore

Abstract Body:

Dilated cardiomyopathy (DCM) is the most common form of cardiomyopathy worldwide. It is characterized by a thinning and weakening of the left ventricular heart walls, resulting in contractile dysfunction. An estimated one-third of DCM cases are inherited, suggesting a strong genetic contribution. A further half of DCM cases remain unexplained by known causes, genetic or non-genetic, and receive the idiopathic classification. With most DCM patients eventually succumbing to sudden cardiac death (SCD), developing methods for early diagnosis is an urgent task.

Epigenetic mechanisms have been suggested to underlie the marked variability in the onset and severity of idiopathic DCM. Nevertheless, the biological relevance of significant hits surfaced by genome-wide association studies (GWAS) and epigenome-wide association studies (EWAS) of heart failure has been hindered by insufficient power and heterogeneity of etiology. Additionally, cross-study replicability of findings has been limited owing to differences in the array technologies employed and resultant variations in CpG sites coverage.

The current study is by far the largest EWAS for DCM as a single etiology of heart failure. This investigation aims to expand the current set of DNA methylation loci with statistically-significant disease associations, utilising the Illumina Infinium Methylation EPIC array which covers >850,000 methylation sites on left-ventricular heart tissues from DCM cases (n=172) and non-heart failure controls (n=174). Concurrent transcriptomic profiling of a subset of samples was conducted via RNA sequencing (RNAseq).

Pilot analysis of the methylation data surfaced 28 independent methylation signals ($P < 1 \times 10^{-5}$) independent of adjustments for confounders, including weight, height, heart mass, diabetes, and hypertension status. Among the 754 genes within 500kb of the 28 sentinel methylation loci, 302 were found to be expressed in our RNAseq dataset. Of the 302 expressed genes, a subset of 90 contained significant expression quantitative trait loci (eQTLs; $P < 1.7 \times 10^{-4}$). The significant eQTLs include an established marker for heart failure with plausible involvement in cardiac remodeling, suggesting an influence of dysmethylation on pathological gene expression in idiopathic dilated cardiomyopathy. With ongoing validation analyses on independent heart tissue samples and generation of causal inferences via applying statistical methods such as Mendelian Randomisation, the findings of this study could facilitate the identification of biomarkers with robust associations with DCM, toward an improved mechanistic understanding of disease and risk stratification.

Epigenetics Posters - Thursday

PB2414*. Epigenome-wide association study in multi-ethnic Asian populations identifies novel markers for incident type 2 diabetes

Authors:

M. Loh^{1,2,3}, D. Tay¹, L. Lakshmanan¹, H. Ng¹, F. Tai¹, R. M. van Dam⁴, X. Sim⁴, J. C. Chambers^{1,2}; ¹Lee Kong Chian Sch. of Med., Singapore, Singapore, ²Imperial Coll. London, London, United Kingdom, ³Natl. Skin Ctr., Singapore, Singapore, ⁴Saw Swee Hock Sch. of Publ. Hlth., Singapore, Singapore

Abstract Body:

Introduction Type 2 diabetes (T2D) is a major public health problem that currently affects 425 million people worldwide. The burden of diabetes is especially high in Asia, with both East and South Asians at increased risk for T2D. This increased risk of T2D in Asians is not accounted for by traditional risk factors such as obesity, diet and physical inactivity. Recent studies have suggested that epigenetic factors, in particular DNA methylation, may be associated with T2D development. However, these studies have been carried out predominantly amongst people of European ancestry, consistent with genetics and other omics data worldwide. In this study, we aim to identify variations in DNA methylation associated with incident T2D among a large multi-ethnic Asian population comprising of 6,168 East Asians, Malays and South Asians from United Kingdom and Singapore.

Methods We performed an epigenome-wide association study using peripheral blood samples from East Asians, Malays and South Asians with incident T2D with age, gender and ethnicity-matched controls from UK and Singapore. DNA methylation was measured on the Illumina 450K and EPIC array, using baseline samples collected before onset of T2D. Epigenome-wide significance was set at $P < 8.62E-8$, a threshold determined via permutation testing. To further evaluate the association between the identified T2D methylation sites with gene expression both in *cis* and in *trans* in our South-East Asian population, we additionally performed RNAseq in approximately 1,200 samples.

Results We identified 420 CpG sites across 314 independent loci to be significantly associated with incident T2D, with relative risk per 1% increase in methylation ranging from 0.87 to 1.18 (P-values: $8.51E-08$ to $9.08E-46$). Among the 314 sentinel markers, only 15 (5%) has been previously reported to be associated with T2D. Our top five hits were cg06500161 in *ABCG1*, cg11024682 in *SREBF1*, cg19693031 in *TXNIP*, cg19758958 in *AHNAK*, and cg19750657 in *UFMI*. Replication testing in independent cohorts is currently underway. Initial functional annotation and expression quantitative trait loci (eQTL) analysis suggest that the identified methylation are likely to be play functionally important roles.

Conclusion Epigenome-wide association studies performed in non-European populations may identify novel methylation markers and provide new insights into the pathways underlying T2D, as well as open up new strategies for risk stratification and prevention of T2D in Asian and other populations.

Epigenetics Posters - Wednesday

PB2415. Epigenome-wide association study reveals CpG sites associated with thyroid function and regulatory effects on *KLF9*

Authors:

A. Teumer^{1,2,3}, A. Weihs¹, L. Chaker⁴, T. C. Martin^{5,6}, J. T. Bell⁶, M. Medici⁴, ThyroidOmics Consortium; ¹Univ. Med. Greifswald, Greifswald, Germany, ²DZHK (German Ctr. for Cardiovascular Res.), Partner Site Greifswald, Greifswald, Germany, ³Med. Univ. of Bialystok, Bialystok, Poland, ⁴Erasmus Med. Ctr., Rotterdam, Netherlands, ⁵Icahn Sch. of Med. at Mount Sinai, New York, NY, ⁶King's Coll. London, London, United Kingdom

Abstract Body:

Thyroid hormones play a key role in cellular growth, development and metabolism, and are known regulators of gene expression through both genomic and non-genomic processes including DNA methylation. The aim of this study was to examine associations between thyroid hormones and DNA methylation.

We carried out an epigenome-wide association study of leucocyte DNA methylation sites from blood in up to 7,073 participants from 8 cohorts of both European and African ancestry individuals from the ThyroidOmics Consortium. Significant associations from the discovery stage were replicated in independent samples. The validated findings were correlated with gene expression levels and genetic variants. Causal influence of thyroid hormones on the DNA methylation levels was assessed by Mendelian randomization.

Epigenome-wide significant associations (p -value $< 1.1E-7$) of 3 CpGs for free T4, 5 for free T3, and 2 for TSH were discovered and replicated in independent cohorts (combined p -values = $1.5e-9$ to $4.3e-28$). The associations included CpG sites annotated to *KLF9* (cg00049440) and *DOTIL* (cg04173586) that overlap with all three traits, with consistent effect directions. Significant associations were also found for CpGs in *FKBP5* for free T4, and at *CSNK1D/LINCO1970* and *LRRC8D* for free T3. Differences in circulating TSH levels have a causal effect on DNA methylation of *KLF9*. DNA methylation of cg00049440 in *KLF9* was inversely correlated with *KLF9* gene expression in blood. The CpG at *CSNK1D/LINCO1970* overlapped with THRA binding peaks in liver cells. The total additive heritability of the methylation levels of the six significant CpG sites is between 25% and 57%. Significant methylation QTLs were identified for CpGs at *KLF9*, *FKBP5*, *LRRC8D* and *CSNK1D/LINCO1970*. We report novel associations between TSH, thyroid hormones and blood-based DNA methylation. This study advances our understanding of thyroid hormone action, and serves as a proof-of-concept that similar integrations of EWAS and other OMICS techniques can provide a valuable tool for unravelling thyroid hormone signaling next to classical *in-vitro* and animal studies.

Epigenetics Posters - Wednesday

PB2416. Epstein-Barr Nuclear Antigen 2 (EBNA2) types 1 and 2 have shared and distinct host DNA binding partners.

Authors:

K. Viel¹, S. Parameswaran², O. Donmez², C. Forney², L. Kottyan², M. Weirauch²; ¹Univ. of Cincinnati, Cincinnati, OH, ²Cincinnati Children S Hosp. Med. Ctr., Cincinnati, OH

Abstract Body:

Epstein-Barr Virus (EBV) is a human gammaherpesvirus that infects 90-95% of the human population and is a risk factor for several autoimmune diseases including Multiple Sclerosis. Furthermore, EBV infection increases one's risk for certain cancers such as Burkitt's lymphoma and nasopharyngeal carcinoma. There are two distinct genetic variations of EBV: EBV1 and EBV2, which differ substantially within Epstein Barr Nuclear Antigen 2 (EBNA2), a transactivator protein that mimics activated Notch. EBNA2 lacks a DNA binding domain, and instead regulates genes through interactions with human transcription factors such as RBPJ, EBF1, and PU.1. Type 1 and type 2 EBNA2 proteins share only 55% amino acid identity between the B95.8 (EBV1) and AG876 (EBV2) virus strains. The difference in genomic binding patterns between EBNA2 type 1 and type 2 is currently unknown. We hypothesized that EBNA2 type 2 has a different genomic binding pattern than EBNA2 type 1 and regulates different genes. To test this hypothesis, we performed ChIP-seq in B cell lines transformed with EBV2 (Jiyoye and AG876) and EBV1 (MutuIII and GM12878) to examine how EBNA2 genomic binding regions vary between EBV1 and EBV2. Through differential peak analysis, we found that ~2629 peaks are shared between EBV1 and EBV2, ~3704 peaks are unique to EBV2, and ~6727 peaks are unique to EBV1. We then computationally identified transcription factor DNA binding motifs that are enriched within these genomic regions. PU.1 motifs are similarly enriched in EBNA2 type 1 and type 2 peaks. Predicted EBF1 motifs are found at a higher frequency in EBNA2 type 1 than EBNA2 type 2, while AP-1 and RBPJ is found at higher percentages in EBNA2 type 2 peaks than EBNA2 type 1 peaks. These findings suggest that type 1 and type 2 EBNA2 have both shared and differential binding partners. As predicted by motif analysis, when we performed anti-PU.1, anti-EBF1, anti-AP-1, and anti-RPBJ ChIP-seq, we found that RBPJ and AP-1 (JUNB) colocalized with EBNA2 type 2 at more genomic regions than EBNA2 type 1. Moreover, in agreement with our predictions, EBF1 colocalized with EBNA2 type 1 at more genomic regions than EBNA2 type 2. Future studies will identify shared and unique disease risk loci that are enriched at EBNA2 from EBV type 1 and type 2.

Epigenetics Posters - Thursday

PB2417*. Expanded studies of a methylation-based COVID-19 classification model to predict severity of disease and its ability to differentiate from other respiratory viruses.

Authors:

B. Peterson¹, W. Zhou², G. F. Harrison³, M. P. Boorgula⁴, M. Campbell⁴, S. Chavan⁴, B. Barnes⁵, R. PORECHA⁵, R. A. Mathias⁶, I. V. Yang⁴, C. Gignoux⁴, A. Taye⁵, A. Monte⁴, K. C. Barnes⁷; ¹Tempus Labs, Inc., Boulder, CO, ²Tempus Labs, Inc., Redwood City, CA, ³Tempus Labs, Inc., Chicago, IL, ⁴Univ. of Colorado, Aurora, CO, ⁵Illumina, San Diego, CA, ⁶Johns Hopkins Univ., Baltimore, MD, ⁷Univ. of Colorado, Anschutz Sch. of Medicin, Aurora, CO

Abstract Body:

SARS-CoV-2 infection triggers many molecular changes including epigenome patterns, in humans. Previously, we demonstrated that DNA methylation signatures differentiate patients with SARS-CoV-2 infection from uninfected individuals. Methylation Risk Scores (MRS) derived from the differential methylation signatures yielded highly predictive scores to determine the presence of SARS-CoV-2 infection (AUC=93.6%) and measure COVID-19 disease severity (AUC=79.1%-84.4%; Konigsberg et al 2021 *Comm. Med.*). In the original study, we observed a positive trend towards specificity of the COVID-19 disease signatures in comparison with other viral upper respiratory infections. However, the study was underpowered to determine the utility of the model to create disease classifiers specific to non-SARS-CoV-2 infections. To expand on this work, we profiled peripheral blood samples from 304 additional patients on the customized Infinium MethylationEPIC array. These patients were either uninfected, infected with SARS-CoV-2, or SARS-CoV-2 negative but infected with other respiratory viruses. Measurements for disease severity, progression (hospitalization, ICU admittance, ventilator use), and vaccination and booster status were extracted from electronic health record data. An epigenome-wide association analysis in this expanded cohort validated genes and pathways that were previously found to be significantly associated with SARS-CoV-2 infection. The previously reported sparse regression based MRS, when tested in this expanded cohort, yielded higher AUCs for case-vs-control status, hospitalization, ICU admission, and progression to death. Additionally, we observed higher predictive scores for other respiratory viruses. In summary, the COVID-19-specific epigenetic signature (in peripheral blood) driven by expression/activation of key immune-related pathways was related to infection status, disease severity, and clinical deterioration. This study provides useful insights for diagnosis and prognosis in patients with COVID-19.

Epigenetics Posters - Wednesday

PB2418. Exploring novel regulatory elements in X-chromosome inactivation.

Authors:

D. Beseiso¹, M. Cloutier¹, M. Hinten², S. Kalantry¹; ¹Univ. of Michigan Med. Sch., Ann Arbor, MI, ²Mayo Clinic, Rochester, MN

Abstract Body:

X-chromosome inactivation equalizes X-linked gene expression between the mammalian sexes by silencing genes on one of the two X chromosomes in females. A long non-coding RNA known as the *X-inactive-specific transcript* (*Xist*) is expressed exclusively from the future inactive X chromosome where it facilitates the deposition of repressive histone marks to initiate gene silencing. Replicated copies of the inactive X chromosome remain silenced through subsequent rounds of cell division making X-inactivation a paradigm of epigenetic transcriptional regulation. Our lab recently found that loss of the *Xist* gene results in a defected X-inactivation in mouse trophoblast stem cells. Unexpectedly, in a separate series of experiments, we found that loss of *Xist* RNA expression, while leaving the underlying DNA sequence intact, results in a significantly milder derepression of silenced genes. These findings may suggest that the *Xist* DNA silences X-linked genes through other mechanisms in addition to the expression of *Xist* RNA. The goal of my work is to explore alternate mechanisms by which silencing of genes on the inactive X chromosome is maintained. The results may shed light on novel regulatory elements of X-inactivation as well as reveal previously uncharacterized mechanisms of gene silencing at the DNA and chromatin levels.

Epigenetics Posters - Thursday

PB2419. Extracellular vesicles microRNAs in traumatic brain injury patients with and without intracranial abnormality

Authors:

R. Vorn¹, C. Lai², C. Turtzo³, C. Devoto⁴, N. Peterkin³, L. Latour³, J. Gill¹; ¹Johns Hopkins Univ., Baltimore, MD, ²Ctr. for NeuroSci. and Regenerative Med., Uniformed Services Univ. of Hlth.Sci., Bethesda, MD, ³Natl. Inst. of Neurological Disorders and Stroke, NIH, Bethesda, MD, ⁴Henry M. Jackson Fndn. for the Advancement of Military Med., Bethesda, MD

Abstract Body:

Mild traumatic brain injury (mTBI) is a major public health burden; however, objective diagnosis remains challenging. Extracellular vesicle (EVs) derived from biofluids is a promising source of biomarkers specific to neurological disorders, including TBI. However, EV-associated microRNAs (miRNAs) have not been adequately studied in acute mTBI patients. Therefore, the objective of this study was to investigate the EV-associated miRNAs from plasma of mTBI patients with and without evidence of intracranial findings. This analysis was a part of the Traumatic Head Injury Neuroimaging Classification (THINC) study and included 27 healthy controls (HCs) and 190 mTBI patients with a mean age of 46 years. Patients presented to the emergency department, received standard of care computed tomography (CT), and received research magnetic resonance imaging (MRI) and peripheral blood draw within 48 hours of injury. For subgroup analysis, TBI patients were classified based on imaging findings, specific to TBI, as CT-MRI- (n= 53), CT-MRI+ (n = 60), and CT+MRI+ (n = 53). EVs-miRNAs profiling was measured using the nCounter miRNA expression panel (human 798 targets) developed by NanoString Technology. We identified 67 significantly deregulated (32 upregulated and 35 downregulated) miRNAs in mTBI patients compared to HCs with the cutoff 0.05 False Discovery Rates and \pm 1.5-fold change. In subgroup analysis, we found 42 miRNAs significantly dysregulated (17 upregulated and 25 downregulated) in mTBI patients with MRI-CT-, 83 miRNAs significantly dysregulated (36 upregulated and 47 downregulated) in mTBI patients with MRI+CT-, and 54 miRNAs significantly dysregulated (26 upregulated and 28 downregulated) in mTBI patients with MRI+CT+ relative to HCs. We observed that 27 miRNAs were overlapping among subgroups and overall TBI patients. Top 10 deregulated miRNAs are miR-1283, miR-496, miR-601, miR-32-5p, miR-376c-3p, miR-6724-5p, miR-1268a, miR-124-3p, miR-296-3p, and miR-579-5p. Our study provided insight into the molecular changes after mTBI and has potential application as diagnostic biomarkers for mTBI.

Epigenetics Posters - Wednesday

PB2420. Finding the Right Match: the Role of ZCWPW1 in Mammalian Meiosis

Authors:

W. Xie, D. Bazzano, S. Hammoud; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Meiosis is an essential process of sexually reproducing organisms to maintain the copy number of chromosomes and generate genetic diversity across generations, characterized by a single round of DNA replication followed by two ensuing divisions. While the second division separates sister chromatids, resembling mitosis, the first division is reductive with the segregation of homologous (maternal and paternal) chromosomes. Faithful segregation of homologs requires a series of coordinated steps: chromosome pairing, synapsis, and homologous recombination (HR) during meiotic prophase I. These events are initiated by the generation of double-strand-breaks (DSBs) at sites specified by PRDM9, a methyltransferase that tri-methylates lysine 4 and 36 of histone 3 (H3K4me3 and H3K36me3). However, it remains poorly understood how H3K4me3/H3K36me3 modifications facilitate pairing and recombination.

Recently, ZCWPW1 is identified as a dual H3K4me3 and H3K36me3 “reader”. The *Zcwpw1* mutants display defects in synapsis, and crossover formation, leading to male sterility consequently. To investigate the function of ZCWPW1 in meiosis, we performed ZCWPW1 immunoprecipitation followed by mass spectrometry and found that ZCWPW1 interacts with a diverse set of proteins including those that function in chromosome movement, chromosome axis assembly, and DNA repair. Specifically, components of the linker of nucleoskeleton and cytoskeleton (LINC) complex and interacting proteins are identified. SUN2, a component of the LINC complex, mislocalizes in the absence of ZCWPW1. While the telomere movement is not abolished, the dynamics is potentially slowed down. Using immunofluorescence combined with fluorescent in situ hybridization (FISH), we found that homolog pairing of chromosome 1 and 19 are impaired in ZCWPW1 knockout mice, with frequent heterologous synapsis and asynapsis.

Given these results, we hypothesize that ZCWPW1 facilitates homolog pairing by interacting with the LINC complex to promote telomere-led chromosome movement and bringing euchromatin into proximity. These studies will help illustrate how a histone methylation writer/reader system contributes to the germline integrity, faithful genomic transmission, and speciation. Additionally, identifying a euchromatic pairing mechanism will shed light on how cells ensure the homolog pairing on a large scale in conjunction with telomere-driven movement and local homology search.

Epigenetics Posters - Thursday

PB2421. First trimester human placenta DNA methylation correlates to maternal estradiol levels

Authors:

T. Gonzalez¹, B. Lee¹, T. Sun¹, K. D. Taylor², J. I. Rotter², Y-D. I. Chen², J. Williams III¹, S. S. Rich³, C. R. Farber³, M. O. Goodarzi¹, J. Cui¹, M. D. Pisarska¹; ¹Cedars-Sinai Med. Ctr., Los Angeles, CA, ²Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, ³Univ. of Virginia, Charlottesville, VA

Abstract Body:

Background: The placenta is critical for normal fetal development and may impact health outcomes later. Gamete manipulation during in vitro fertilization (IVF) has previously been implicated to alter DNA methylation in the placenta. However, recent studies demonstrate minimal differences, but suggest underlying infertility and/or other treatments may be contributory. We found elevated 17-beta-estradiol and progesterone levels in the late first trimester of subjects pregnant through fertility treatments, compared to those conceived spontaneously. To determine if elevated hormonal states drive some of the methylation differences, we investigated if maternal hormones correlate to human placenta DNA methylation in first trimester, regardless of infertility status. **Methods:** Pregnant subjects were recruited into the SMAART cohort. Late first trimester placenta was collected from leftover sample after prenatal diagnostic testing, chorionic villus sampling. Maternal plasma was collected in the same visit and ELISA assays were previously performed to measure maternal levels of 17-beta-estradiol and progesterone (PMID: 30445606). Here, DNA methylation of first trimester placenta was measured using the Illumina Infinium MethylationEPIC Array. Probes were excluded if identified as masked (unreliable), if not on autosome chromosomes, or if located at single nucleotide polymorphism sites, leaving 741,145 CpG sites for analysis. We used data from N=88 subjects (37 spontaneous conceptions, 25 non-IVF fertility treatments, and 26 IVF fertility treatments) and a generalized linear model to find correlations between placenta DNA methylation and maternal hormones. Covariates of fetal sex and maternal age were considered. **Results:** In the generalized linear model without covariates, DNA methylation correlated positively with 17-beta-estradiol in maternal plasma at 5 CpG sites with genome wide significance at $P < 9E-8$. Significant sites were cg20868379 (*LRRC20* gene), cg13873263 (no gene context), cg06929324 (*VRTN* gene), cg08140055 (*EPHB1* and *KY* genes), and cg22611570 (*SLC2A9* gene). There were no CpG methylation sites that correlated with progesterone levels. Fetal sex did not correlate to either estradiol ($P=0.889$) or progesterone ($P=0.0667$) and was not used as a covariate. Maternal age also did not correlate to estradiol ($P=0.452$) or progesterone ($P=0.234$). **Conclusion:** Maternal 17-beta-estradiol levels, but not progesterone levels, correlate to CpG site methylation in first trimester human placenta. Clinically, embryo transfers into a physiologic hormonal environment may minimize these outcomes.

Epigenetics Posters - Wednesday

PB2422. Functional Analysis of CHD2 and CHD8 *De Novo* Mutations and Related Molecular Events Involved in Neurodevelopmental Disorders

Authors:

T. Muhammad¹, S. F. Pastore¹, T. T. Madanagopal², A. Mikhailov², J. B. Vincent¹; ¹Univ. of Toronto, Toronto, ON, Canada, ²Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada

Abstract Body:

Background: Neurodevelopmental disorders (NDDs) such as ID and ASD are associated with impaired intellectual and adaptive functioning, deficits in social communication accompanied by the presence of limited interests and repetitive behaviors. A number of genes that have been identified to regulate chromatin remodeling-an active and key epigenetic mechanism of altering the structure and function of chromatin- have been implicated in both disorders. Chromodomain helicase DNA-binding (CHD) proteins are important factors for remodeling chromatin and for gene regulation. CHD2 and CHD8 pathogenic variants have been extensively reported for ID and ASD. **Objectives:** The initial goals of this study are to evaluate the effects of *CHD2* and *CHD8* pathogenic missense mutations on DNA/chromatin binding dynamics, on protein stability and gene expression, and to assess the effects of mutations in these genes on neuronal morphology and differentiation. **Methods:** We performed a review of the literature to identify de novo missense mutations, and checked for prediction of pathogenicity using SIFT, PROVEAN, Mutation taster, and Polyphen. Next, we cloned *CHD2* and *CHD8* into vectors for transfection into cell lines. We examined methyltransferases via immunoblotting followed by RT-qPCR and ChIP experiments to evaluate important genes involved in development. We have generated in silico data for the effects of SNVs on the CHD2/CHD8 protein surface energy, stability, flexibility and binding dynamics using docking and MD simulations analysis. **Results:** Preliminary findings revealed that these mutations affect the expression of active histone H3.3, key methyltransferases such as RbBP5, Set8, Set7/9, ESET etc., that are required for numerous development-related gene expressions and regulation. Preliminary ChIP data suggests that these mutations affect the binding dynamics of the CHD2 and CHD8 proteins. **Conclusion:** These pathogenic mutations affect the expression of active histone H3.3, dysregulate various methyltransferases, and affect binding dynamics of CHD2 and CHD8 proteins. We are now utilizing the CRISPR/Cas9 system to knockout CHD2 and CHD8, and introduce point mutations, in HEK293T, P19 embryonic cells and iPSCs to get isogenic cell lines for detailed mechanistic studies.

Epigenetics Posters - Thursday

PB2423*. Gene regulatory network synchronizes genetic and epigenetic signals, prioritizes GWAS SNPs, and identifies repurposable drug candidates for multiple sclerosis

Authors:

A. Manuel, Y. Dai, Z. Zhao; The Univ. of Texas Hlth.Sci. Ctr., Houston, TX

Abstract Body:

Background: Multiple sclerosis (MS) is an autoimmune disease that leads to demyelination and neurological disability. Although the exact cause of MS is undetermined, both genetic and environmental risk factors have been established. Genome-wide association studies (GWAS) have identified 233 single nucleotide polymorphisms (SNPs) associated with MS with genome-wide significance. Epigenetic studies have pinpointed differentially methylated CpG sites in the genomes of individuals with MS. However, the interplay between genetic factors and epigenetic regulation remains elusive. **Methods:** We applied systems biology approaches to harmonize the effects of genome-wide genetic and epigenetic signals in MS. We extended the dense module search of GWAS (dmGWAS) network model to integrate SNP-level association signals from the largest MS GWAS of 14,802 MS cases and 26,703 controls, DNA methylation profiles from 140 MS cases and 139 controls, and the human interactome. We obtained differentially methylated genes by aggregating additive effects of differentially methylated CpG sites within respective promoter regions. Literature-curated transcription factor (TF) interactions were used to reconstruct a gene regulatory network (GRN). Colocalization of the MS GWAS and methylation quantitative trait loci (mQTL) was performed to assess the genes of GRN. Cell-type specificity of the GRN was also investigated by the Web-based Cell-type Specific Enrichment Analysis of Genes (WebCSEA). Lastly, we performed drug target enrichment analysis using the Therapeutic Target Database (TTD). **Results:** The MS-associated GRN was comprised of 25 genes and several TF interactions. We highlight genome-wide significant SNPs with GWAS-mQTL colocalization pairs: rs6032663, rs6065926, and rs2024568 of *CD40* locus, rs9913597 of *STAT3* locus, and rs887864 and rs741175 of *CIITA* locus. Interestingly, *CD40* and *STAT3* had aligned mQTL and eQTL signals, indicating these SNPs may be affecting epigenetic mechanisms and gene expression. WebCSEA analysis showed that the GRN was enriched (p-value = 0.0016) in T follicular helper (Tfh) cells. The GRN was also enriched (p-value = 3.89×10^{-4}) with three successful drug targets and eight clinical trial drug targets, which revealed repurposable candidates for MS treatment: vorinostat (*HDAC1* inhibitor), napabucasin (*STAT3* inhibitor), and others. **Conclusion:** We prioritize a GRN (25 genes) with genetic and epigenetic signals, Tfh cell-specificity, and drug target genes. We pinpoint GWAS SNPs that may alter DNA methylation and gene expression patterns in MS. We propose repurposable candidates for MS, which warrant further investigation.

Epigenetics Posters - Wednesday

PB2424*. Genetic variation in correlated regulatory region of immunity

Authors:

D. Avalos^{1,2}, G. Rey², D. M. Ribeiro¹, A. Ramisch², E. T. Dermitzakis², O. Delaneau¹; ¹Univ. of Lausanne, Lausanne, Switzerland, ²Univ. of Geneva, Geneva, Switzerland

Abstract Body:

Studying the interplay between genetic variation, epigenetic changes and regulation of gene expression in immune cells is important to understand the modification of cellular states in various conditions, including immune diseases. Here, we built cis maps of regulatory regions with coordinated activity - Cis Regulatory Domains (CRDs) - in neutrophils, monocytes and T cells. For this, we leveraged (i) whole-genome sequencing (WGS), (ii) chromatin immunoprecipitation sequencing (ChIP-seq), (iii) DNA methylation (450k arrays), and (iv) transcriptional profiles (RNA-seq) from the BLUEPRINT consortium, for up to 200 individuals. Our study uncovers 9287, 7666 and 5480 histone CRDs (hCRDs) and 6053, 6112, 5701 methyl CRDs (mCRDs) in monocytes, neutrophils and T-cells, respectively. We discovered 15294 hCRD-gene and 6185 mCRD-gene associations (5% FDR). Only 33% of hCRD-gene associations and 37% of mCRD-gene associations were shared between cell-types, revealing the dynamic nature of regulatory interactions and how similarly located regulatory regions modulate the activity of different genes on different cell types. We mapped Quantitative Trait Loci associated with CRD activity (CRD-QTLs) and found that 89% and 70% of these hCRDs and mCRDs are under genetic control highlighting the importance of genetic variation to study the coordination of cellular regulatory programs. We found CRD-QTLs to be enriched in cell-type-specific transcription factor binding sites, such as SPI1, STAT3, RFX1, SOX4, ATF3 for neutrophils and monocytes and TCF4 and BCL11A for T-cells, in line with the Human Protein Atlas. We integrated PChi-C data, which showed that most significant associations discovered within gene-CRD associations and co-expressed genes associated with the same CRD, involving large genomic distances, tend to happen between genomic regions in close spatial proximity. Finally, we mapped trans regulatory associations between CRDs, which enabled the discovery of 207 trans-eQTLs across cell types. Overlapping our hits with trans eQTLs from eQTLGen Consortium meta-analysis in whole blood revealed 81 trans-eQTLs shared between the two studies. Overall, we show that mapping functional regulatory units using population genomics data allows discovering important mechanisms in the regulation of gene expression in immune cells and gain a greater understanding of cell-type specific regulatory mechanisms of immunity.

Epigenetics Posters - Thursday

PB2425. Genome-wide DNA methylation profiling in subcortical regions reveals epigenetic signatures associated with PTSD and MDD

Authors:

H. Li¹, J. Wang², D. Cruz³, J. Modliszewski⁴, D. Corcoran⁵, Traumatic Stress Brain Research Group, J. Krystal^{6,7}, R. Duman⁸, H. Zhao^{1,2}, D. Williamson³, M. J. Girgenti^{6,7}; ¹Dept. of Biostatistics, Yale Sch. of Publ. Hlth., New Haven, CT, ²Program of Computational Biology and Bioinformatics, Yale Univ., New Haven, CT, ³Dept. of Psychiatry and Behavioral Sci., Duke Univ. Med. Ctr., Durham, NC, ⁴Duke Univ., Durham, NC, ⁵Duke Univ, Durham, NC, ⁶Dept. of Psychiatry, Yale Univ. Sch. of Med., New Haven, CT, ⁷Natl. Ctr. for PTSD, U.S. Dept. of VA, New Haven, CT, ⁸Yale Univ., New Haven, CT

Abstract Body:

Methylation of DNA (DNAm) is a critical regulator of genome architecture, gene expression, and cell function. These processes are important for mammalian development, genomic imprinting, aging and response to external stimuli such as stress. Post-traumatic stress disorder is a debilitating psychiatric disorder with an approximately 7% prevalence in the general population and occurs in the aftermath of a traumatic event, making PTSD one of the best psychiatric disorders to be understood from an epigenetic standpoint. We generated DNAm data from 6 postmortem brain regions (subiculum, dentate gyrus, CA subfields, basolateral amygdala, medial amygdala and central amygdala nuclei) from 171 individual donors using the Illumina Methyl-seq system. This resource provides genomic coverage across (1,065,750 probes representing 22,544 genes, a large sample size (117 cases versus 54 controls), across two unique diagnostic cohorts (PTSD and MDD) in regions not typically analyzed in large genomic cohorts (amygdala and hippocampus) in an effort to comprehensively measure DNA methylation changes across the stressed brain. We find differential DNAm signatures across all regions including many non-overlapping, sex-specific differences. Gene set enrichment identified pathways including glucocorticoid signaling, inflammation and GABAergic signaling as mechanisms differentially enriched in PTSD and MDD. DNAm signals aggregate near PTSD risk variants for genes such as *ELFN1*, *CRHI*, and *MAD1L1*. Taken together, we present a unique and powerful resource for exploring DNA methylation changes in human postmortem subcortical regions important for stress disorder pathology.

Epigenetics Posters - Wednesday

PB2426*. Genome-wide dysregulation of R-loops in Ataxia Telangiectasia neurological pathogenesis.

Authors:

K. Westover, Y. Hou, Y. Li, B. Yao; Emory Univ., Atlanta, GA

Abstract Body:

Ataxia Telangiectasia (AT), a multisystemic neurodegenerative disease characterized by decreasing motor coordination, mental development, immune defects, and telangiectasia of the eyes, affects up to 1 in 40,000 to 100,000 people worldwide. A recessive early childhood onset disorder, AT is caused by mutations within the ataxia telangiectasia mutated (ATM) threonine/serine kinase which plays crucial roles within the DNA damage response (DDR). However, the precise molecular mechanisms underlying AT pathogenesis and how ATM loss-of-function leads to deficient DDR remain elusive. R-loops, three stranded RNA-DNA structures composed of an DNA-RNA hybrid and a non-template DNA strand, have emerged as key components of double strand break (DSB)-induced DDR. Mounting evidence has documented critical roles of R-loops in both causing and responding to DSBs. As DSBs and the failure of their repair play major roles in the pathology of AT, R-loop dysregulation is likely to contribute to AT pathogenesis. One recently identified kinase substrate of ATM is methyltransferase like 3 (METTL3) protein, a N6-methyladenosine (m6A) methyltransferase. m6A on the RNA strand of R-loops is present inside nuclei and affects R-loop formation during DSB repair. The relationship between ATM-METTL3 phosphorylation in response to DNA damage and regulation of R-loop formation through m6A deposition, which could play crucial roles in AT pathogenesis, has yet to be defined. Our preliminary data has demonstrated a global trend of R-loops decreasing in AT patient-derived neurons compared to healthy controls. ~20% of these downregulated loci were rescued in an isogenic line where the ATM mutation had been corrected. We hypothesize that in AT, the lack of METTL3 phosphorylation by ATM could globally dysregulate R-loop formation and underlie AT progression. We have successfully inserted wild-type, phospho-mimetic, and phospho-deficient METTL3 into plasmids for future experiments to determine the role of ATM-phosphorylated METTL3 on R-loop formation. We are currently generating iPSC-derived motor neurons from age-matched healthy controls, AT patients, and their isogenic lines with the pathogenic mutations corrected by genome editing systematically identify critical R-loop loci that are associated with AT and mechanistically explore the role of ATM truncations in AT progression through METTL3-dependent R-loop regulation.

Epigenetics Posters - Thursday

PB2427*. Genome-wide evaluation of the effect short tandem repeat variation on local DNA methylation

Authors:

A. Martin Trujillo, P. Garg, N. Patel, B. Jadhav, A. J. Sharp; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Short tandem repeats (STRs) represent a class of repetitive sequences that are widely distributed throughout the human genome and contribute significantly to genetic diversity, including disease-causing variation. While the functional consequences of STRs on gene expression have been extensively assessed, their impact on epigenetics has been poorly studied. Here, we hypothesized that STRs can act as methylation quantitative trait loci (mQTL), i.e. STRs that are able to regulate DNA methylation profiles (mSTRs). To test this hypothesis, we performed association analysis between STR length and DNA methylation variation using genotypes from 131,635 polymorphic STRs derived from Whole Genome Sequencing (WGS) and the Illumina Infinium MethylationEPIC Bead Chip, respectively. After accounting for biological and technical covariates as well as multiple testing, we identified 11,870 STRs that were associated with DNA methylation levels of 11,774 unique CpGs in a discovery cohort of 245 individuals. Importantly, we validated ~90% of the detected STR:CpGs associations in a second independent cohort of 484 individuals, underscoring the robustness of the detected signals and our method. mSTRs were enriched for regulatory elements such as promoters (OR=2.57) and enhancers (OR=2.52), supporting their potential role in regulating genome function. Subsequent fine-mapping analysis indicated that 585 of these mSTRs represent the main genetic driver responsible for DNA methylation variation. Furthermore, using STR genotypes obtained from 1000 Genomes Project samples, we observed an enrichment for population divergent alleles at fine-mapped mSTRs, suggesting that they might have been subject to selection as a result of environmental pressures. Finally, by integrating GWAS-signals in our analysis, we identified dozens of fine-mapped mSTRs linked to a wide variety of human traits that range from blood-related traits to neurodegenerative diseases, providing evidence to support that variation at these STRs could underlie a fraction of human traits. Overall, our findings expand the catalog of functional variants in the genome and show that a fraction of STRs act as regulators of genome function by modulating DNA methylation levels in cis, which ultimately can represent the molecular mechanism underlying phenotypic variation and disease.

Epigenetics Posters - Wednesday

PB2428*. Genome-wide mapping implicates 5-hydroxymethylcytosines in diabetes and Alzheimer's disease

Authors:

Z. Zhang¹, A. Beadell², A. W. Capuano³, D. Bennett³, C. He², W. Zhang¹, Z. Arvanitakis³; ¹Northwestern Univ., Chicago, IL, ²Univ. of Chicago, Chicago, IL, ³Rush Univ. Med. Ctr., Chicago, IL

Abstract Body:

Background: Diabetes mellitus (DM) is a recognized risk factor for dementia including due to Alzheimer's disease (AD). Because DM is a potentially modifiable risk factor, a greater understanding of the mechanisms linking DM to the clinical expression of AD may provide insights into the much needed dementia therapeutics or intervention approaches. Epigenetic analysis offers a powerful approach, and previous studies including our team's exploration in circulating cell-free DNA (cfDNA) suggested that the under-investigated 5-hydroxymethylcytosines (5hmC) is now emerging as a promising measure to investigate in DM-related conditions. Methods: We used samples and data from the Rush Memory and Aging Project. Using the 5hmC-Seal technique, a highly sensitive chemical labeling technique developed by our team, we performed genome-wide profiling of 5hmC in cfDNA samples (n=74) from antemortem blood samples and brain tissue (n=75) genomic DNA (gDNA) from postmortem prefrontal cortex tissue from 80 deceased individuals across four groups: AD (neuropathologically defined), DM (clinically defined Type 2 DM), DM with AD (AD+DM), and non-AD/non-DM controls. Differential analysis and machine learning approaches were used to investigate 5hmC signatures associated with different conditions and summarize epigenetic scores to evaluate diagnostic potential for AD in cfDNA. Pathway enrichment analysis was performed on differential 5hmC signatures to explore biological mechanisms underlying DM-associated AD. Results: We identified distinct 5hmC signatures and biological pathways, such as Wnt signaling and inflammatory mediator regulation of TRP channels, as well as amino sugar and nucleotide sugar metabolism, that are associated with AD or AD+DM compared to non-AD/non-DM controls or DM alone. We further demonstrated the potential diagnostic value of 5hmC profiling in circulating cfDNA. Specifically, a 4-gene model showed capacity for distinguishing AD+DM from DM (AUC [area under the curve]= 82.7%; 95% CI [confidence interval], 67.6-97.9%), while an 11-gene model distinguished AD from non-AD/non-DM controls (AUC = 90.0%; 95% CI, 80.4-99.6%). Conclusions: Genome-wide 5hmC profiling uncovered epigenetic features and biological pathways related to nucleotide sugar metabolism that link DM to DM-associated AD. Our findings also demonstrated the potential of utilizing 5hmC in circulating cfDNA as diagnostic biomarkers or disease monitoring tools, with the ultimate goal of preventing or ameliorating DM-associated AD dementia and improving clinical outcomes. This study was supported partially by RF1AG074549 from the NIH.

Epigenetics Posters - Wednesday

PB2430. Genome-wide neonatal epigenetic changes associated with maternal exposure to the COVID-19 pandemic

Authors:

K. Kocher^{1,2,3}, S. Bhattacharya^{4,5}, N. Andescavage^{6,5}, C. Lopez^{6,5}, M. Almalvez^{1,5}, D. Henderson^{6,5}, E. Vilain^{1,5,3}, C. Limperopoulos⁶, E. Delot^{1,3}; ¹Children's Natl. Res. and Innovation Campus, Washington, DC, ²Children's Natl. Hosp., Washington DC, DC, ³The George Washington Univ., Washington, DC, ⁴Children's Natl. Res. and Innovation Campus, Washington DC, DC, ⁵Children's Natl. Hosp., Washington, DC, ⁶Children's Natl. Hosp., Washington, DC

Abstract Body:

With the sudden outbreak of COVID-19 in early 2020 came a prolonged period of quarantine and shutdown of society leading to widespread psychological distress for many. During gestation, stressors to the fetus, including maternal psychological distress, can fundamentally alter the neonatal epigenome and may be associated with long-term impaired neurological or developmental outcomes. In this study we aimed to determine whether there are unique epigenetic signatures in newborns associated with the stress of the COVID-19 pandemic. We investigated DNA methylation differences in newborns who experienced otherwise healthy pregnancies that occurred during the COVID-19 pandemic (Project RESCUE). Widespread differential methylation was found between newborns born and *in utero* during the pandemic (delivered after June 2020) and a cohort of age- and sex-matched healthy controls, delivered before the start of the pandemic (before December 2019). In contrast, there were no apparent epigenetic differences associated with maternal COVID-19 infection during pregnancy. Differential methylation was observed among genomic sites that underpin important neurological pathways and have been previously reported in the literature to be differentially methylated as a result of prenatal stress, such as *NR3C1*. Our study reveals that the onset and continuation of the COVID-19 pandemic has fundamentally altered the epigenomes of newborns born during this time, even in otherwise healthy pregnancies, which should be taken into account in any current and future DNA methylation study. It will also be critical to document longitudinal neurodevelopmental outcomes of these cohorts to understand if these altered epigenetic signatures predict resilience or susceptibility to the stressors.

Epigenetics Posters - Thursday

PB2431. Genome-wide screening of epigenetic variations in the EpiSign Knowledge Database

Authors:

R. Relator¹, M. Levy¹, J. Kerkhof¹, H. McConkey^{1,2}, B. Sadikovic^{1,2}; ¹London Hlth.Sci. Ctr., London, ON, Canada, ²Western Univ., London, ON, Canada

Abstract Body:

The EpiSign Knowledge Database (EKD) contains more than 10,000 genome-wide peripheral blood DNA methylation profiles including individuals with genetic neurodevelopmental disorders and unaffected controls (general population samples) with varying age and racial background. An expanding number of genetic disorders included in these cohorts have been shown to be associated with distinct epigenetic profiles, with over 120 such episignatures currently identified. While episignatures represent effective diagnostic biomarkers, they don't include assessment of individual, genomic locus-specific epigenetic changes in individual patients. To date, many studies have shown that epigenetic variation, also referred to as epivariants or epimutations, are associated with various human diseases. However, there is limited information on their prevalence in patients with known and suspected genetic neurodevelopmental syndromes. In this study we have catalogued the epivariants in the EKD by performing a robust outlier analysis. From Illumina Infinium EPIC arrays, hypo/hypermethylated regions are defined with at least three contiguous probes within 1000bp. For each sample with a known episignature disorder, we investigate all predefined regions and compare methylation levels with a fixed set of reference controls. We aggregate regions identified meeting the criteria of epivariations, followed by visual inspection. The identification of epivariants, which are usually hypo- or hypermethylation events located in gene promoter regions, can provide better understanding of disease etiology and aid in the development of regional biomarkers for prediction models in clinical diagnostics.

Epigenetics Posters - Wednesday

PB2432. Global identification of cortical circRNAs and circRNA clusters associated with Alzheimer's disease

Authors:

F. Wang¹, Y. Li¹, P. Martinez-Feduchi¹, J. Kuehner¹, Y. Feng², B. Yao¹; ¹Dept. of Human Genetics, Emory Univ. Sch. of Med., Atlanta, GA, ²Dept. of Pharmacology and Chemical Biology, Emory Univ. Sch. of Med., Atlanta, GA

Abstract Body:

Circular RNAs (circRNAs) have recently emerged as novel class of regulatory RNAs with spatial and temporal expression control during development and aging. Several key cis-regulatory elements and trans-factors were identified to contribute to the production of circRNA, including Alu repetitive elements in flanking introns and RNA-binding proteins (RBPs), respectively. Many circRNAs are evolutionary conserved from invertebrate to mammals, suggesting their essential biological functions. A great number of circRNAs are highly enriched in the brain and have been shown to play important roles in brain development, maturity, and functions. Recently data shed light on their global dysregulation and potential roles in neurodegenerative disorders, such as Alzheimer's disease (AD). However, how circRNAs participate in AD-related pathophysiology remains poorly elucidated. Our recent study established a robust platform to identify circRNA landscape in human neurons and oligodendrocytes. Using this platform, we identify and quantify circRNA landscape from early development to aging process in 5xFAD mice cortex. We identified a group of differentially expressed circRNAs in AD mouse cortex and found many of these AD-related circRNAs changed of their expression in human AD postmortem brains. Many circRNA-forming loci produce multiple circRNAs by different combinations of exon sequences. These "clustered" circRNAs may play additive effects in regulating miRNA or RBP activities through their common sequences. In addition to individual circRNAs that were significantly altered in AD, we identified 440 circRNAs clusters that were significantly differentially expressed in AD mouse brain. These circRNA clusters contained many statistically insignificant low expressing circRNAs that could additively contribute to AD pathogenesis. In summary, our study provides novel mechanistic link between circRNAs, circRNA clusters and AD pathogenesis during aging progression.

Epigenetics Posters - Thursday

PB2433*. High dimensional co-expression network analysis unravels transcriptomic drivers of diverse biological systems

Authors:

S. Morabito, V. Swarup; Univ. of California Irvine, Irvine, CA

Abstract Body:

The parameters that dictate the onset, progression, and genetic basis of human disease are encoded in biological networks. To date, many single-cell transcriptomic analyses remain at the level of individual genes rather than biological networks, partially due to a lack of sufficient tools for gene network analysis in high dimensional data. Here, we describe hdWGCNA, an analytical framework for co-expression network analysis in high dimensional transcriptomics data such as single cell RNA-seq and spatial transcriptomics. We optimized the bioinformatics protocol for co-expression network analysis for high dimensional transcriptomic data, and we maximize the usability of this pipeline via an open source R package that is directly compatible with Seurat formatted data objects. We show that hdWGCNA yields cell-type-specific co-expression modules in the human cortex, and we demonstrate that these networks are reproducible across independent datasets. We applied hdWGCNA to inhibitory neurons from published single-cell RNA-seq data of the prefrontal cortex from autism spectrum disorder (ASD) patients, identifying gene modules that contribute to neuronal dysfunction, as well as gene modules containing key ASD GWAS hits. We performed a consensus co-expression network analysis of microglia in Alzheimer's Disease (AD) using three independent single-nucleus RNA-seq datasets, revealing distinct networks that are critical in the transition from homeostatic to disease-activated cells. Additionally, gene modules identified in single cell data can be projected into relevant spatial datasets to discover their spatial expression patterns, and hdWGCNA is also capable of constructing co-expression networks in spatial transcriptomics datasets themselves. Furthermore, hdWGCNA offers a statistical framework for comparing networks across different biological systems and data modalities. Together, we demonstrate that hdWGCNA is a powerful tool for uncovering the transcriptomic drivers of diverse biological systems.

Epigenetics Posters - Wednesday

PB2434. High-dimension to high-dimension screening for detecting genome-wide epigenetic regulators of gene expression.

Authors:

H. Ke¹, Z. Ren², J. Qi³, S. Chen³, G. C. Tseng², Z. Ye³, T. Ma¹; ¹Univ. of Maryland, Coll. Park, College Park, MD, ²Univ. of Pittsburgh, Pittsburgh, PA, ³Univ. of Maryland, Baltimore, Baltimore, MD

Abstract Body:

The advancement of high-throughput technology characterizes a wide range of epigenetic modifications across the genome involved in disease pathogenesis via regulating gene expression. The high-dimensionality of both epigenetic and gene expression data make it challenging to identify the important epigenetic regulators of genes. Conducting univariate test for each epigenetic-gene pair is subject to serious multiple comparison burden, and direct application of regularization methods to select epigenetic-gene pairs is computationally infeasible. Applying fast screening to reduce dimension first before regularization is more efficient and stable than applying regularization methods alone. Here we propose a novel screening method based on robust partial correlation to detect epigenetic regulators of gene expression over the whole genome, a problem that includes both high-dimensional predictors and high-dimensional responses. Compared to existing screening methods, our method is conceptually innovative that it reduces the dimension of both predictor and response, and screens at both node (epigenetic features or genes) and edge (epigenetic-gene pairs) levels. We develop data-driven procedures to determine the conditional sets and the optimal screening threshold, and implement a fast iterative algorithm. Simulations and applications to long non-coding RNA and microRNA regulation in Kidney cancer and DNA methylation regulation in Glioblastoma Multiforme illustrate the validity and advantage of our method.

Epigenetics Posters - Thursday

PB2435. High-resolution analysis identifies ancestry-specific chromatin 3D interactions

Authors:

W. Xu^{1,2}, L. Wang^{3,4}, X. Liu¹, O. Oron³, F. Rajabli^{3,4}, L. Lu¹, D. Dykxhoorn^{3,4}, A. Griswold^{3,4}, M. Pericak-Vance^{3,4}, J. Young^{3,4}, F. Jin^{1,5}, J. Vance^{3,4}; ¹Dept. of Genetics and Genome Sci., Sch. of Med., Case Western Reserve Univ., Cleveland, OH, ²The BioMed. Sci. Training Program (BSTP), Sch. of Med., Case Western Reserve Univ., Cleveland, OH, ³John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, ⁴John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, ⁵Dept. of Computer and Data Sci., Dept. of Population and Quantitative Hlth.Sci., Case Comprehensive Cancer Ctr., Case Western Reserve Univ., Cleveland, OH

Abstract Body:

Introduction: Alzheimer's disease (AD) has ancestry-specific risk. We and others have shown that the local ancestry (LA) block surrounding *ApoE* confers differential genetic risk for *ApoE* ε4 carriers in Non-Hispanic Whites (NHW) and African American (AA). However, the mechanism underlying the ancestry-specific AD risk remains unclear. 3D genome architecture plays an important role in transcriptional regulation via long-range chromatin looping. Genome-wide Hi-C analyses have revealed highly dynamic genome architecture during development and disease progression. The variation of 3D genome architecture on different ancestry background has not been characterized, especially at the level of chromatin loops, which are most likely representing enhancer-promoter interactions. Mapping chromatin loop is a major challenge in the 3D genome research. Due to the complex bias structure and severe data sparsity, it is generally believed that multibillion-read sequencing depth is required for kilobase (kb)-resolution loop calling. **We recently developed *DeepLoop* which allows robust mapping of chromatin loops with <10-fold Hi-C reads.** **Methods:** Chromatin Hi-C libraries were constructed with 4-cutter enzyme on astrocytes derived from induced pluripotent stem cells (iPSCs) from four individuals with different ancestry composition. *DeepLoop* was used to call chromatin loops at 5kb-resolution. LA at each locus across the genome was calculated using RFMix with SNP-array data. A reference panel with diverse populations from 1000 Genomes was used for European (EU), African (AF) and Amerindian (AI) ancestries. **Results:** On average, ~500 million pair-end reads were obtained for each library. Among the four samples, the proportion of AF ancestry (AF%), EU ancestry (EU%), and AI ancestry (AI%) ranges from 0.3% to 14.9%, 72.5% to 99.7%, and 0% to 13.6%, respectively. In 5 out of 6 pair-wise comparisons, we detected a significant enrichment of sample-specific chromatin loops in regions with different LA ($P < 0.05$, Fisher's exact test; two-side). Some genes residing in these ancestry-enriched chromatin loops are relevant to brain functions, e.g. *SLC1A5*. **Conclusion:** These early results revealed that genome architecture can be variable between different ancestry backgrounds, suggesting that LA is a major source for differential chromatin loops between individuals. Future efforts will explore if and how such differences may contribute to ancestry-specific AD risk.

Epigenetics Posters - Wednesday

PB2436. Human epi- and transcriptomic trajectories in the postprandial state.

Authors:

R. Costeira¹, L. D. Ruiz², L. Sinke³, C. Christiansen¹, Y. Raza¹, M. Tomlinson¹, B. Heijmans³, E. Slagboom³, K. Small¹, M. Waldenberger⁴, S. Berry¹, J. M. Ordovas⁵, J. T. Bell¹; ¹King's Coll. London, London, United Kingdom, ²IMDEA Food Inst., Madrid, Spain, ³Leiden Univ. Med. Ctr., Leiden, Netherlands, ⁴Helmholtz Zentrum München, Munich, Germany, ⁵Tufts Univ., Boston, MA

Abstract Body:

Background and objectives: The postprandial state, i.e., the period following a meal, is characterized by a highly individualized response to food that remains little explored at the molecular level. DNA methylation (DNAm) is a key regulator of gene function and has the potential to explain some of the inter-individual variation observed postprandially through changes in the activation and expression of genes. Within the JPI-DIMENSION consortium, we aimed to characterise the changing epi- and transcriptomic postprandial landscapes and elucidate the metabolic flexibility of humans in response to food.

Methods: To do this, whole blood EPIC DNAm profiles were obtained for altogether 270 participants of the ZOE PREDICT 1 (n = 120 individuals) and CORDIOPREV (n = 150 individuals) studies at three and two timepoints, respectively. Time-matched whole blood transcriptomes were further obtained for 50 participants of ZOE PREDICT 1 using Illumina RNA sequencing technologies. Epi- and transcriptome-wide association analyses were carried at >756,000 CpG-sites and >23,000 genes. First, glycemic (t_{30m}) and lipemic (t_{4h}) methylation and gene expression level changes were compared to fasting (t₀) in ZOE PREDICT 1. Second, peak lipemia DNAm level changes were meta-analysed between the ZOE PREDICT 1 and CORDIOPREV studies. Linear mixed effects models were used, and results are adjusted for age, sex, BMI, smoking, blood cell type proportions, technical and cohort-specific variables (FDR = 10%).

Results: Overall, 19 differentially methylated positions (DMPs) showed postprandial trajectory changes in ZOE PREDICT 1. The peak signal was in the *CAMTA1* gene, with decreased methylation at 30m after test meal (cg23021268; $b=0.311\pm0.050$, $p=2.40E-08$). ZOE PREDICT 1 and CORDIOPREV meta-analysis identified 125 DMPs across 94 genes that changed postprandially at 4h, including in the cholesterol efflux gene *ABCG1* (cg27518648: $b=0.078\pm0.016$, $p=1.11E-06$). Pathway analysis identified enrichment in the PKA pathway, as well as in pathways of cardiac and metabolic signaling. Preliminary transcriptomic results identified gene expression changes in ZOE PREDICT 1 over the 3 timepoints analysed. Ongoing analyses focus on exploring the gene expression data in the context of linking specific expression changes to the DNAm trajectories observed.

Conclusions: In summary, we observe significant changes in DNAm and gene expression at target genomic regions postprandially. Signals were observed in metabolically relevant genes that help elucidating the link between the individual response to food and the development of cardiovascular disease.

Epigenetics Posters - Thursday

PB2437. Human gene regulatory evolution is driven by divergence in both *cis* and *trans*.

Authors:

T. Hansen¹, S. Fong², J. Capra², E. Hodges¹; ¹Vanderbilt Univ., Nashville, TN, ²Univ. of California San Francisco, San Francisco, CA

Abstract Body:

Phenotypic differences between closely related species are frequently driven by variation in gene expression patterns. At the transcriptional level, regulatory differences that drive differential gene expression can occur in *cis*-, that is mutations that affect DNA regulatory element sequence, or in *trans*-, mutations that affect the cellular environment (e.g., transcription factor expression levels). However, to date studies investigating differential gene regulatory function typically have not been able to distinguish the effects of *cis*- and *trans*-regulation. Since a *trans*-effect can offset a *cis*-effect and visa-versa, it is critical to analyze the two effects independently to understand the mode of gene regulatory evolution. To address this challenge, we applied ATAC-STARR-seq to investigate divergent enhancer activity, chromatin accessibility, and transcription factor (TF) binding between human and rhesus macaque lymphoblastoid cell lines (LCLs). The modular design of the ATAC-STARR-seq assay allows us to decouple each species' genome sequence from its cellular environment and test the activity of all accessible human and macaque sequences in both human and macaque cellular environments. This enables global identification of species-specific regulatory regions driven exclusively by either *cis*- or *trans*-regulatory differences.

We discovered that *cis* and *trans* effects on gene regulatory activity occur at similar frequencies among 29,531 shared accessible regions. This strongly contrasts with the current presumption based on small-scale studies that the majority of gene regulatory divergence occurs in *cis*. Furthermore, we find that *cis* and *trans* effects most often affect the same regulatory element, suggesting they cooperate to drive most of the divergent gene regulation between human and macaque LCLs. Leveraging TF footprinting, we observe that ~57% of human specific *trans* effects are explained by differential expression of key TFs, including IRF family members and NF- κ B, which alters their availability in the respective cellular environments. By contrast, *cis* effects are enriched for regions with accelerated substitution rates characteristic of positive selection, indicating DNA sequence divergence at those loci is functionally important for driving human evolution.

Altogether, our high-resolution genome-wide study reveals that evolutionary divergence in the *trans*-regulatory environment drives a substantial degree of the gene regulatory differences between human and macaque LCLs, and these observations highlight an underappreciated role for *trans*-regulatory divergence in human evolution.

Epigenetics Posters - Wednesday

PB2438. Identification of diagnostic DNA methylation epesignatures associated with large structural chromosomal variations in patients with neurodevelopmental disorders.

Authors:

K. Rooney¹, R. Relator¹, M. Levy¹, H. McConkey¹, J. Kerkhof², M. Alders³, P. Henneman⁴, G. Cappuccio⁵, N. Brunetti-Pierra⁵, M. Niceta⁶, M. Tartaglia⁶, P. Prontera⁷, G. Merla⁸, A. Brusco⁹, G. B. Ferrero⁹, M. Tedder¹⁰, B. DuPont¹⁰, B. Sadikovic¹; ¹London Hlth.Sci. Ctr., London, ON, Canada, ²London Hlth.Sci. Ctr., London, ON, Canada, ³Amsterdam UMC, Amsterdam, Netherlands, ⁴Univ. of Amsterdam, Academic Med. Ctr., Amsterdam, Netherlands, ⁵Federico II Univ., Naples, Italy, ⁶Bambino Gesù Children's Hosp., Rome, Italy, ⁷Univ. and Hosp. of Perugia, Perugia, Italy, ⁸Univ. of Naples Federico II, Naples, Italy, ⁹Univ. of Torino, Torino, Italy, ¹⁰Greenwood Genetic Ctr., Greenwood, SC

Abstract Body:

Background: An expanding number of neurodevelopmental disorders (NDDs), associated with genetic alterations, have demonstrated distinct changes in the DNA methylation profile. These epigenetic changes, called epesignatures, are highly consistent and specific amongst individuals affected by the same underlying genetic aberration. This study assesses DNA methylation profiles in cohorts of individuals with more than 20 common recurrent copy number variants (CNVs) associated with genomic disorders. **Methods:** Genome-wide DNA methylation analysis was performed on peripheral blood samples from cohorts of patients with known CNV disorders using the Illumina 850K BeadChip array. Each cohort was assessed for methylation changes against a subset of age and sex matched controls. Methylation levels for each CpG probe were measured and 100-500 differentially methylated probes were used to define the disorder epesignature. The robustness and sensitivity of the selected probes were tested using clustering and cross validation methods. A support vector machine learning model was constructed to confirm the specificity for the epesignature to classify each disorder. **Results:** Patients with CNVs, associated with specific genomic disorders, exhibit distinct clinical epesignatures that can be differentiated from each other and from other NDDs with established epesignatures. These epesignatures are highly sensitive and specific biomarkers that can be used as a clinical diagnostic screening test. Assessment of breakpoints provides evidence of epigenetic regulatory genes and regions in some of the CNV disorders. **Conclusion:** We mapped epesignatures in CNV disorders that provide a specific, differentiating biomarker for these conditions. These epesignatures have the potential to improve the diagnostic yield of clinical testing in patients with NDDs, and future work to assess molecular insights into disease pathology through differentially methylated regions is underway.

Epigenetics Posters - Thursday

PB2439. Identifying functional interactors of conserved meiotic histone reader ZCWPW1

Authors:

D. Bazzano¹, W. Xie¹, C. Zuckerman¹, J. Nandakumar¹, F. Cole², S. Hammoud¹; ¹Univ. of Michigan, Ann Arbor, MI, ²MD Anderson Cancer Ctr., Univ. of Texas, Houston, TX

Abstract Body:

Spermatogenesis is a highly regulated process necessary for the generation of haploid gametes that transmit genetic and epigenetic information from one generation to the next. Mature spermatid differentiation relies on a specialized form of cell division called meiosis, where homologous chromosomes from a diploid genome pair together, recombine genetic information through induced DNA double-strand breaks (DSBs) and separate into genetically distinct haploid daughter cells. Homologous recombination is tightly controlled, since defects in the DSB repair process result in de novo mutations, large structural variations and aneuploidies. The sites of homologous recombination in mammals, called crossovers, are determined by PRDM9, a zinc-finger H3K4/K36 methyltransferase. Although PRDM9 is recruited to thousands of sites in the human and mouse genome, only 200-300 sites recruit downstream DSB repair factors and ~20 of these sites are resolved as crossovers. Currently, the mechanism that link sites of H3K4/K36 trimethylation with recombination is largely unknown. Preliminary data from our lab and others has shown that a H3K4/K36 reader expressed in meiotic prophase, known as ZCWPW1, has defects in homologous chromosome pairing and recombination. Despite these characterizations, the mechanism of homolog engagement and DSB repair promoted by ZCWPW1 remains unsolved. Here, we propose that ZCWPW1 can mediate pairing by reducing the homologue search grid in the nucleus and bridging homologous chromosomes together prior to DSB formation through self-interaction. Another non-mutually exclusive possibility is that ZCWPW1 interacts with structural components or DSB repair factors to promote crossover formation and stabilize interhomolog contacts. To test these ideas, we developed genetic and biochemical approaches to determine whether ZCWPW1 can interact with itself and/or other proteins involved in meiotic DSB repair. These approaches include co-immunoprecipitation experiments in HEK293T cells, yeast two-hybrid assay and purification of recombinant protein to perform IP pull down experiments and SEC-MALS. Together, these experiments will provide new insight into the role of ZCWPW1 during crossover formation and provide further clarity into the mechanisms of homolog pairing and engagement in meiosis. The knowledge gained from this project will further characterize the conserved epigenetic mechanisms underlying heredity, chromosome segregation and genomic fidelity in the germline.

Epigenetics Posters - Wednesday

PB2440. Immune cell-type level epigenome-wide association analysis on chronic HIV infection highlights hallmark genes for HIV pathogenesis and cancer

Authors:

X. Zhang¹, Y. Hu², A. Justice³, C. Yan², B. Aouizerat⁴, K. Xu⁵; ¹Yale Univ., New Haven, CT, ²NCI, Rockville, MD, ³Yale Sch. of Medicine, New Haven, CT, ⁴New York Univ., New York, NY, ⁵Yale Sch. Med., New Haven, CT

Abstract Body:

Background: Previous epigenome-wide association studies (EWAS) have identified CpG sites associated with chronic HIV infection from the heterogeneous blood cells or from CD4+ T cells, which offer the limited knowledge of cell type-specific methylation aberrant in other immune cells for HIV infection. **Methods:** Applying a deconvolution method that is validated by capture methylation sequencing, we conducted a cell-type based EWAS and identified DNA methylation aberrant for chronic HIV infection among five immune cell types in blood: CD4+, CD8+, B, Natural Killer (NK), and Monocytes in a cohort: Veteran Aging Cohort Study (VACS) (N=718). The VACS cohort included HIV-negative and HIV-positive participants who were on antiviral therapy. **Results:** A cell-type level EWAS revealed the distinct patterns of HIV-associated methylation CpG sites with 67.9% of CpG sites unique to individual cell type (False discovery rate, FDR <0.05). We identified 2,212 CpG sites in CD4+ T cells and 21 CpG sites in monocytes for HIV infection. One of the top significant CpG, cg16411857, on *NLRC5* showed a large effect size in CD4+, B, and monocytes. Significant CpG sites were located on the genes that are involved in immunity and HIV pathogenesis (e.g. *RUNX3* in CD4+, *STAT3* in B, *IL7* in NK, *LCK* in monocyte). More importantly, HIV-associated CpG sites were overrepresented on hallmark genes involved in cancer pathology (q<0.05) (e.g. *BCL* family, *STAT3*, *PRDM16*, *ESR1*, *DNMT3A*). HIV-associated CpG sites are enriched on the pathways involved in HIV pathogenesis and carcinogenesis such as Kras-signaling and apoptotic pathways. **Conclusions:** Our findings provide new insights on cell type specific epigenetic modifications in the HIV infected host genome that support a notion of pathogen-induced epigenetic cancerization and further elucidate mechanisms of HIV and its comorbid with cancer.

Epigenetics Posters - Thursday

PB2441*. Investigating aberrant DNA methylation in pediatric epilepsies

Authors:

C. LaFlamme¹, E. Almanza Fuerte¹, S. Russ-Hall², A. Schneider², University of Washington Center for Mendelian Genomics, D. Miller³, M. Galey³, Z. Wang¹, B. Sadikovic⁴, L. Sadleir⁵, S. Berkovic², I. Scheffer², H. Mefford¹; ¹St. Jude Children's Res. Hosp., Memphis, TN, ²Univ. of Melbourne, Melbourne, Australia, ³Univ. of Washington, Seattle, WA, ⁴LHSC, London, ON, Canada, ⁵Univ. of Otago, Dunedin, New Zealand

Abstract Body:

Developmental and epileptic encephalopathies (DEEs) are a group of disorders characterized by drug-resistant seizures and developmental slowing or regression with epileptiform activity on EEG. Advances in sequencing technology have enabled Mendelian genetic diagnosis in ~50% of patients with the potential for developing targeted therapies. Those without known pathogenic DNA variants are deemed “unsolved.” DNA methylation is an important epigenetic modification implicated in various neurodevelopmental disorders, such as Fragile X syndrome. Rare differentially methylated regions (DMRs) affect gene expression and can be influenced by DNA sequence variations, such as GC-rich repeat expansions impossible to detect by standard sequencing modalities, but have not been studied in DEEs. Additionally, as a part of EpiSign analysis, distinct genome-wide “episignatures” have been derived for ~120 monogenic neurodevelopmental disorders, including >20 genetic epilepsy syndromes and DEEs. Episignatures indicate the presence of a pathogenic variant in a specific gene. Thus, a patient with unsolved DEE could display an episignature informing the gene to examine for missed or misinterpreted coding and noncoding variants. The diagnostic utility of episignature analysis for unsolved DEE is not known. We hypothesize that DNA methylation analysis can be used to uncover novel causes of DEE by (i) investigating rare DMRs as an alternative disease mechanism and (ii) using recently established methylation episignatures of monogenic disorders to inform the identification of previously hidden, novel pathogenic variants. We generated Illumina EPIC methylation array data for >400 patients with unsolved DEEs, identified rare DMRs using a robust outlier approach, and performed episignature analysis. We have validated 15 DMRs to date using targeted bisulfite sequencing and nanopore long-read sequencing. We then identified potential underlying DNA defects of DMRs and candidate pathogenic variants associated with episignatures using a combination of short and long-read sequencing. We find that patients with unsolved DEEs harbor rare outlier DMRs and episignatures that inform potentially disease causative DNA variants. To investigate the effects of DMRs and candidate pathogenic variants on gene expression, we will perform RNA-seq of patient-derived or genetically engineered cell lines. Our findings will provide insights into novel epigenetic mechanisms underlying DEEs, facilitate molecular diagnoses of patients with unsolved DEE, and potentially lead to improved diagnostics and the discovery of novel therapeutic targets for these devastating disorders.

Epigenetics Posters - Wednesday

PB2442. Investigation of the effects of alcohol exposure on the chromatin binding and localization of cohesin and CTCF.

Authors:

C. Cummings, M. Rowley; Univ. of Nebraska Med. Ctr., Omaha, NE

Abstract Body:

Alcohol exposure is associated with a wide array of medical problems, including both acquired and congenital phenotypes. Examples of these phenotypes include increased incidence of multiple cancer subtypes and diagnoses under the fetal alcohol spectrum disorders umbrella. While much work has attempted to elucidate the downstream molecular mechanisms induced by ethanol exposure, there are large gaps in our understanding. One promising question is how ethanol exposure affects epigenetics, including the impact of ethanol on DNA methylation, histone modifications, and chromatin organization. The three-dimensional structure of chromatin is formed by multiple co-occurring levels of genome organization and is impacted by architectural proteins and chromatin modifications. Patients with genetic defects in the architectural proteins CTCF or cohesin present with overlapping syndromes characterized primarily by a neurodevelopmental disorder with growth defects and to a lesser degree multiple congenital anomalies. In cell culture models of ethanol exposure, we performed ChIP-Seq experiments and found alterations to the chromatin binding of cohesin. This indicates alteration of insulator proteins and may suggest changes to the regional chromatin organization and effects on local gene regulation. In summary, ethanol treatment alters the localization of the architectural protein cohesin on chromatin. This likely influences nearby chromatin organization and may correspond to the observed downstream transcription changes known to be induced by ethanol treatment. We are currently evaluating the dose sensitivity and persistence of these observed effects, and their molecular mediators. Together, this data reveals a previously unknown mechanistic-functional explanation of ethanol exposure phenotypes, indicating that the phenotypes may be due in part to disrupted chromatin organization. Intriguingly, this work may provide a link between the well-described cohesinopathy phenotypes and fetal alcohol spectrum disorders, explaining their phenotypic similarities.

Epigenetics Posters - Thursday

PB2443. Investigation of Tissue Level of Molecular Pathways Effective in Vascular Calcification at Chronic Renal Insufficiency Patients

Authors:

E. Arslan, P. Ata; Marmara Üniversitesi, Istanbul, Turkey

Abstract Body:

Objective: Chronic renal insufficiency (CRI) is a public problem affecting more than two million people. Vascular calcification (VC), emerging from pathological mineral deposition at medial and intimal parts of the vessels, affects morbidity and mortality. Although it is an end result of normal aging process, VC emerges from pathological processes effective in CRI. **Material and Methods:** Herein we investigated hsa-miR-30a-5p and hsa-miR-223-3p with RNA PCR kit at waste vascular tissue taken from 37 patients during renal transplantation operation. For expression levels of target miRNA analysis qRT-PCR technique was used in Biorad Real time PCR platform. **Results:** There were 8 females and 29 male patients with an average age of 36 +/- 14 years of age. For normalization of expression levels hsa-miR-145-5p was used as internal control which is expressed at vascular tissue at physiological condition. **Conclusion:** Hsa-miR-223-3p is an important miRNA at inhibition of (*IL-6*)/*STAT3* signaling pathway that is effective in induction of vascular calcification. Hsa-miR-30a-5p targeting *RUNX2* signalling effects osteoblastic differentiation. We have detected that both Hsa-miR-223-3p and Hsa-miR-30a-5p expressions were increased in vascular tissue. **Keywords:** Chronic renal insufficiency, vascular calcification, epigenetic control of gene expression.

Epigenetics Posters - Wednesday

PB2444. KDM6A/B inhibition rescues an osteoblast phenotype and altered gene expression in a novel mouse model of Weaver Syndrome with skeletal overgrowth.

Authors:

J. Fahrner¹, W-Y. Lin¹, L. Boukas², C. W. Gao³, P. Kushwaha¹, B. Spiegelberg¹, H. T. Bjornsson⁴, R. C. Riddle¹, K. D. Hansen⁵; ¹Johns Hopkins Sch. of Med., Baltimore, MD, ²Johns Hopkins Univ., Baltimore, MD, ³Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ⁴Univ. of Iceland, Reykjavik, Iceland, ⁵Johns Hopkins Univ, Baltimore, MD

Abstract Body:

Mendelian disorders of the epigenetic machinery (MDEMs) result from germ-line mutations in genes encoding chromatin modifying machinery. They are characterized by intellectual disability and abnormal growth, which may manifest as growth retardation or overgrowth. Weaver syndrome (WS) is a classic overgrowth MDEM and exhibits striking tall stature, mild to moderate intellectual disability, and characteristic facial features. WS results from heterozygous, typically *de novo* variants in *EZH2*, which encodes a histone methyltransferase writer that methylates lysine 27 on histone H3 (H3K27me) as part of the Polycomb repressive complex 2 (PRC2). Most pathogenic variants in *EZH2* are hypomorphic missense changes. Others' work revealed that homozygous ablation of *Ezh2* in mice led to embryonic lethality, whereas heterozygotes had no apparent phenotype. Conditional models have not fully recapitulated the WS phenotype; however, a recent constitutional knock-in mouse model exhibited mild overgrowth. We used CRISPR gene editing and homology-directed repair to generate a mouse model of WS with the most common causative variant, R684C. WS mice (*Ezh2*^{R684C/+}) exhibit overgrowth, reminiscent of the human phenotype. Specifically, high-resolution micro-CT of long bones reveals skeletal overgrowth, and *in vivo* labeling reveals increased osteoblast activity in *Ezh2*^{R684C/+} compared to *Ezh2*^{+/+}. Staining of osteoblasts after isolation of precursors from mouse bone marrow and differentiation *in vitro* confirms increased bone matrix deposition in *Ezh2*^{R684C/+} compared to *Ezh2*^{+/+}. Genome-wide transcriptome profiling of *Ezh2*^{R684C/+} and *Ezh2*^{+/+} osteoblasts with RNA-seq reveals differential expression of known PRC2 targets. Most notably, we show extensive overexpression of genes involved in bone morphogenetic protein (BMP) signaling, which is important in osteogenesis. *Ezh2*^{R684C/R684C} mouse embryonic fibroblasts show global loss of H3K27me3, confirming the variant as hypomorphic. Notably, counteracting the deficient EZH2 writer of H3K27me3 by inhibiting the opposing eraser using a KDM6A/B inhibitor rescued the osteoblast phenotype and partially corrected expression of select BMP pathway genes. Our findings establish a role for altered BMP signaling and H3K27me3 in WS overgrowth and, importantly, point to KDM6A/B inhibition as a potential targeted treatment for WS. Ongoing ChIP-seq studies will further define the precise role of H3K27me3, and planned preclinical therapeutic trials with this epigenetic drug may pave the way for the first treatment for individuals with WS, as well as for related MDEMs and more common forms of overgrowth.

Epigenetics Posters - Thursday

PB2445. Learning a generalized regulatory model from paired chromatin accessibility and transcriptome

Authors:

L. Xiong^{1,2}, M. Kellis^{1,2}, ¹Computer Sci. and Artificial Intelligence Lab., Massachusetts Inst. of Technology, Cambridge, MA, ²Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract Body:

The fundamental mechanisms that diversify the cells into various cell types and generate cell-type specific gene expression from the same DNA sequences remains unclear. Recent multi-modal single-cell sequencing technologies jointly profiling chromatin accessibility and transcriptome in the same cell enable us to investigate the regulatory mechanisms that link the chromatin change and gene expression in single cell resolution. However, the searching space of the regulatory relationships from millions of chromatin accessible regions and tens of thousands of genes is huge and complex. Hard to distinguish the missing signals in scATAC-seq data between biological ones or technical issues makes the question even more challenging due to the sparse nature of scATAC-seq data. Variations from cell to cell and batch effects existing in both scATAC-seq and scRNA-seq are also troublesome confounding factors. Currently, researchers are calculating gene activity score based-on distance using single-cell ATAC-seq data to predict the gene transcriptome level. However, there is no standard gene regulatory model that adopts all of the genes and situations vary a lot from gene to gene. Here, we developed a deep learning method to by building a generalized regulatory model that decodes the cis-regulatory code of gene expression and allows to accurately predict expression of tens of thousands of genes for each cell using epigenetic information. First, we trained our model on million-sized multi-omics ADRD datasets and a million-sized public scATAC-seq human atlas dataset to obtain a generalized model among multiple cell types across various tissues. Second, we identified cell subpopulations, gene modules and chromatin accessible elements modules and identified their connections on the module level by clustering and visualized the learned representations for cell, gene and chromatin accessible regions in the UMAP embeddings. Third, we tried to interpret the cis-regulatory code of each gene by identify the most contributing regulatory elements in the gene neighborhood, and constructed a regulatory network based-on these regulatory elements and their target genes. Fourth, we expected to predict the outcome of the in silico perturbation based-on the regulatory network and validate them with perturbation experiments.

Epigenetics Posters - Wednesday

PB2446. Learning the cis and trans sequence features of gene regulatory divergence between human and rhesus macaque

Authors:

S. Fong¹, T. Hansen¹, J. Capra², E. Hodges³; ¹Vanderbilt Univ., Nashville, TN, ²Univ. of California San Francisco, San Francisco, CA, ³Vanderbilt Univ. Sch. of Med., Nashville, TN

Abstract Body:

Motivation: The evolution of non-coding gene regulatory elements is a major driver of species divergence, yet the mechanisms underlying this evolution genome-wide are poorly understood. Recently, we systematically categorized gene regulatory activity divergence between closely related species due to *cis* changes—mutations that directly affect the regulatory potential of an individual element and the expression of its target gene—or in *trans*—changes that alter the cellular environment (e.g., transcription factor abundance) and expression of multiple genes. We found that divergent elements commonly require both species-specific *cis*- and *trans*-variation for activity in human and rhesus macaque lymphoblastoid cell models. However, the underlying sequence determinants that produce species-specific *cis*- and *trans*-activity are unknown.

Results: Here, we investigate sequence features that distinguish human-specific *cis*- from *trans*- gene regulatory activity generated by massively parallel ATAC-STARR-seq reporter assays. First, we develop two supervised machine learning classifiers to separately model the sequence features underlying *cis* or *trans* gene regulatory activity divergence between humans and rhesus macaques, and validate model predictions with differential transcription factor footprinting analyses. Then, we apply these evolutionary classifiers to investigate elements with both phenotypic *cis*- and *trans*- divergent features and dissect the contributions of *cis*-like or *trans*-like sequence features to divergent activity.

Conclusions: Our work rigorously models the sequence features and associated mechanisms underlying gene regulatory divergence between human and macaque genome-wide. Highlighting the sensitivity of these sequences to both genetic and cellular environment changes, we provide valuable insights into the molecular mechanisms that have produced species-specific gene regulatory activity and have contributed to the divergence of humans and rhesus macaques.

Epigenetics Posters - Thursday

PB2447. Linking genetic variants associated with gluteofemoral fat storage to determine their role in adipogenesis

Authors:

J. Belanich¹, C. Tomaselli², A. Hecq³, M. G. Urena¹, B. Emanuelli¹, T. O. Kilpeläinen¹; ¹Univ. of Copenhagen, København, Denmark, ²Pompeu Fabra Univ., Barcelona, Spain, ³Univ. of Mons, Mons, Belgium

Abstract Body:

While obesity as a whole is linked to an increased risk of type 2 diabetes, this blurs and disregards individual body fat distribution, which plays a role in individual risk and health outcomes. By examining the physical location of adipose tissue, gluteofemoral fat has been found to be protective, preventing fat deposition in critical areas such as visceral, intra-hepatic, and intra-pancreatic fat depots. Additionally, there have been several studies that have indicated an inverted linkage; those individuals who express decreased levels of gluteofemoral fat show increased insulin resistance. With this, we aim to identify and experimentally follow up on SNPs in order to explain the mechanistic actions behind the protective gluteofemoral fat. The overall goal is to further understand and parse the mechanisms underlying lower gluteofemoral fat and higher risk of cardiometabolic disease, independent of BMI. We took forward all genome-wide significant SNPs associated with WHR or WHR adjusted for BMI in the largest GWAS published to date, and identified SNPs that are associated with lower BMI but higher WHR. We prioritized candidate causal genes in the corresponding genetic regions, and assessed the genes' roles in differentiation and maturation of mouse and human preadipocytes by microscopy and quantitative methods. Altogether 252 independent loci showed an association with lower BMI but higher WHR. We prioritized six candidate causal genes for initial perturbation studies using siRNA. The studies showed an inhibition of adipogenesis for four of the six tested genes: ADAMTS9, ABHD15, EMILIN2 and COL18A1. Following this, small interfering RNA (siRNAs) were used in mouse and human pre-adipocyte cultures to knockdown the expression of the discovered target genes. The cultures were differentiated into white adipocytes, and were examined at various stages of differentiation in order to better examine the genes' roles in growth, lipid deposition, and how they affect the expression of other genes. In further experiments, we are determining the effects on lipid metabolism and insulin signaling through downregulation and dCas9 overexpression, and will examine the resulting expression profile data. Through this, we hope to explain the paradoxical findings on genetic variants associated with a leaner body weight yet an unfavorable body fat distribution and cardiometabolic risk profile, and begin to identify new targets for drug development and treatments.

Epigenetics Posters - Wednesday

PB2448*. Loss of DNMT3A or TET2 activity in stimulated macrophages alters transcription factor binding and enhances inflammatory gene expression.

Authors:

K. B. Rodrigues, J. Gopakumar, M. Maurer, Z. Weng, T. Eulalio, S. Mitchell, D. Nachun, D. Estrada, T. Mazumder, L. Ma, S. Montgomery, S. Jaiswal; Stanford Univ., Stanford, CA

Abstract Body:

A source of increased age-associated inflammation is Clonal Hematopoiesis of Indeterminate Potential (CHIP). CHIP occurs when a somatic mutation arises in a bone marrow stem cell and promotes clonal expansion of that stem cell and its progeny. Furthermore, the function of macrophages derived from these bone marrow stem cells is thought to be affected by CHIP mutations, possibly explaining the increased risk of cardiovascular diseases in people with CHIP. Strangely, the mutations that cause CHIP most frequently occur in genes with opposite function: loss-of-function mutations in *DNMT3A* and *TET2*. These enzymes methylate and demethylate DNA, respectively. To evaluate how DNMT3A and TET2 contribute to inflammatory responses in the context of CHIP, we exposed bone marrow derived macrophages from *Dnmt3a* KO, *Tet2* KO, and wild type (WT) control mice to a low dose of the inflammatory stimuli lipopolysaccharide (LPS). We then measured gene responses with RNA-seq across 12 time points from 0 to 24 hours, paired with ATAC-seq and whole genome methyl-seq for multiple time points. A key subset of inflammatory genes (including *Il1b*, *Il6*, and *Il23a*) were expressed around twice that of WT in both KOs by 2 hours of stimulation, despite the opposing activity of the enzymes on DNA methylation. Further, DNA methylation changed in WT upon stimulation in cis-regulatory regions of many of these inflammatory genes. Additionally, KO of *Dnmt3a* or *Tet2* affected DNA methylation both before and during stimulation. DNA footprinting and motif enrichment analysis from ATAC-seq similarly showed that the binding of transcription factors before and during stimulation was affected by loss of either enzyme. This supports a mechanism whereby the activity of DNMT3a and TET2 alters DNA methylation, modifying binding of nearby transcription factors. Our integrative multi-omic analysis demonstrates how DNMT3A and TET2 modulate DNA methylation to limit inflammatory gene expression in healthy macrophages. Targeting these sites of inflammation-associated DNA methylation defects is a suggested therapeutic avenue for blocking the pathogenic role of CHIP for many inflammatory diseases of aging.

Epigenetics Posters - Thursday

PB2449*. Mapping chromatin accessibility QTL in 138 liver tissue samples identifies coordinated regulation, links regulatory elements to genes, and predicts mechanisms at GWAS loci.

Authors:

K. Currin¹, H. J. Perrin¹, A. S. Etheridge^{1,2}, S. Vadlamudi¹, A. A. Alkhawaja¹, K. A. Broadaway¹, J. D. Rosen¹, A. S. Chaudhry³, P. J. Gallins⁴, Y. Zhou⁴, F. Wright⁴, E. G. Schuetz³, F. Innocenti⁵, K. L. Mohlke¹; ¹Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, ²Eshelman Sch. of Pharmacy and Ctr. for Pharmacogenomics and Individualized Therapy, Univ. of North Carolina, Chapel Hill, NC, ³Dept. of Pharmaceutical Sci., St. Jude Children's Res. Hosp., Memphis, TN, ⁴Bioinformatics Res. Ctr., North Carolina State Univ., Raleigh, NC, ⁵AbbVie Inc., South San Francisco, CA

Abstract Body:

Chromatin accessibility maps have been used to identify regulatory elements and plausible molecular mechanisms at GWAS loci. However, chromatin accessibility maps for many tissues are based on one or a few individuals, which limits identification of regulatory elements that vary across individuals due to genetic and environmental factors. To address this limitation, we profiled chromatin accessibility using ATAC-seq in frozen liver tissue from 138 individuals. We identified 358,304 consensus accessible chromatin regions (peaks) present in at least 5% of individuals. To identify genetic effects on chromatin accessibility, we mapped chromatin accessibility QTL (caQTL) with FastQTL using variants within 100 kb of peak centers and identified 27,289 peaks with a caQTL (caPeaks, FDR<5%). Using colocalization (linkage disequilibrium (LD) $r^2 \geq 0.8$) between caQTL and existing liver eQTL (1,181 individuals), we linked 3,182 caPeaks to 2,032 genes. We also identified dozens of GWAS-caQTL colocalizations, including 122 for total cholesterol, 112 for body mass index, 27 for coronary artery disease, and 45 for type 2 diabetes. We observed widespread coordinated genetic regulation of caPeaks, with 7,988 caPeaks sharing a genetic signal (LD $r^2 \geq 0.9$) with at least one other caPeak. Most of the 2,941 coordinated caPeak sets (69%) consist of 2 peaks, and 238 (8%) of sets contain 5 or more peaks. 607 coordinated caPeak sets contained both promoter and distal peaks, which can be used to link peaks to target genes. A large set of 23 coordinated caPeaks near *SORT1* is associated with rs12740374, previously identified as an eQTL for *SORT1*, *CELSR2*, and *PSRC1*, and a well-characterized GWAS variant for cardiovascular traits. This set of 23 peaks includes a peak at both the *SORT1* and *MYBPHL* promoters, providing evidence for coordinate regulation of multiple peaks and genes at this locus. At another locus, a caQTL signal is colocalized with a GWAS signal for total cholesterol and an eQTL for *MICAL2*. This caQTL signal contains 3 coordinated peaks, including a peak at the *MICAL2* promoter, providing multiple lines of evidence supporting *MICAL2* as the target gene and implicating 3 peaks in its regulation. We and others have previously shown that caQTL often disrupt transcription factor (TF) motifs, suggesting that many caQTL may result from altered TF binding. Our results demonstrate that mapping chromatin accessibility in increased sample sizes identifies widespread genetic effects on chromatin accessibility and that these data can be combined with eQTL and GWAS to predict variants that alter chromatin accessibility leading to effects on gene expression and disease-relevant traits.

Epigenetics Posters - Wednesday

PB2450. Methylation-based machine learning classification of osteoarthritis in the Multicenter Osteoarthritis Study.

Authors:

P. Okoro¹, I. K. Haugen², J. A. Lynch³, M. Nevitt³, C. E. Lewis⁴, J. Torner⁵, D. T. Felson⁶, M. S. Yau^{1,7}; ¹Hebrew SeniorLife - Harvard Med. Sch., Roslindale, MA, ²Ctr. for treatment of Rheumatic and Musculoskeletal Diseases (REMEDY), Diakonhjemmet Hosp., Oslo, Norway, ³Dept. of Epidemiology and Biostatistics, Univ. of California San Francisco, San Francisco, CA, ⁴Dept. of Epidemiology, Univ. of Alabama, Birmingham, AL, ⁵Dept. of Epidemiology, Univ. of Iowa, Iowa City, IA, ⁶Section of Rheumatology, Dept. of Med., Boston Univ. Sch. of Med., Boston, MA, ⁷Div. of Gerontology, Dept. of Med., Beth Israel Deaconess Med. Ctr., Boston, MA

Abstract Body:

Purpose: Osteoarthritis (OA) is the most common form of arthritis and a leading cause of physical disability in older individuals. Currently, there are no biomarkers available to assess OA disease status. This may be attributed to the highly heterogeneous nature of OA and possible underlying OA endotypes. Epigenetic mechanisms like DNA methylation may play an important role in OA pathophysiology. For example, expression of *GDF5*, a well-known OA gene, is modulated by demethylation of the 5' untranslated region. There may be other loci under epigenetic regulation that could provide mechanistic insights into OA. Thus, we aim to build methylation-based joint-specific and multi-joint OA case classifiers using machine learning.

Methods: We analyzed data from a subset of participants in the Multicenter Osteoarthritis Study (MOST) who had bilateral hand and knee x-rays and methylation array data available (n=671). Hand OA was defined as the presence of 1) three or more hand joints with Kellgren-Lawrence (KL) grade ≥ 2 where at least two are part of the same DIP, PIP, or CMC1 joint group, 2) KL grade ≥ 2 in at least one DIP, and 3) bilateral involvement in ≥ 1 of the DIP, PIP, or CMC1 joint groups. Knee OA was defined as KL grade ≥ 2 in one or both knees. Multi-joint OA was defined as the presence of both hand and knee OA. About 23% had hand OA, 25% had knee OA, and 8% had multi-joint OA. DNA methylation was measured on peripheral whole blood with the Illumina Infinium MethylationEPIC 850K array. We removed poor quality probes and conducted quantile normalization. We also removed probes overlapping SNPs, cross-reactive probes, and probes on sex chromosomes, resulting in 774,460 CpG probes. The CpG methylation beta-values were log₂ transformed to M-values for modeling. We developed Elastic net-penalized classification models on the final datasets. Following 80%-training, 20%-testing data split, we carried out 10-fold cross validation on the training data; optimized models were applied on the test data. Model performance was evaluated using receiver operator characteristic area under the curve (AUC-ROC).

Results: The methylation profiles classified OA with AUC-ROC \pm SD as follows: Hand (Train-set=0.57 \pm 0.09; Test-set=0.61), Knee (Train-set=0.52 \pm 0.15; Test-set=0.46), Multi-joint (Train-set=0.64 \pm 0.19; Test-set=0.63).

Conclusion: Our models performed better with multi-joint OA, underscoring its possible systemic etiology, which may be better captured with blood DNA methylation biomarkers. While models performed better with multi-joint joint OA, AUCs were generally low. We plan to further improve and validate these models in other independent cohorts.

Epigenetics Posters - Thursday

PB2451*. MIMOSA: A method for improved methylome imputation increases power to identify CpG site-phenotype associations

Authors:

H. Melton¹, C. Wu¹, Z. Zhang¹, L. Wu²; ¹Florida State Univ., Tallahassee, FL, ²Univ. of Hawaii Cancer Ctr., Honolulu, HI

Abstract Body:

Inter-individual variation in DNA methylation (DNAm) has been linked to many complex traits through conventional epidemiological studies. However, its role in disease mechanisms is not yet well-characterized, primarily due to several common biases encountered in observational studies, including selection bias, unmeasured confounding factors, and reverse causation. One strategy to decrease such limitations is employing methylome-wide association studies (MWAS), one type of instrumental variable regression that involves two steps. First, MWAS builds DNA methylation prediction models for each CpG site using a reference panel with individual-level genetic and DNA methylation data. Second, MWAS tests the association between predicted DNA methylation levels and traits of interest. While appealing, current MWAS methods rely on an individual-level reference panel to build DNAm prediction models, thus limiting statistical power due to the limited sample size of these reference panels. To fill this gap, we introduce the Method to Impute Methylome Obliging Summary-level mQTLs and Annotation information (MIMOSA), a novel MWAS approach that unifies summary-level DNAm reference panels and functional annotation information with trait-specific genome-wide association study (GWAS) results to pinpoint CpG sites that are likely causally associated with the trait of interest. Briefly, we build DNAm imputation models in blood for each CpG site using summary-level mQTL data ($n=27,750$ individuals) from the Genetics of DNA Methylation Consortium (GoDMC). Likely due to sample size limitations in individual-level mQTL data, only imputation models with $R^2 > 0.01$ are typically included in MWAS. In contrast, MIMOSA further includes imputation models with $0.005 < R^2 < 0.01$, corresponding to CpG sites with lower heritability. To build more accurate DNAm imputation models, we further include functional annotations from the Functional Annotation of Variants - Online Resource (FAVOR) as prior information to prioritize genetic variants that are more likely to play a role in DNAm.

Through both simulation studies and analysis of GWAS summary statistics for complex traits, we demonstrate that MIMOSA successfully and substantially increases the accuracy of DNAm imputation in blood and identifies markedly more CpG sites-trait associations than existing methods. To conclude, we present case studies for Alzheimer's Disease and Prostate Cancer and identify 111 and 444 significantly associated CpG sites, respectively.

Epigenetics Posters - Wednesday

PB2452. Multiscale phase separation by explosive percolation with the single chromatin loop resolution

Authors:

D. Plewczynski¹, K. Sengupta¹, A. Agarwal¹, M. Denkiewicz¹, A. Mollah¹, T. Szczepinska¹, R. DSouza², Y. Ruan³; ¹Univ. of Warsaw, Warsaw, mazovia, Poland, ²Dept. of Computer Sci., Univ. of California, Davis, CA, ³The Jackson Lab. for Genomic Med., Farmington, CT

Abstract Body:

We propose models of dynamical human genome folding into hierarchical components in human lymphoblastoid and stem cell lines. Our models are based on explosive percolation theory. The chromosomes are modeled as graphs where CTCF chromatin loops are represented as edges. The folding trajectory is simulated by gradually introducing loops to the graph following various edge addition strategies that are based on topological network properties, chromatin loop frequencies, compartmentalization, or chromatin epigenomic features. Finally, we propose the genome folding model as a biophysical pseudo-time process of chromatin loops formation guided by a single scalar order parameter, which value is calculated by Linear Discriminant Analysis from chromatin features to classify the compartments efficiently. The chromatin phase separation, where fiber is condensing in three-dimensional space into topological domains and compartments, is observed when the critical number of contacts is reached. Overall, our in silico model integrates the high-throughput 3D genome interaction experimental data with the novel theoretical concept of phase separation, which allows us to model event-based time dynamics of chromatin loop formation and folding trajectories. Availability: <https://github.com/SFGLab/percolation> Contact: d.plewczynski@cent.uw.edu.pl

Epigenetics Posters - Thursday

PB2453*. NANOME: A Nextflow pipeline for haplotype-aware allele-specific consensus DNA methylation detection by nanopore long-read sequencing

Authors:

Y. Liu¹, Z. Pan^{1,2}, C. Chatzipantsiou³, T. Slocum¹, L. Karuturi¹, S. Li^{1,2,4}; ¹The Jackson Lab. for Genomic Med., Farmington, CT, ²Uconn Hlth., Farmington, CT, ³Lifebit Biotech Ltd., London, United Kingdom, ⁴Univ. of Connecticut, Storrs, CT

Abstract Body:

Nanopore long-read sequencing expanded the capacity of long-range, single-base, and single-molecule DNA-methylation (DNAMe) detection and haplotype-aware allele-specific epigenetic phasing. Previously, we benchmarked and ranked the robustness of seven computational tools for DNAMe detection using nanopore sequencing. Overall, the top performers were Megalodon, Nanopolish, DeepSignal, and Guppy. The methylation detection at regions with discordant non-singleton DNAMe patterns, intergenic regions, low CG density regions, and repetitive regions exhibited lower performance. The state-of-the-art algorithms for allele-specific methylation detection using nanopore sequencing, e.g., NanomethPhase and PRINCESS, only incorporate on single methylation calling tool - Nanopolish. Thus, it is desirable to further enhance the performance of the DNA methylation calling for long-range epigenetic phasing. Furthermore, long-read sequencing analysis for the mammalian genome requires much more computing resources than next-generation sequencing. Thus, scalability and reproducibility are desirable for nanopore sequencing data analysis. Currently, no pipeline integrates and automates nanopore sequencing DNA methylation detection using cloud computing. Therefore, we developed NANOME, the first Nextflow-based container environment (Docker and Singularity) for consensus DNA methylation detection. We designed a consensus DNA methylation predictive model using XGBoost, a gradient boosting meta-learning algorithm which integrate the output of the top three performers from our benchmarking results. The consensus model improves the F1-score accuracy of DNA methylation detection by 3-5% at single-molecule resolution and enhances the mean square error (MSE) by 9%-13% at single-base resolution. The consensus model also detects more CpGs (1-7%) than other tools. Combing the variants calling by Clair3 and long-read phasing by WhatsHap, NANOME detects the haplotype-aware allele-specific DNA methylation of known imprinting control regions using two normal human B-lymphocyte cell lines (NA12878 and NA19240). Moreover, the pipeline supports tera-base scale data analysis with a single command. NANOME is an open-source, reproducible, and end-to-end pipeline for whole-genome DNAMe detection, and is compatible with multiple HPC clusters and cloud OS platforms. A web-based interface is available using Lifebit CloudOS for cloud-based analysis, monitoring processes, and visualizing results from a single command line. NANOME is a complete step forward for DNA methylation detection and long-range epigenetic phasing.

Epigenetics Posters - Wednesday

PB2454. Nascent RNA reveals new links between genetic variation and disease at multiple levels of regulation

Authors:

C. Buen Abad Najjar, B. J. Fair, Y. Zeng, G. Mossian, S. Lozano, J. Staley, Y. I. Li; Univ. of Chicago, Chicago, IL

Abstract Body:

RNA sequencing is the most popular approach to study the effects of genetic variants on gene expression and splicing. Many expression QTLs (eQTLs) have been traced back to changes in chromatin accessibility or promoter activity. However, many regulatory steps between chromatin activation and the final mRNA molecule, as well as a myriad of non-coding RNAs, are not captured by standard sequencing of mature mRNA. Here we use RNA sequencing of nascent RNA (chRNA-seq) on 86 LCL Yoruba cell lines to explore the effect of genetic variation on these previously “invisible” regulatory steps. As expected, we find that chRNA-seq recovers the majority of the eQTLs and splicing QTLs (sQTLs) that are captured by standard RNA-seq (73% and 79%, respectively). In addition, we identified a large set of molQTLs that have not yet been studied. In total, 430 splicing and intron retention QTLs were associated with mature mRNA eQTLs; of these, 126 were exclusively discovered in chRNA-seq. Many of these chRNA-seq specific QTLs are linked to steady-state eQTLs. Some of these effects occur through nonsense-mediated decay, while others are associated with effects in chromatin marks. We also discover hundreds of novel links between genetic variation, expression of non-coding RNAs, including previously unannotated transcripts, and that of protein-coding genes. Using colocalization of molQTLs and GWAS data, we find that many new variants are linked to disease susceptibility. In conclusion, profiling nascent RNA reveals a large class of previously unknown links between genetic variation and disease at multiple levels of regulation.

Epigenetics Posters - Thursday

PB2455. Network analysis across cellular reprogramming states.

Authors:

C. Thangavelu¹, T. M. Norden-Krichmar²; ¹Univ. of California Irvine, Irvine, CA, ²Univ of California Irvine, Irvine, CA

Abstract Body:

Background: Cellular reprogramming is a process in which somatic cells, such as fibroblasts, are converted into pluripotent stem cells (iPSCs). In this study, we integrated time course chromatin accessibility and gene expression data to create time-resolved regulatory networks as cells undergo reprogramming in order to further investigate the regulatory patterns involved. **Methods:** RNA and ATAC-sequencing data that was generated as mouse fibroblasts transitioned to iPSCs was obtained from the publicly available GEO database, GSE101905. Biological replicates were merged and differentially accessible peaks were identified (bedtools2 v2.29.2). For each timepoint, motifs that were the most highly enriched in accessible peaks were identified (HOMER v4.11.1) Pre-processed RNA-seq files were consolidated (TxImport v1.18.0). ATAC-seq SRA files were aligned to the mouse mm10 genome using bowtie2 v2.3.2. ATAC-seq files were sorted and indexed (samtools v1.10) and mitochondrial DNA alignments were excluded (removeChrom Harvard ATAC-seq module). PCR duplicates were then removed (Picard v2.24.1) and mouseENCODE blacklisted genomic regions (Boyle Lab Blacklist.v2) were filtered out (bedtools2 v2.29.2). For each sample, ATAC-seq BAM files were integrated with TPM gene expression files (PECA2 v3.0.1) to generate network files (Cytoscape v3.9.1). **Results:** Timepoint comparison showed key network changes across fibroblast, intermediate, and pluripotent states, including the silencing of the somatic transcription program and the activation of the pluripotency gene regulatory network. These changes were consistent with motif enrichment findings across the different timepoints. **Conclusions:** Integrating transcriptomic and epigenomic data revealed distinct network changes as fibroblasts reprogrammed to a pluripotent state. This integrative analysis expands upon studies that use single data types to construct networks by leveraging multi-omics to gain novel insights into the mechanism of reprogramming.

Epigenetics Posters - Thursday

PB2456. Nucleosome spike-in controls identify best-in-class antibodies and enable reliable next-generation epigenomic mapping approaches

Authors:

N. Husby, L. M. Lewis, K. E. Maier, E. Weinzapfel, I. K. Popova, B. J. Venters, K. Novitzky, D. N. Maryanski, K. L. Rodriguez, M. R. Marunde, C. P. Lin, A. Gladney, R. Watson, Z. B. Gillespie, A. Adeleke, S. Gopinath, H. Willis, L. Sun, M. Cheek, M. J. Meiners, J. R. Bone, M. Cowles, Z-W. Sun, M-C. Keogh; EpiCypher Inc, Durham, NC

Abstract Body:

Mapping the genomic localization of transcriptional signaling proteins and epigenetic modifications is critical to understand chromatin function in basic biology and disease. However, applications in clinical research and other high impact projects have been limited by available technologies, such as ChIP-seq. Despite its widespread use, ChIP-seq suffers from multiple drawbacks, including high cell inputs, low reliability/accuracy, and high costs. The accuracy of antibodies to chromatin targets, particularly histone post-translational modifications (PTMs), is also a major concern. To enable improved ChIP-seq studies, we developed panels of DNA-barcoded recombinant nucleosome spike-in controls, which represent the physiological targets of chromatin mapping assays. We first validated these tools in ChIP-seq by characterizing >400 commercially available histone PTM antibodies, revealing alarming rates of cross-reactivity and poor enrichment (>70%; data available at www.ChromatinAntibodies.com). These problems with ChIP-seq emphasize the broad need for alternative epigenomic mapping methodologies as well as the development of compatible spike-in control standards. The recent development of new immunotethering assays, such as Cleavage Under Targets and Release Using Nuclease (CUT&RUN), deliver high signal-to-noise mapping data using a fraction of the required cells, sequencing depth, and costs compared to ChIP-seq. We show that CUT&RUN enables reliable genomic profiling for a variety of targets, including histone post-translational modifications (PTMs), transcription factors, and traditionally intractable ChIP targets, such as chromatin remodelers. We have also recently begun developing nucleosome spike-in controls for CUT&RUN assays and are using them to identify best-in-class histone PTM antibodies and optimize workflows. This antibody validation effort, coupled with routine use of the spike-in controls for continuous assay monitoring, improves assay rigor and reproducibility to ultimately advance CUT&RUN into the next frontier of genomic mapping research.

Epigenetics Posters - Wednesday

PB2457*. Peripheral immune roles in bipolar disorder: genomic, epigenomic, and phenotypic integration

Authors:

L. Hou¹, Y. Li², Y. Park³, X. Xiong⁴, M. Frye⁵, J. Biernacka⁵, T. Ordog⁵, M. Kellis⁶; ¹MIT & Broad Inst., Cambridge, MA, ²McGill Univ., Montreal, QC, Canada, ³Univ. of British Columbia, Vancouver, BC, Canada, ⁴MIT, Brookline, MA, ⁵Mayo Clinic, Rochester, MN, ⁶MIT, Cambridge, MA

Abstract Body:

Aims: Immune dysfunction is implicated in bipolar disorder (BD) but remains poorly studied. Here, we systematically chart the epigenomic landscape of immune alterations in BD in 188 whole-genome-sequenced and deeply phenotyped individuals, to characterize the genetic variants, epigenetic alterations, and gene-regulatory circuits underlying immune contributors to BD. Methods: We profile and integrate 940 histone mark profiles targeting active (H3K27ac), repressed (H3K27me3), enhancer (H3K4me1), promoter (H3K4me3), and transcribed (H3K36me3) signatures from blood samples (buffy coat) of 94 BD patients and 94 healthy controls, with whole-genome sequencing and medical records. We annotate 450k cis-regulatory elements (CREs), study their variation across individuals, and cluster BD-informative features into 23 latent factors, used to group individuals into medical-record-informed patient subtypes. We also integrate BD-associated genetic variants, BD-differential CRE activity, chromatin interactions, and expression quantitative trait loci, to recognize BD-associated gene-regulatory circuits. Results: We find 33k CREs in 2257 genes showing significant differences between BD and controls, enriched in inflammation-related processes. Our epigenomics-based clustering reveals 5 patient subgroups correlated with different medical terms: addiction-related, inflammation/infection-enriched, polyethylene-glycol-3350-enriched (constipation) and glucose intolerance-enriched, hypertension-enriched and a final group with no associated medical term. Our gene-regulatory analysis reveals two main biological immune drivers of BD: inflammatory/autophagy-related circuits specific to blood immune cells, and cellular-structural-related circuits shared between blood immune cells and brain neural cells. Conclusion: Our study demonstrates the power of genome-epigenome-phenotype integration, and the important roles that immune cells play in complex mental disorders, with implications for biomarker discovery, disease or treatment progression monitoring, and for potential interventions targeting causal peripheral immune contributors to BD.

Epigenetics Posters - Thursday

PB2458. Rare inherited Copy Number Variants as genetic modifiers of Developmental Disorders.

Authors:

M. Atzori¹, L. Hannes^{1,2}, E. Pelgrims¹, A. Swillen^{1,2}, A. Sifrim^{1,3}, J. Breckpot^{1,2}, C. Attanasio¹; ¹KU Leuven, Leuven, Belgium, ²Univ. Hosp. Leuven, Leuven, Belgium, ³KU Leuven Inst. for Single Cell Omics (LISCO), Leuven, Belgium

Abstract Body:

Neurodevelopmental disorders (NDDs) are a group of cognitive and/or behavioural conditions that have onset in childhood. They are often characterized by variable phenotype expressivity, incomplete penetrance, and pleiotropy of clinical features. Rare inherited Copy-Number Variants (CNVs), which are usually discarded as putatively pathogenic in the diagnostic set up, could play a pivotal role in the sensitization of the genome and modulation of NDD phenotypic traits. To test this hypothesis, we have designed a family-based study combining transcriptomics and chromatin conformation analyses in carriers and non-carriers. Specifically, we have recruited 6 family trios, for 6 unique rare CNVs that range from 118kb to 1Mb in size and map preferentially to gene-poor regions of the genome ($n \leq 10$). In all families the CNV was transmitted to the affected child by a seemingly healthy parent. EBV-transformed lymphoblastoid cell lines were established for each individual and used for gene expression (RNA-seq) and chromatin structure analysis (Capture Hi-C). To complement these data, whole-genome sequencing (WGS) and deep familial phenotyping are also being performed to identify additional putative NDD causative genetic variants and correlate CNV familial segregation with NDD phenotypes. Our multi-layered analyses are still ongoing, however for our family 2 (F2) the results show that the rare inherited CNV is not associated to other *de novo*, recessive, X-linked or paternally inherited variant or SNV, indels and additional CNVs. Deep phenotyping analysis, instead, indicates that the segregation of the 450kb deletion CNV on chromosome 4 correlates with a progressive aggravation of cognitive and behavioral abilities, with mild symptoms in the carrier parent and severe one in the affected child. At the level of gene expression analyses, our data show differential gene expression for genes within ($n = 1$) and in the flanking regions ($n = 2$) of the CNV. This points toward a role for genes not yet associated to NDDs, such as *ZNF330*, which is disrupted by the deletion, and *SETD7*, which is upstream of the CNV. Finally, Capture Hi-C data seem to reveal a unique chromatin interaction profile in the F2 index patient compared to the parents and additional non-carrier controls. Interestingly, the F2 index patient shows a depletion of interactions at the promoter of *ZNF330* compared to the non-carrier controls which could explain the decreased of expression observed in RNA-seq data. These preliminary results are still subjected to further study. Overall, however, our data on F2 seems to point toward a contributory effect of rare inherited CNVs to DD's phenotypic variability.

Epigenetics Posters - Wednesday

PB2459. Reduced-Representation Methylation Sequencing (RRMS): Oxford Nanopore's alternative to RRBS enables accurate and easy to use methylation calling in CpG islands, shores, shelves, and promotor regions using a MinION flowcell.

Authors:

P. Rescheneder¹, R. Esteban¹, I-A. Vasilescu¹, S. Juul², D. Stoddart¹; ¹Oxford Nanopore Technologies, Oxford, United Kingdom, ²Oxford Nanopore Technologies, New York, NY

Abstract Body:

Nanopore sequencing enables direct detection of methylated cytosines (e.g., at CpG sites). CpG sites frequently occur in clusters called CpG islands (CGI) and >60% of human genes have their promoters embedded within CGIs. Changes in methylation patterns within promoters is associated with changes in gene expression. Reduced representation bisulphite sequencing (RRBS) is a method for methylation analysis without the need to sequence the whole genome but is expensive and time consuming. Furthermore, the complex library preparation method does not specifically target any promoter region. Oxford Nanopore's Adaptive sampling (AS) enables targeting regions of interest (e.g., CGIs) by rejecting off-target reads during sequencing with no requirement for upfront sample manipulation. Here we introduce Reduced-Representation Methylation Sequencing (RRMS). We use AS to target 310 Mb of the human genome that are enriched for CpGs including ~28,000 CpG islands, ~50,600 shores and ~42,700 shelves as well as ~21,600 promotor regions. We performed RRMS on five replicates of a metastatic melanoma cell line and its normal pair and a triple negative breast cancer cell-line pair and compared results to RRBS. RRMS resulted in high confidence methylation calls for 7.3 - 8.5 million CpGs per sample. RRBS in comparison yielded 1.7 - 2.5 high confidence calls per sample. Methylation frequencies called by RRMS and RRBS are highly similar for CpGs covered by both technologies ($R > 0.967$). We show that coverage across target regions is extremely uniform and reproducible with RRMS. As a result, we were able to detect in total 62 mega bases of differentially methylated regions (DMRs) between tumour and normal pairs of which a high proportion overlaps with cancer genes. In comparison RRBS yielded 20 mega bases of DMRs. When investigating the DMRs further we identified the de-novo methyltransferase *DNMT3A* promoter to be methylated in both tumours but not in their normal counter parts. These methyltransferases are essential for establishing and maintaining normal levels of methylation, and so their dysregulation can contribute to cancer development. Finally, we demonstrate that the short rejected reads can be used to call copy-number variation across the whole genome. Furthermore, target regions can easily be extended to specific genes of interest to enable SNP and SV calling if required. Combined with its ease of use and ability to scale to a high number of samples, these characteristics make RRMS perfectly suited to investigate methylation differences in large cohorts as well as provide deeper insight into the mechanisms behind diseases like cancer or monitoring tumour progression.

Epigenetics Posters - Thursday

PB2460. RFX transcription factors regulate genes involved with primary cilium in glioma

Authors:

S. Chua^{1,2}, I. Inoue^{2,1}; ¹The Graduate Univ. for Advanced Studies, Hayama, Japan, ²Natl. Institute of Genetics, Mishima, Japan

Abstract Body:

The brain and spinal tumors derived from the neuron-supporting glial cells called glioma is classified as astrocytoma, oligodendroglioma, and the highest-grade glioblastoma. Glioblastoma cell lines and patient tumors have abnormal primary cilium or completely lack it. The primary cilium is crucial for glioma proliferation and drug sensitivity as it houses the sensors for extracellular signals and regulates signaling pathways. We hypothesize a gene is causing the defect to the primary cilium and identifying this gene may have the potential to restore ciliary function. To investigate, we used publicly available ATAC-sequencing, RNA-sequencing, and clinical data of glioma. We found a ciliary transcription factor Regulatory factor X (RFX) Xbox motif enriched in the active chromatin regions of patients. We also found that *RFX1*, *RFX2*, and *RFX3* are differentially expressed in patients compared to normal brain samples. The abundance of the Xbox motif indicates that the irregularity of *RFX1-3* expression may affect numerous regulated genes, which our gene ontology analysis showed are important for ciliogenesis. Furthermore, we show that *RFX1* and *RFX3* expression are prognostic markers for low-grade glioma and glioblastoma patients, respectively, and can be used to determine high- and low-risk groups. This study provides target genes for manipulating primary cilium formation or length and opens the potential of ciliotherapy as a therapeutic strategy in glioma.

Epigenetics Posters - Wednesday

PB2461*. Role of TET1-mediated epigenetic modulation in Alzheimer's disease

Authors:

M. Armstrong¹, Y. Jin², T. S. Wingo², A. P. Wingo^{2,3}, P. Jin²; ¹Emory University, Atlanta, GA, ²Emory Univ. Sch. of Med., Atlanta, GA, ³Atlanta VA Med. Ctr., Decatur, GA

Abstract Body:

Alzheimer's disease (AD) is a neurodegenerative disease with development and progression influenced by a complex interplay between environmental, epigenetic, and genetic components. DNA methylation (5mC) and hydroxymethylation (5hmC) are DNA modifications that function as tissue-specific and temporal regulators of gene expression. These epigenetic modifications are dynamically regulated by TET family enzymes in response to environmental conditions and provide a link between environmental factors and gene expression. Prior epigenetic studies have identified 5mC and 5hmC changes associated with AD and more recently, rare AD-associated variants have been identified in *TET2* through deep sequencing. We performed *TET1* targeted sequencing on 349 early-onset AD (EOAD) and 940 control samples from the National Alzheimer's Coordinating Center and National Centralized Repository of Alzheimer's Disease and Related Dementias (NACC/NCRAD) cohort. Through gene-wise burden analysis, we identified significant enrichment of rare *TET1* variants associated with AD ($p=0.04$). Using the RUSH Religious Orders Study and Rush Memory and Aging Project (ROSMAP) cohort, we observed a significant increase in expression of *TET2* ($p = 0.0009$), *TREM2* ($p<0.0001$), and a near significant increase in expression of *TET1* ($p=0.0523$) in the AD group. We further profiled 5hmC using a subset of AD and control samples. Analysis of 5hmC identified DhMRs in key genes responsible for regulating the methylome *TET3*, *DNMT3L*, *DNMT3A*, and *MECP2*. To further test the role of Tet1 in AD pathogenesis, we utilized the 5xFAD mouse model with Tet1 KO allele to examine how the loss of Tet1 influences AD pathogenesis. We observed significant changes in behavior, neuropathology, 5hmC, and RNA expression associated with the loss of Tet1. Loss of Tet1 significantly increased amyloid plaque burden in the 5xFAD mouse ($p = 0.0197$) and increased mobility in the Forced-swim assay ($p = 0.0369$). At the molecular level, we observed significant DhMRs enriched in pathways responsible for neuronal projection organization, dendritic spine development and organization, and myelin assembly. RNA-Seq analysis revealed a significant increase in the expression of AD-associated genes such as *Mpeg1*, *Ctsd*, and *Trem2*. Taken together, our results suggest that the TET enzymes responsible for regulating the methylome, in particular TET1, could contribute to the pathogenicity of AD: where the loss of TET function increases AD-associated pathology.

Epigenetics Posters - Thursday

PB2462. Scoring individual cells in snATACseq dataset based on GWAS fine-mapping identifies disease-relevant cell-types in neurodegenerative diseases

Authors:

A. Mohamed¹, B. van de Geijn², J. Blischak³, M. McCarthy³, B. Friedman³, T. Bhangale³; ¹Hoffman La Roche Ltd, Calgary, AB, Canada, ²Genentech, Foster City, CA, ³Genentech, South San Francisco, CA

Abstract Body:

Since most common variants causal to complex traits likely act through their impact on non-coding regulatory elements, approaches for integrating GWAS results with scATAC-seq data offer great potential to provide insights into the underlying cell-types and mechanisms. One technique for such integration, recently described by Yu et al (bioRxiv 2022, SCAVENGE), is to calculate a trait relevance score (TRS) for individual cell based on the overlap of its open chromatin with putative causal variants identified using fine-mapping (FM) of GWAS loci, followed by correlating the scores with cell-types/sub-populations, cell-state, and trajectory. We extend this approach to include signals outside of established GWAS loci using genome-wide FM, substantially boosting the attributable proportion of trait heritability and allowing us to apply TRS to neurodegenerative disease traits (NT).

We fine-mapped summary statistics from the largest GWAS for four NTs: Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), and multiple sclerosis (MS). For comparison, we fine-mapped four non-neurological traits (non-NT): RBC distribution width, hypertension, blood glucose, and height. FM was performed in 3Mb tiled windows across the genome to obtain posterior inclusion probabilities (PIP) for all common variants. We obtained snATAC- and snRNA-seq data for 191,890 nuclei from postmortem brains of 12 AD and 11 healthy brains from the work by Morabito et al (Nat Genet 2022) and used the ArchR framework for data processing, analysis and integration. We then used SCAVENGE to obtain TRS based on PIPs and snATAC for each combination of nucleus and trait. snRNA-seq data was used to predict neurons and various glia nuclei based on previously known markers.

We find significantly high TRS (Δ TRS > 1 and p value < 1E-7) relative to non-NT traits for: 1. microglia (MG) in MS and AD; 2. Astrocytes (ASC) in AD; 3. Excitatory (EXC) neurons in AD; 4. Oligodendrocytes (ODC) in AD. ASC showed significantly depleted scores for ALS and MS. Similar patterns were observed in analyses stratified by AD cases and controls. Cases had a higher value of TRS compared to controls for subclusters of activated 1. MG TRS across all NT but not in non-NT, 2. ASC and OPC TRS for AD, and 3. ODC for ALS.

The role of microglia in AD and MS has been well described in the past, but our approach was able to capture additional cell-populations which could point to distinct pathways and mechanisms. In summary, our technique systematically integrates genome-wide causal variant information from GWAS with snATAC-data and can offer novel insights into the causal mechanisms.

Epigenetics Posters - Wednesday

PB2463*. Serum miR-451a is up-regulated in sarcopenic patients compared to healthy controls

Authors:

S. Agostini¹, R. Mancuso¹, L. Citterio¹, F. Trecate¹, M. Clerici^{1,2}; ¹IRCCS Don Gnocchi Fndn., Milano, Italy, ²Univ. of Milano, Milan, Italy

Abstract Body:

Sarcopenia is a geriatric syndrome characterized by loss of muscle strength and mass, associated with declining physical function. miRNAs are short molecules of RNA able to modulate gene expression; they can be found in serum and often are useful biomarkers of several diseases. The aim of the present work is to verify if the circulatory miR-451a is differentially expressed in sarcopenic patients compared to age- and sex- matched controls (HC). miR-451a expression was evaluated in serum of 45 severe sarcopenic patients (SPPB<3) 50 mild sarcopenic patients (3≤SPPB≤9) and of 136 HC by droplet digital PCR (ddPCR), an innovative technique that can measure miRNA expression as absolute values and not relative to a reference gene as traditional qPCR. We found that the expression of circulatory miR-451a was significantly higher ($p<0.05$) in serum of severe sarcopenic patients (5.45×10^3 ; 2.85×10^3 - 1.26×10^4 c/ng) and mild sarcopenic patient (6.59×10^4 ; 1.12×10^4 - 2.5×10^5 c/ng) compared of HC (2.31×10^4 ; 1.94×10^3 - 2.01×10^5 c/ng). The severe sarcopenic patients underwent a rehabilitation program of 30-days, and, interestingly, the rehabilitation caused a significantly increase of circulatory miR-451a expression (1.22×10^4 ; 5.15×10^3 - 3.00×10^4 c/ng; $p=0.01$), and, when the difference of the expression before and after rehabilitative treatment (delta) was calculated for miRNAs in relationship with SPPB, a significant negative correlation was found between the delta of miR-451a expression and the delta of SPPB ($p<0.05$). In conclusion our results suggest that the measurement of serum miR-451a can discriminate between HC sarcopenia, both for mild or severe symptoms, and can be a useful biomarker for the rehabilitative outcome for the disease.

Epigenetics Posters - Thursday

PB2464. Simultaneous sequencing of genetics and epigenetics provides opportunities for new biological insights.

Authors:

C. Lumby¹, S. Balasubramanian², P. Burns¹, T. Charlesworth¹, P. Creed¹, J. Füllgrabe¹, W. Gosal¹, J. D. Holbrook¹, D. Morley¹, S. Yu¹; ¹Cambridge Epigenetix, Cambridge, United Kingdom, ²Univ. of Cambridge, Cambridge, United Kingdom

Abstract Body:

There is more to DNA than the genetic alphabet A, C, G and T. Epigenetics is of growing scientific interest and plays a causal role in cell fate, ageing, response to environment and disease development. The ability to accurately determine both genetic and epigenetic letters, and their interaction with each other, is important for deriving new biological insights.

In humans, chemical modification of cytosines represents the most common form of epigenetic letters. Of these, 5-methylcytosine and 5-hydroxymethylcytosine are the most frequent and exhibit distinct biological functions. Current approaches for determining modified cytosines apply next-generation sequencing in combination with a pre-sequencing protocol that differentially converts genetic bases dependent on epigenetic status. However, the ability to discriminate epigenetic state comes at a cost of incomplete genetic determination; inherently, it is impossible to determine full genetics and epigenetics from a system limited to only four states of information. Partial genetic sequencing results in suboptimal read mapping and variant discovery.

An alternative approach is to perform multiple analyses on aliquots from the same sample. However, multiple analyses are not only expensive in terms of sample volume, time and reagent costs, but also yield suboptimal information due to the inherent inaccuracy of data integration and the loss of phasing between genetic and epigenetic letters.

We present a technology, 5-Letter seq, that addresses these challenges by expanding the number of information states in next-generation sequencing from four to sixteen. This is accomplished by utilising pairs of bases (some of which have been differentially converted) to code for each of sixteen states. The expansion of the information content to sixteen states allows direct, digital and phased discrimination of genetic and epigenetic letters on the same read. Sixteen state encoding also enables suppression of sequencing derived artefacts resulting in very high accuracy for both genetics and epigenetics. The generation of phased high-quality variant call data and epigenetic information enables 5L-seq to examine unique biological phenomena. One functionally important example is allele specific methylation, in which differential methylation levels are observed across alleles at a heterozygous variant site. Additionally, our method facilitates the exploration of methylation at CpG sites absent from the reference genome by using variant call data to define sample specific CpGs. Overall, 5L-seq delivers a more comprehensive and accurate genetic and epigenetic assessment than was previously possible.

Epigenetics Posters - Wednesday

PB2465*. Single-cell sequencing links reduced glucose metabolism to cocaine addiction-like behavior in rats

Authors:

J. Zhou¹, G. McVicker², F. Telese¹; ¹Univ. of California San Diego, San Diego, CA, ²Salk Inst. for Biological Studies, La Jolla, CA

Abstract Body:

The United States faces an epidemic of substance use disorders with an alarming increase in overdose deaths involving stimulants, such as cocaine. However, we still have limited knowledge of the cell type-specific mechanisms underlying cocaine addiction. Prior studies examining the effects of cocaine on transcriptional changes in rodent brains primarily analyzed bulk tissues, hindering our understanding of how distinct cell types respond to cocaine. Here we applied single-nucleus RNA-seq (snRNA-seq) and ATAC-seq (snATAC-seq) to the amygdala of heterogeneous stock (HS) rats trained to self-administer cocaine under extended access conditions and subjected to 5 weeks of abstinence. Rats were classified as having a low or high addiction index based on several behavioral measures of addiction severity. We identified major neuronal and glial cell types in the amygdala and based on this annotation analyzed differences in the transcriptomic and epigenomic data coming between the high addiction index rats versus the low addiction index rats. Pathway enrichment analysis of cell type-specific DEGs revealed that several signaling pathways were perturbed by cocaine use in a cell type-specific manner, including changes in glucose metabolism in neuronal cell types. We experimentally perturbed a step in the glycolysis pathway and successfully rescued key cellular and behavioral measures associated with addiction-related phenotypes. This may be due to changes in levels of methylglyoxal, a partial agonist for the GABA-A receptor, indicating that disruption of GABA transmission is associated with differences in cocaine addiction. To better understand the mechanisms driving this phenomenon, we examined the cell type-specific differentially accessible genomic regions between nuclei from high and low addiction index rats. Indeed, we found that these regions were enriched for TSS and promoter regions compared to background and that DEGs were enriched for having differentially accessible promoter regions in the same cell. We identified enriched transcription factor motifs in the cell type-specific differentially accessible OCRs and performed a co-accessibility analysis to link differential OCRs to DEGs. Overall, we identified a number of candidate regulatory mechanisms driving perturbation of glucose metabolism in neuronal cells. Our work has unveiled a new cell-type-specific pathway associated with cocaine addiction and the mechanisms driving its perturbation to advance understanding of the molecular basis of the neuroadaptations induced by long-term use of cocaine.

Epigenetics Posters - Thursday

PB2466. Single-molecule architecture and heterogeneity of human telomeric DNA and chromatin

Authors:

A. Stergachis, D. Dubocanin, A. Seden Cortes, J. Ranchalis, T. Real, B. Mallory; Univ. of Washington, Seattle, WA

Abstract Body:

Telomeres are essential for linear genomes, yet their repetitive DNA content and somatic variability has hindered attempts to delineate their genetic and chromatin architectures. We performed single-molecule chromatin fiber sequencing (Fiber-seq) on human cells with a fully resolved genome, enabling nucleotide-precise maps of the genetic and chromatin structure of all telomeres. Telomere fibers are predominantly comprised of three distinct chromatin domains that co-occupy individual DNA molecules - multi-kilobase telomeric caps, highly accessible telomeric-subtelomeric boundary elements, and subtelomeric heterochromatin. Extended G-rich telomere variant repeats (TVRs) punctuate nearly all telomeres, and telomere caps imprecisely bridge these degenerate repeats. Telomeres demonstrate pervasive somatic alterations in length, sequence, and chromatin composition, with TVRs and adjacent CTCF-bound promoters impacting their stability and composition. Our results detail the structure and function of human telomeres.

Epigenetics Posters - Wednesday

PB2467. SnapHiC-G: identifying long-range enhancer-promoter interactions from single cell Hi-C data via a global background model

Authors:

W. Liu¹, W. Zhong², P. Giusti-Rodriguez³, M. Hu⁴, Y. Li¹; ¹Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ²Merck & Co., Inc., Rahway, NJ, ³Univ. of Florida Coll. of Med., Gainesville, FL, ⁴Cleveland Clinic, Cleveland, OH

Abstract Body:

Harnessing the power of single cell genomics technologies, single-cell Hi-C and its derived co-assays such as sc-methyl-Hi-C and sn-m3c-seq, provide powerful tools to measure spatial proximity between cis-regulatory elements in individual cells. In our recently developed SnapHiC method, we combined both global and local background models to identify chromatin loops from scHi-C data. As a result, most SnapHiC-identified chromatin loops are CTCF-anchored interactions, while the sensitivity to identify cell-type-specific long-range enhancer-promoter interactions is limited. To fill in this gap, we propose SnapHiC-G, a new computational approach based solely on the global background model to identify long-range enhancer-promoter interactions from scHi-C data. We applied SnapHiC-G to analyze scHi-C datasets generated from mouse embryonic stem cells (mESCs) and human brain cortical cells, and showed that SnapHiC-G outperformed SnapHiC and other existing methods designed for bulk Hi-C data (FitHiC2, FastHiC, HiC-ACT, HiC-DC+), demonstrating higher sensitivity in identifying long-range enhancer-promoter interactions. Moreover, SnapHiC-G-identified enhancer-promoter interactions were significantly enriched with brain eQTL-gene pairs, using scHi-C data from human brain cortical cells and brain eQTL data from the CommonMind Consortium. Finally, we showed that SnapHiC-G can identify putative target genes of GWAS variants, and genetic heritability for neuropsychiatric diseases was enriched for SNPs within SnapHiC-G identified interactions in a cell type-specific manner. SnapHiC-G is a powerful tool for characterizing cell-type-specific enhancer-promoter interactions in single cells from complex tissue samples, and fostering the interpretation of noncoding GWAS variants. The SnapHiC-G software is freely available at <https://github.com/wjzhong/SnapHiC-G>.

Epigenetics Posters - Thursday

PB2468. SpliceVI: a visualization tool for predicting protein consequences based on SpliceAI data.

Authors:

Y. Cho, J. Kim, H. Lee, G. Seo; 3billion, Seoul, Korea, Republic of

Abstract Body:

SpliceAI, an *in silico* tool that calculates how likely a given variant may alter splicing, outputs Δ scores and relative positions from the variant in question for each Δ type [acceptor loss (AL), acceptor gain (AG), donor loss (DL) and donor gain (DG)] within a preset distance. Variants with high Δ score are likely to be interpreted as potential null, just like essential splice site variants are commonly interpreted as. However, the actual protein consequence (PC) may as well be an inframe change and therefore, even though SpliceAI output is a likelihood, predicting the PC accordingly is important. Currently, the user has to manually figure out the possible PC based on SpliceAI output. Here, we developed a user-friendly visualizer, SpliceVI, that displays the genomic positions of the Δ type relative to the given variant and outputs predicted PC. SpliceVI uses 'raw' SpliceAI scores with max distance of 1Kb as default. The MANE transcript is used for predicting the PC. The following rules are applied if the Δ scores are >0 for a pair of Δ type: 1) AL/AG or DL/DG: the region in between is spliced out if AG or DG is upstream of AL or DL in the flanking intron or exon, respectively, and retained if AG or DG is downstream of AL or DL in the flanking exon or intron, respectively, 2) AL/DL: the exon in between is skipped if AL and DL flank the same exon and AL is upstream of DL, and the intron in between is retained if AL and DL flank the same intron and AL is downstream of DL, 3) AG/DG: pseudoexon is created if AG and DG are more than a base apart, reside within an intron and AG is upstream of DG, and the region in between is spliced out if DG and AG are more than a base apart, reside within an exon and DG is upstream of AG. Then, the frame of the transcript and if/where a new stop-codon gets created are determined. For other scenarios such as 1) only one of the 4 Δ type has a Δ score, 2) DL/AL does not occur at a canonical splice site 3) the Δ type pair spans more than 1 exon or intron, the SpliceVI will only give a message that a manual inspection is required. For validation, we compared the PC prediction by SpliceVI to that by manual inspection, using 100 variants with Δ score >0.6 . All 100 were concordant, and it took 113 and 0.5 seconds on average by manual inspection and SpliceVI, respectively. SpliceVI is a reliable and convenient tool that not only provides visualization of where a cryptic splice site (CSS) may occur but also predicts what the PC could be based on SpliceAI data and therefore, utilization of SpliceVI will reduce human burden and error in interpreting a potential CSS variant. Lastly, even though SpliceVI currently only takes SpliceAI results as input, it could be flexible for data from other tools.

Epigenetics Posters - Wednesday

PB2469. Stability of genetic regulation of gene expression over time.

Authors:

Y. Raza¹, J. S. El-Sayed Moustafa¹, A. Ramisch², D. Wang¹, G. Leday³, Y. Jiao⁴, G. Nicholson⁴, M. Stevens⁵, M. Abdalla⁴, C. Menni¹, C. Holmes⁴, T. Spector¹, M. McCarthy⁴, S. Richardson³, E. Dermitzakis², K. Small¹; ¹King's Coll. London, London, United Kingdom, ²Univ. of Geneva, Geneva, Switzerland, ³Med. Res. Council Biostatistics Unit, Cambridge, United Kingdom, ⁴Univ. of Oxford, Oxford, United Kingdom, ⁵Univ. of Geneva, Geneva, Switzerland

Abstract Body:

Ageing is associated with a decline in function and an increased risk of disease. Unravelling the molecular changes that occur over time can help us understand the process of ageing. Whilst cross-sectional studies have identified strong links between ageing and changes in gene expression, few studies have explored the changes that occur within an individual over time. Given the heterogeneity of ageing, exploring individual trajectories allows us to better understand disease risk factors and the progression of ageing.

To investigate variation of gene expression we instigated the MultiMuTHER study of the TwinsUK cohort; a longitudinal dataset measured across three time points between 2009-2017, comprised of 335 individuals (mean age 60 years) with gene expression (whole blood RNASeq of 16,292 genes) and metabolomics (Metabolon profile 1,197 metabolites). The study also captured short-term technical and biological variation. We found a systemic effect of time on the transcriptome, with 40% of genes showing a linear association with time.

We aimed to understand whether the genetic regulation of gene expression is stable or changes over time using allele specific expression (ASE) to consider allelic imbalance within an individual. Following haplotype phasing of RNASeq reads using phASER (phasing and Allele Specific Expression from RNA-seq), to generate data across heterozygous SNPs and subsequently aggregated across haplotypes, ASEP (Allele-Specific Expression in a Population) was utilised to test for ASE in individuals and to compare differences in ASE across time.

Following multiple testing correction (FDR 5%) we observed a median of 9% of genes which display ASE, with 1045, 1021, and 1053 genes exhibiting ASE at timepoints 1, 2 and 3, respectively. 552 of these genes were common across all three timepoints, indicating ~ 500 unique genes with ASE at each timepoints. We further observed 694 genes common between timepoints 1 and 2, 671 common genes between timepoints 2 and 3, and 676 common genes between timepoints 1 and 3. We tested for differential ASE between timepoints to understand the stability of ASE over time. We found differential ASE in 18 genes between timepoints 1 and 3, with 27 genes in differential ASE between timepoints 1 and 2 and 20 genes between timepoints 2 and timepoints 3. This indicates genetic regulation of most genes remain stable over time, with a small number displaying changes over time. We observed genes which exhibit changes in ASE but did not show a change in gene expression levels indicating complex regulation. Further work will utilise methods to incorporate environment and phenotypic measures to consider gene-environment interactions.

Epigenetics Posters - Thursday

PB2470*. Subsetting systemic lupus erythematosus patients based on clustering of DNA methylation at the time of disease flare

Authors:

M. Horton¹, J. Nititham¹, K. Taylor², L. Trupin², P. Katz², C. Ye³, J. Yazdany², M. Dall'Era⁴, L. F. Barcellos⁵, L. A. Criswell⁶, C. Lanata¹; ¹NIH/NHGRI, Bethesda, MD, ²Univ. of California, San Francisco, San Francisco, CA, ³Univ. of California, San Francisco, San Francisco, CA, ⁴Univ. of California, San Francisco, San Francisco, CA, ⁵Univ. of California, Berkeley, Berkeley, CA, ⁶NIH/NIAMS, Bethesda, MD

Abstract Body:

Current treatments for systemic lupus erythematosus (SLE) do not adequately prevent disease progression. This lack of efficacy, in part, relates to the molecular heterogeneity of SLE. The objective of this study was to use DNA methylation from flaring SLE patients to identify distinct biological subtypes of SLE associated with disease activity. Fifty-three SLE flaring patients were studied. A flare was defined by the treating rheumatologist and characterized with the SLE Disease Activity Index (SLEDAI) score. Blood samples and clinical data were collected at the time of flare and three months later. Genome-wide methylation profiles were generated using Illumina's Infinium Human MethylationEPIC BeadChip. After CpG quality control, differentially methylated positions (DMPs) associated with SLEDAI score were identified using *limma*. Models accounted for the paired design and adjusted for confounders such as blood cell proportions, batch effects, age, sex, medications and genetic PCs. Consensus hierarchical clustering was performed on the 5,000 most significant DMPs to identify patient subgroups. Clinical features and DMPs unique to each subgroup were identified using t-tests, chi-square, and *limma*. Gene ontology enrichment analysis was performed on the set of DMPs unique to each subtype. Participants were predominantly female (85.5%) and of diverse racial and ethnic backgrounds. Three clusters of patients were identified using DNA methylation data. Mean SLEDAI was lowest for Cluster 1 (7.5 (sd=7.2)) and increased for each subsequent group (11.0 (sd=7.2) and 14.7 (sd=3.9) for Cluster 2 and Cluster 3, respectively). The most significant gene ontology pathways for Cluster 1 included phosphorylation processes and regulation of defense response by viruses. For Cluster 2, notable DMPs were found in *EBF1* (B-cell transcription factor), *C12orf66* (regulator of mTOR signaling, the target of rapamycin), and *SWTI* (riboendonuclease expressed in basophils). Top pathways included import into the nucleus and adrenergic receptor signaling. For Cluster 3, notable DMPs include *KMT2C* (methyltransferase expressed in neutrophils), *RFXP3* (receptor of relaxin, associated with monocyte recruitment and differentiation), and *ITSN2* (regulator of clathrin-coated vesicles and induction of T cell antigen receptor endocytosis). Pathways for Cluster 3 included synaptic transmission and several neuronal membrane pathways. Three biologically distinct subgroups of flaring SLE patients were identified using DNA methylation data. This subtyping might be used to better inform treatment decisions and targeted therapies based on relevant underlying biological pathways.

Epigenetics Posters - Wednesday

PB2471. TDP-43 depletion results in aberrant gene expression and TE activation via dysregulation of R-loop homeostasis.

Authors:

Y. Hou, Y. Li, P. Jin, B. Yao; Dept. of Human Genetics, Emory Univ. Sch. of Med., Atlanta, GA

Abstract Body:

TAR DNA-binding protein 43 (TDP-43; also known as TARDBP) is a DNA and RNA-binding protein involved in many aspects of RNA metabolism. TDP-43 pathological aggregation is a hallmark of many neurodegenerative disorders. However, the detailed mechanistic role underlying TDP-43 pathology in neurons remains largely elusive. R-loop, a three-stranded DNA:RNA hybrid structure formed when a single-stranded RNA invades the DNA double strand and hybridizes with a complementary DNA strand, has been revealed to be required for several cellular processes, such as regulation of gene expression and DNA damage repair. Recent studies have established a direct link between TDP-43 dysfunction and aberrant regulation of R-loops. To investigate the long-term impact of TDP-43 loss-of-function in neurons, we first established TDP-43 stable knockdown (TDP-43 KD) neuronal cells by infecting TDP-43 shRNA containing lentivirus into SH-SY5Y neuroblastoma cells. Genome-wide dysregulated R-loops were identified in the absence of TDP-43 by DNA-RNA immunoprecipitation (DRIP) sequencing. Given 5-hydroxymethylcytosine (5hmC) is an important epigenetic mark that regulates gene expression, and it has been shown to positively correlate with R-loops on genomic loci, we correlated abnormal R-loops/5hmC co-occurring peaks with aberrant gene expression and found a strong positive correlation between dysregulation of gene expression and imbalanced R-loops and 5hmC, suggesting an important role of TDP-43 in modulating gene expression via R-loop and 5hmC alteration. We also found TDP-43 loss induced transposable elements activation, which has been demonstrated as a phenotype observed in many neurodegenerative disorders and may have deleterious effects by disturbing the transcriptional landscape and causing cellular dysfunctions. Together, TDP-43 depletion modulates gene expression and TE activation via R-loop regulation, which may contribute to neurodegeneration.

Epigenetics Posters - Thursday

PB2472. The association of cigarette smoking with DNA methylation and gene expression in human tissues

Authors:

L. I. Tamayo¹, L. Tong¹, F. Jasmine¹, K. G. Muhammad¹, K. Demanelis², M. Oliva³, L. Chen¹, B. L. Pierce¹; ¹Univ. of Chicago, Chicago, IL, ²Univ. of Pittsburgh, Pittsburgh, PA, ³AbbVie, Mettawa, IL

Abstract Body:

Background: Cigarette smoking has detrimental effects on many aspects of human health. Associations between smoking and features of the human epigenome have been described previously, including associations with DNA methylation (DNAm). Most prior studies assessed DNAm in leukocytes, and those studies identified numerous smoking-associated regions (e.g., *AHRR*). In this study, we identify smoking-associated DNAm features in non-blood tissue types, tissues that are typically inaccessible in human studies and therefore understudied. **Methods:** We generated DNAm data using the Illumina EPIC array for 929 Genotype-Tissue Expression (GTEx) tissue samples representing 9 tissue types (lung, colon, ovary, prostate, whole blood, breast, testis, kidney, and muscle). For each tissue type, the association of smoking status (ever/never) with DNAm (and existing gene expression data) was estimated using linear models, adjusting for age, sex, BMI, race/ethnicity, ischemic time, and surrogate variables (FDR of 0.05). We conducted gene set enrichment analysis using missMethyl (gometh) and hallmark gene sets from MSigDB (FDR of 0.05). **Results:** We identified 6350 smoking-associated CpGs in lung (n=212) and 2735 in colon (n=210), the majority of which (6219) had not been reported previously. For all 7 other tissue types (sample sizes 46-153), no clear associations were observed (based on FDR); however, CpGs showing the smallest P-values in lung, colon, ovary, and kidney showed clear enrichment for CpGs reported in past studies of smoking and DNAm in leukocytes. For ~1500 loci, smoking was associated with both DNAm and local gene expression (in lung). For smoking-associated loci detected in both lung and colon (e.g., *AHRR*, *CYP1B1*), we observed general consistency across tissues in terms of the location of smoking-associated CpG clusters (and the direction of association), but differences in patterns of association for CpGs within clusters. Seventeen hallmark gene sets were enriched for genes that annotated to smoking-associated CpGs (in lung), with the top 10 sets including a xenobiotic metabolism gene set and 9 cancer-related gene sets (e.g., apoptosis, p53). At least 4 smoking-associated regions in lung were impacted by lung methylation-QTLs that co-localize with GWAS signals for lung function (FEV1/FVC) in UK Biobank. This suggests that smoking-related epigenetic alterations may mediate the effects of smoking on lung health. **Conclusion:** A multi-tissue approach for studying the effects of smoking on DNAm has identified novel smoking-associated regions in disease-relevant tissues, inducing effects that are both tissue-specific and shared across tissues.

Epigenetics Posters - Wednesday

PB2473. The interplay of epigenetic, genetic, and traditional risk factors on blood pressure: Findings from the Health and Retirement Study.

Authors:

X. Zhang, F. Ammous, L. Lin, S. Ratliff, W. Zhao, S. Kardia, J. Smith; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Elevated blood pressure (BP) is a major public health burden and a leading cause of morbidity and mortality. BP is complex phenotype often resulting from an interplay between genetics and the environment. Little is known about how the epigenome, a cellular mechanism that regulates gene expression and is modifiable by the environment, interacts with traditional or genetic risk factors to influence BP. In this study, we leveraged 13 previously reported DNA methylation sites (CpGs) that have been significantly associated with BP and evaluated their association with systolic and diastolic BP, individually and aggregated into methylation risk scores (MRS) in 3,070 participants (437 African Americans (AA), 2,021 European Americans (EA), and 612 of other ancestries) aged ≥ 50 years from the Health and Retirement Study (HRS). We also evaluated interactions between the MRS, and genetic risk score (GRS), age, sex, and education on BP. We performed both full sample (trans-ancestral) and ancestry-specific analysis (for EA and AA). Using linear regression models adjusted for age, sex, education, parental education, type 2 diabetes, smoking, alcohol use, exercise, and genetic PCs we detected significant associations with either systolic or diastolic BP ($P < 0.05$) in 10 of the 13 CpGs in the full sample analysis. MRS_{SBP} was a significant predictor of SBP in the full sample ($P = 2.6 \times 10^{-5}$) and EA ($P = 0.001$) with adjusted effect estimates of 1.7 (full sample) and 1.6 (EA) mmHg increase in SBP for each 1 standard deviation (SD) increase in the risk score. MRS_{DBP} was also associated with DBP in the full sample ($P = 1.8 \times 10^{-6}$) and in ancestry-specific analysis (EA, $P = 7.2 \times 10^{-5}$; AA, $P = 0.033$), with adjusted effect estimates ranging between 1.1 (full sample and EA) and 1.4 (AA) mmHg increase in DBP for each 1 SD increase in MRS_{DBP} . After further adjustment for GRS, both MRS_{SBP} and MRS_{DBP} remained significant predictors of BP in EA. In the full sample and EA, the effects of MRS_{SBP} and MRS_{DBP} on BP decreased as age increased (all $P_{interaction} < 0.05$). For the full sample, this was equivalent to an effect estimate of 2.9 vs. 0.95 mmHg increase for each 1 SD increase in MRS_{SBP} and 1.9 vs. 0.65 mmHg increase for each 1 SD increase in MRS_{DBP} at the 25th and 75th percentile of age, respectively. In AA, MRS_{SBP} was associated with higher SBP in females but not in males ($P_{interaction} = 0.01$). We noted no significant interactions between the MRS and education nor the GRS. In summary, we show that MRS_{SBP} and MRS_{DBP} were significant predictors of BP after accounting for genetic and traditional risk factors, with evidence of interactions with age and sex and variation by genetic ancestry.

Epigenetics Posters - Thursday

PB2474. The role of genetic and epigenetic mechanisms in the definition of the circular RNA landscape in multiple sclerosis.

Authors:

E. Paraboschi^{1,2}, F. Airi², G. Liberatore³, G. Soldà^{1,2}, E. Nobile-Orazio^{3,4}, R. Asselta^{1,2}; ¹Humanitas Univ., Milan, Italy, ²IRCCS Humanitas Res. Hosp., Milan, Italy, ³Neuromuscular and Neuroimmunology Service, IRCCS Humanitas Res. Hosp., Milan, Italy, ⁴Univ. of Milan, Milan, Italy

Abstract Body:

Alternative splicing (AS) is a post-transcriptional mechanism that increases the information content of the transcriptome through the expression of different mRNAs from single genes. AS directly competes with the backsplicing (BS) process, a mechanism that leads to a non-canonical splicing of genes to generate non-coding, circular RNAs (circRNAs). AS and BS processes are intertwined, since their functions are based on the same spliceosomal machinery, and they are tuned by the same trans-acting factors. A dysregulation of AS/BS has been implicated in human disorders, including multiple sclerosis (MS), a chronic autoimmune neuroinflammatory and neurodegenerative disease in which autoreactive T lymphocytes attack antigens of the central nervous system.

The aim of this project is to unravel the mechanisms involved in the dysregulation of AS/BS processes in MS by integrating different omics data.

We decided to specifically evaluate the BS profile in CD4⁺/CD8⁺ T lymphocytes, considering their essential role in the disease pathogenesis, and recruited 20 MS patients and 20 controls. A first analysis conducted on CD4⁺ T lymphocytes showed 176 differentially expressed circRNAs (P<0.05).

Interestingly, they derive from genes implicated in chromatin organization, neuroinflammation, and regulation of gene expression by epigenetic mechanisms, thus highlighting the importance of epigenetic features in circRNAs determination. Then, given the differential expression of circRNAs in MS patients, and the existence of a genotype-dependent regulation, we explored a possible role of the genetic background on circRNAs modulation, focusing on known MS-associated genetic variants. We evidenced the existence of significant QTLs, thus suggesting possible correlations between the genetic background and circRNA expression profiles.

In conclusion, we described the circRNA expression profile of specific T lymphocyte classes in MS patients, suggesting that MS-associated variants may tune the expression levels of circRNAs acting as circ-QTLs. The analysis of the BS landscape in CD8⁺ T lymphocytes, as well as the correlation with the methylation profile is currently being performed. All omics data will be integrated in a unified framework to infer regulatory relationships, and to gain a detailed insight on the cellular states characterizing the disease. This approach will unmask the biological processes that are driving the development of the disease phenotype and that would not be observable if the single omics were analyzed separately.

Acknowledgments. This work is supported by Fondazione Regionale per la Ricerca Biomedica (FRRB), Early Career Award.

Epigenetics Posters - Wednesday

PB2475. Tobacco smoke-derived metabolites in plasma are associated with cardiometabolic traits, are genetically regulated, and may modulate adipose tissue DNA-methylation in African Americans.

Authors:

S. Das, M. Comeau, C. Langefeld; Wake Forest Univ. Sch. of Med., Winston-Salem, NC

Abstract Body:

Genetic factors influence susceptibility to cardiometabolic diseases. However, lifestyle factors including smoking, are important environmental risk factors for these diseases. Interestingly smoking-cessation leads to weight gain. Genetic factors influence smoking behavior and addiction, and metabolic pathways may determine the systemic effects of smoking by influencing the level of tobacco smoke-derived metabolites. As compared to self-reported smoking, plasma levels of tobacco-derived metabolites provide empirical and accurate evaluation of chronic smoking or exposure to tobacco smoke. Untargeted LC-MS profiling (Metabolon) of fasting plasma in 253 non-diabetic African Americans from AAGMEx cohort quantified 1124 metabolites, including 124 xenobiotics, and allowed us to evaluate the role of environmental exposures, including smoking in determining gluco-cardio-metabolic traits. Three tobacco metabolites: cotinine, hydroxycotinine, and 2-hydroxyfluorene sulfate were associated with obesity and insulin sensitivity. In linear regression analyses adjusted for age, sex, and admixture, hydroxycotinine level was inversely and directly associated with BMI ($\beta = - 5.09$, $p = 7.88 \times 10^{-7}$), and insulin sensitivity ($\beta = 3.53$, $p = 5.14 \times 10^{-4}$), respectively. Genome-wide association analysis identified significant quantitative trait loci for these metabolites. SNP rs2835863_G in intron of *KCNJ6* ($P = 2.05 \times 10^{-7}$), and rs7609906_A in intron of *FAMI9A4* ($P = 1.5 \times 10^{-7}$) was most significantly associated with hydroxycotinine and cotinine, respectively. Previous studies suggest that *KCNJ6* variants associate with susceptibility to nicotine dependence/failure of smoking cessation, and *FAMI9A4* is involved in addiction behavior. Altered DNA-methylation is an epigenetic change that reflects exposure to lifestyle factors including cigarette smoking. We identified association ($p < 0.0001$) between plasma hydroxycotinine level with DNAm levels of 75 CpG sites in adipose tissue. These CpG sites were annotated to 48 genes, including 5 sites in *AHRR* gene reported previously in multiple EWAS's for smoking. Thus plasma levels of tobacco smoking-derived metabolites are under genetic regulation, and these metabolites are involved in smoking-induced altered DNAm in adipose tissue of African Americans.

Epigenetics Posters - Thursday

PB2476. Transcriptome and chromatin accessibility dynamics across 25 brain regions identify novel susceptibility gene sets for neuropsychiatric disorders

Authors:

P. Dong¹, J. Bendl², J. Fullard³, G. Hoffman³, P. Roussos³; ¹Icahn school of Medicine at Mount Sinai, New York, NY, ²Icahn Sch. of Med. at Mount Sinai, New York City, NY, ³Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Transcriptome and epigenome dynamics across brain regions and cell types are strongly associated with neuropsychiatric disease. We profiled chromatin accessibility and gene expression in two cell types (neurons and non-neurons) across 25 distinct brain regions. We found extensive molecular features and regulome alterations across brain regions. Leveraging a system biology approach, we identified brain region-specific chromatin co-accessibility and gene-coexpression modules that are robustly associated with neuropsychiatric disease. The genes associated with a promoter module are enriched for regulatory, instead of synaptic, function but are strongly overrepresented for common variants associated with different neuropsychiatric disorders, including Autism spectrum disease, bipolar disorder, and schizophrenia, as well as rare variants and fine-mapped gene sets. After conditioning using schizophrenia-associated cell types, the association with schizophrenia common variants is still highly significant. Similarly, the module still overrepresents common variants following the removal of all known schizophrenia-associated pathways. Our analysis highlights the function of non-synaptic genes and the polygenic basis of neuropsychiatric disorders.

Epigenetics Posters - Wednesday

PB2477. Transfer learning of a methylation based COVID19 severity model for DNA viral infection by eczema herpeticum.

Authors:

W. Zhou¹, M. P. Boorgula², M. Campbell², S. Chavan², G. Harrison¹, B. R. Peterson¹, N. Rafaels², B. Barnes³, R. PORECHA³, R. Mathias⁴, A. Taye³, K. Barnes^{1,2}; ¹Tempus Labs, Chicago, IL, ²Univ. of Colorado, Denver, CO, ³Illumina, Inc., La Jolla, CA, ⁴Johns Hopkins Univ., Baltimore, MD

Abstract Body:

Methylation risk score (MRS) is an algorithmic aggregation of methylation states that are increasingly used to infer clinical phenotypes. An advantage of MRS is that it captures multi-factorial predispositions as methylation states are influenced by both genetic and environmental factors. Whether MRS can be universally applied across closely relevant phenotypes remains unexplored. Additionally, factors to be taken into account for reaching a satisfactory model generalization are undetermined. DNA and RNA viruses have evolved to utilize epigenetic modifications including DNA methylation as a means to maximize viral gene expression and to escape host immunity during infection. Previously, we customized Illumina's MethylationEPIC array to enhance immune response detection. We profiled peripheral blood samples from patients infected with SARS-CoV-2 (a RNA virus) thereby developing a COVID-19 MRS model to predict disease severity. In the current study, we aimed to assess the utility of COVID-19 MRS model in DNA viral infections. The clinical outcome of focus was atopic dermatitis (AD)-associated eczema herpeticum (ADEH), a severe complication in AD caused by disseminated herpes simplex virus (HSV, a double-stranded DNA virus). We profiled peripheral blood DNA samples with the customized MethylationEPIC array from 258 ADEH+ and 100 ADEH- patients with disease severity measures. We further compared data to the COVID-19 positive cases and negative controls from recently reported study (Konigsberg et al 2021, Comm. Med.). Firstly, we demonstrated the COVID-19 MRS model accuracy on ADEH severity prediction. We then stratified model performance based on cohort characteristics to understand contributing factors that influence model transferability and generalizability. Next, we trained a separate MRS model for specificity to ADEH cases alone. Both models were compared to quantify specificity of methylation patterns induced by different viral types. In summary, our study reveals the occurrence of both common and differential methylation states in infectious diseases caused by different viral classes. In addition, our study provides evidence to delineate host DNA modifications utilized by unique viruses to cause infections in humans. Overall, this work serves as a case study to evaluate the validity and utility of MRS in its implementation towards infectious disease healthcare.

Epigenetics Posters - Thursday

PB2478. Transposable elements are associated with the variable response to influenza infection

Authors:

X. Chen¹, A. Pacis², K. Aracena³, S. Gona³, T. Kwan⁴, C. Groza⁵, Y-L. Lin³, R. H. M. Sindeaux⁶, V. Yotova⁶, A. Pramatarova⁴, M-M. Simon⁴, T. Pastinen^{7,8}, L. Barreiro^{3,9,10}, G. Bourque^{1,2,4,8}; ¹Inst. for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto Univ., Kyoto, Japan, ²Canadian Ctr. for Computational Genomics, McGill Univ., Montreal, QC, Canada, ³Dept. of Human Genetics, Univ. of Chicago, Chicago, IL, ⁴McGill Genome Ctr., Montreal, QC, Canada, ⁵Quantitative Life Sci., McGill Univ., Montreal, QC, Canada, ⁶Ctr. de Recherche, CHU Sainte-Justine, Université de Montréal, Montreal, QC, Canada, ⁷Genomic Med. Ctr., Children's Mercy Hosp. and Res. Inst., Kansas City, MO, ⁸Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, ⁹Section of Genetic Med., Dept. of Med., Univ. of Chicago, Chicago, IL, ¹⁰Committee on Immunology, Univ. of Chicago, Chicago, IL

Abstract Body:

Influenza A virus infections are frequent every year and result in a range of disease severity between individuals. The variability in the human response is a combination of viral- and host-factors but many of the underlying mechanisms remain largely elusive. Given that transposable elements (TEs) have been shown to contribute to the activation of innate immunity, we wanted to explore their potential role in this variability. We profiled the transcriptional program before and after influenza infection using RNA-seq in monocyte-derived macrophages from 39 human individuals and observed significant inter-individual variability in viral load post-infection (from 3.8% to 57.8% of RNA reads). Using the same cells, we profiled the epigenetic landscape using ATAC-seq and ChIP-seq and identified a set of TEs with either enhanced (37 families) or reduced (39 families) accessibility in response to infection. Notably, of the enhanced families, 15 showed high variability in chromatin accessibility changes between individuals. We showed that enhanced and reduced TE families were enriched for various histone modifications (H3K27ac, H3K4me1, H3K4me3 or H3K27me3) and in proximity to genes that are regulated in the response to infection. Motif analysis revealed an association with known immune regulators in families stably enriched across individuals and, in contrast, an enrichment of putatively novel host factors in highly variable families, including ZNF460 and other KRAB-ZNFs. Finally, we observed an association between TE transcripts levels before infection and viral load post infection. Using this information combined with the new host factors identified, we built a predictive model for viral load post infection suggesting that TEs, and host factors regulating TEs, contribute to the variable response to infection.

Epigenetics Posters - Wednesday

PB2479. Trinity of chromatin architects - The coordination of CTCF, RNAPOL2 and Cohesin in shaping the genomic landscape

Authors:

A. Agarwal¹, D. Plewczynski^{1,2}, M. Chilinski²; ¹Ctr. Of New Technologies, Univ. Of Warsaw, Warsaw, Poland, ²Warsaw Univ. of Technology, Warsaw, Poland

Abstract Body:

Genome-wide architectural landscapes of chromatin in the nucleus can be identified by the advanced high-throughput sequencing-based 3C-type methods such as Hi-C, ChIA-PET, and HiChIP. The spatial organisation of chromatin in the nucleus is stabilised by structural proteins, such as the CCCTC-binding factor (CTCF), RNAPOL2 and cohesin complex. These proteins play an essential role in establishing long-range chromatin interactions (chromatin loops), facilitating topologically associating domain formation and allowing for the coordination of genes with their corresponding regulatory elements. Here, we discuss the exact role of CTCF, RNAPOL2 and cohesin in shaping chromatin multiscale three-dimensional architecture, in particular how static architecture defined by CTCF is re-shaped by the dynamical activity of cohesin (LEM: loop extrusion model), and re-organized during transcriptional activity by RNAPOL2. We analyse in detail CTCF, cohesin and RNAPOL2 binding sites that account for the topological regulation of chromatin loops, the dynamics of loop extrusion, and phase separation condensates related to the transcriptional factories.

Epigenetics Posters - Thursday

PB2480. Uncovering Novel Functions of Histone Demethylase KDM5 Through a Genome-wide Approach

Authors:

M. Yheskel¹, J. Cadiz¹, S. Sidoli¹, J. Secombe²; ¹Albert Einstein Coll. of Med., Bronx, NY, ²Albert Einstein Coll. of Med., New York, NY

Abstract Body:

Mutations throughout human histone lysine demethylase KDM5C are associated with intellectual disability (ID). KDM5C's main canonical function is the removal of H3K4me3, a histone mark associated with transcriptional activation. Recent studies have shown that KDM5 has both demethylase-dependent and independent functions in neurons. Nevertheless, little is known about the molecular consequences of ID-associated mutations in KDM5. Thus, we have modeled patient KDM5C ID-associated mutations in *Drosophila melanogaster* to understand how they affect KDM5 binding, demethylase activity, transcriptional regulation, and protein binding. Here, we assay genome-wide KDM5 binding and local changes to H3K4me3, and how these mutations affect transcription of mRNA. Furthermore, we use the novel technique of proximity labeling to reveal that KDM5 associates with insulator proteins that help partition the genome in 3D-space to ensure proper regulation of genes. Key among these insulator proteins is CTCF, the sole insulator found in mammals which has been shown to result in ID when mutated. KDM5 extensively co-localizes with insulator proteins and topologically associated domain (TAD) boundaries. Interestingly, TAD boundaries are found near housekeeping genes such as ribosomal protein genes (RPGs); a class of gene that is transcriptionally downregulated ID-mutant strains. Together, these analyses reveal that ID-associated KDM5 mutations may impart their deleterious effects through dysregulation of genome organization in a conserved manner.

Epigenetics Posters - Wednesday

PB2481. Using allele-specific expression to assess intra- and inter- individual heterogeneity in X-chromosome inactivation within human tissues

Authors:

H. Markus¹, B. Jiang², L. Carrel³, D. Liu²; ¹Penn State Coll. of Med., Hershey, PA, ²Pennsylvania State Univ. Coll. of Med., Hershey, PA, ³Penn State Univ Col Med, Hershey, PA

Abstract Body:

Functional genomics and disease association studies of X-linked genes remains understudied due to the unique biology of X chromosome. In humans, males carry one copy of the X-chromosome while females carry two. X-chromosome inactivation (XCI) epigenetically silences one copy of the X in females to maintain balanced gene dosage with males. Yet up to 10% of genes escape XCI and are transcribed from both Xs. Another 30% of genes variably escape XCI in a subset of tissues or individuals. Understanding how XCI states vary between individuals and tissues will help elucidate the contribution of X-linked genes to human diseases and pinpoint causal tissue types. We recently developed a statistical method XCIR to identify XCI escape genes from RNA-seq data. XCIR estimates sample skewing (i.e., the fraction of cells where a particular X chromosome is inactivated) by examining allele-specific expression (ASE) from a gene set known to be silenced by XCI. XCI states are then inferred for each X gene and concluded to escape XCI if the gene ASE is statistically significantly more balanced than sample skewing. Using XCIR we assessed XCI states in samples from Genotype Tissue Expression (GTEx) Project (v8p release). We limit our analysis to samples with XCI skewing > 70:30, which provides XCIR sufficient power to infer XCI states per gene within a sample. In total we assess XCI states for 542 X-linked genes in 1,775 transcriptomes from 274 female donors across 44 tissues and infer XCI states for all genes with transcribed well-expressed SNPs. Across all tissues/samples, 78% of our inferred XCI states agreed with previous male-female differential gene expression results across genes and tissue types. In addition, we provide XCI states for 2,686 gene and tissue pairs not previously described among which 1,421 are from female-specific tissues. We also annotate 39 X-linked genes not previously described. Among these genes, ARSD-AS1 escaped XCI and PNMA6F, RAP2C-AS1, PLXNB3-AS1, DIAPH2-AS1 and LINC00106 variably escaped in more than 1 tissue. Lastly, using a method similar to the analysis of variance, we examine how the XCI states vary between tissue types in the same individual, and between different individuals of the same tissue type. Using paired t-test, we found XCI states exhibit much higher variability between different individuals across tissue types (p-value < 2.2x10⁻¹⁶), pointing to possible influence of genetic variation. Overall, we provide updated and accurate XCI states of genes based on ASE across human tissues. We believe these results will increase our understanding of the extent of intra- and inter- individual variability of XCI and help us study its effect on sex-biased diseases.

Epigenetics Posters - Thursday

PB2482*. Using iPSC-derived microglia and oligodendrocytes to study ancestry-specific gene expression in the context of Alzheimer's Disease (AD).

Authors:

S. Moura¹, A. Ramirez¹, J. Shepherd¹, S. Simon¹, M. Muniz¹, P. Whitehead¹, L. Adams¹, T. Starks², P. Mena¹, M. Illanes-Manrique³, S. Tejada¹, A. Griswold¹, J. Young¹, G. Byrd⁴, M. Cornejo-Olivas⁵, B. Feliciano-Astacio⁶, M. Pericak-Vance¹, J. Vance¹, D. Dykxhoorn¹; ¹Univ. of Miami, Miami, FL, ²Wake Forest Sch. of Med., Winston-Salem, NC, ³Neurogenetics Res. Ctr., Inst. Natl. de Ciencias Neurologicas, Lima, Peru, ⁴NC A&T State Univ, Greensboro, NC, ⁵Inst. Natl. de Ciencias Neurologicas, Lima, Peru, ⁶Univ. Central del Caribe, Bayamón, Puerto Rico

Abstract Body:

Genome-Wide Association Studies (GWAS) studies in AD have been performed mainly in Non-Hispanic Whites (NHW). However, GWAS studies in admixed populations such as (African and NHW ancestries) and Hispanics (African, NHW, and Amerindian ancestries) are now underway. We have shown that different ancestry groups carry distinct genomic architectures that could lead to differing genetic susceptibility to AD. Thus, to study the functional mechanisms of GWAS studies in these admixed populations, we are identifying the impact of ancestry-specific genetic variants on the regulome and transcriptome across different cell types of the central nervous system. As part of this study, we analyzed the regulatory maps of iPSC-derived microglia (iMG) and oligodendrocytes (iOLs) derived from individuals with high global African (AF), Amerindian (AI), or NHW ancestry. iPSCs lines were derived from AD patients and non-cognitively impaired (NCI) controls which have >90% genomic content from different ancestral populations - Non-Hispanic Whites, African, and Amerindian. These iPSC lines were validated for pluripotency and chromosomal stability and were differentiated into MG and OLs. These differentiated cell populations were validated via immunocytochemistry (ICC) and qRT-PCR for the expression of cell type-specific markers. Bulk ATAC-seq, Hi-C, and RNA-seq were performed to study the regulatory architecture of the genomes from these cells in both case and control lines and their associated impact on gene expression. We optimized the differentiation of iMG and iOLs and validated it using ICC and qRT-PCR for lineage-specific markers. We analyzed and compared the regulatory maps in AF and AI ancestries by aligning chromatin accessibility (ATAC-seq) and promoter-enhancer interaction (Hi-C) data in iMG and iOLs. Moreover, we investigated the downstream effects the regulatory architecture might have on transcription (RNA-seq) specific for these cell types and ancestral backgrounds. This is one of the first studies to provide ancestry-specific regulatory data on the ancestral backgrounds that contribute to a large part of the US population. Our data provide new insights into ancestry-specific risk factors and candidate genes in AD pathophysiology, specifically, we report novel data on chromatin accessibility and transcriptomics data in AF and AI iPSC-derived microglia and oligodendrocytes, with the latter not having been a main research focus in AD pathology.

Epigenetics Posters - Wednesday

PB2483. Using machine learning to predict 3D genome organization across thousands of diverse individuals reveals conservation and population differentiation.

Authors:

E. Gilbertson¹, E. McArthur², C. Brand¹, D. Rinker², J. Capra¹; ¹Univ. of California San Francisco, San Francisco, CA, ²Vanderbilt Univ., Nashville, TN

Abstract Body:

A central goal of human population genetics is to quantify all aspects of genotypic and phenotypic variation across diverse populations, including human variation in genome folding. Quantifying human genetic and phenotypic variation in diverse populations is necessary to enable equitable science and medicine. DNA sequence variants that influence genome folding can have massive regulatory impacts on disease by altering chromatin interactions. However, current methods for interpreting the effect of non-coding variants do not consider this mechanism because we do not understand how 3D folding relates to function and how genetic variants affect this critical function. Understanding the breadth of chromatin contact diversity across human populations is critical for interpreting 3D genome variants, yet it is prohibitively costly to experimentally determine chromatin-interactions at a population-scale. Here, we use machine learning to predict 3D chromatin organization from genome sequence and map the diversity of 3D folding across thousands of modern humans. To explore the variation of human 3D folding patterns, we derive a metric for quantifying population-level 3D genome differentiation motivated by the sequence-based F_{ST} . Applying the 3D F_{ST} genome-wide, we identify regions with significant differentiation in contact maps between human populations that may contribute to phenotypic differences between populations. We find that 3D folding similarity genome-wide follows patterns of sequence divergence and that pressure to maintain 3D folding patterns has broadly constrained chromatin-interacting sequence evolution. However, we also identify loci with significant variation in 3D genome organization that associates with observable expression differences and phenotypic diversity. In summary, we infer 3D genome folding maps across diverse human populations and quantify the variational and functional implications of diverse folding patterns. We anticipate that these maps of 3D folding diversity will provide a reference for future work on the interpretation of 3D genome folding variation across human populations.

Epigenetics Posters - Thursday

PB2484. Validation and Methylome analysis of the CRISPR-edited DUX4 locus in immortalized myoblast cell lines modeling FSHD using long-read sequencing.

Authors:

J. Sakr, X. Kong, N. Nguyen, K. Yokomori, A. Mortazavi; Univ. of California, Irvine, Irvine, CA

Abstract Body:

Facioscapulohumeral muscular dystrophy (FSHD) is one of the three most common forms of muscular dystrophy in adults. The major form of FSHD (FSHD1) is associated with contraction of D4Z4 macrosatellite repeats on chromosome 4q35 with loss of DNA methylation in haplotypes with a distal “permissive” specific SNP on the last D4Z4 repeat that creates a polyadenylation site for the transcription factor DUX4, which is then misexpressed in muscle. The chromosome 10q subtelomere contains an almost identical repeat array but contractions on this locus are non-pathogenic. These highly homologous repeats on both chromosomes make modeling the disease and characterizing the mechanisms especially challenging with short-read sequencing. We assessed the contraction status of the D4Z4 repeat of several FSHD model cell lines using the Oxford Nanopore cas9 kit to selectively sequence the genome region associated with the disease. By using long reads, we are able to capture complete contracted D4Z4 repeat region and the distal permissive allele in a single read. Furthermore, by directly sequencing the genome, we are able to study the epigenetic signatures of the disease. We are further analyzing the methylation status in the contracted repeats to assess the extent of methylation loss and its correlation with Dux4 and Dux4-target expression in order to select the most representative immortalized cell models of FSHD1.

Epigenetics Posters - Wednesday

PB2485. Variable *cis*-regulatory landscapes associate with gene function and expression patterns

Authors:

M. Benton¹, J. Capra², D. M. Ruderfer³; ¹Baylor Univ., Waco, TX, ²Univ. of California San Francisco, San Francisco, CA, ³Vanderbit Univ. Med. Ctr., Nashville, TN

Abstract Body:

Motivation: *Cis*-regulatory elements (CREs) regulate gene expression by binding transcription factors and controlling transcription across diverse cell types. Multiple CREs can cooperate to regulate the expression of a single gene, and individual CREs can have multiple gene targets. However, despite evidence that the groups of CREs associated with genes—the CRE landscapes—are ubiquitous, we do not understand how CRE landscape variability across genes relates to the strength and tissue-specificity of expression for those gene targets. Furthermore, genetic variation in CREs has been implicated in the etiology of complex disease, yet, current strategies do not consider the impact of CRE variation in the context of the CRE landscape. **Results:** Here, we quantify the CRE landscape composition across ten human tissues and relate landscape attributes to the expression of genes in the landscapes. We integrate three-dimensional chromatin conformation data with functional and evolutionary characterization of human CREs to assess CRE landscape attributes such as the number of active CREs, DNA sequence conservation, and the tissue-specificity of the component CREs. For example, we find that expressed genes have larger CRE landscapes than non-expressed genes in the same tissue and that genes with an increased proportion of tissue-specific elements are more likely expressed in specific tissues. We also find that CRE landscapes with genes under constraint on their expression (e.g., loss-of-function intolerant and housekeeping genes) have more evolutionarily conserved sequences than expressed genes overall. We hypothesized that genes with large CRE landscapes would be more robust to changes in expression due to the potential for CRE redundancy within the landscape. However, we do not observe a strong relationship between the size of the CRE landscape and expression variability across individuals. Nonetheless, we do observe a relative depletion for overlap with eQTL in larger CRE landscapes. **Conclusions:** Overall, this work highlights how differences in gene function, expression, and evolutionary constraint are reflected in the features of their CRE landscapes. Thus, considering the CRE landscape of a gene is an essential component to understanding gene expression dynamics across biological contexts. In the future, quantifying the effects of CRE alteration in the context of these CRE landscapes will facilitate interpretations of the effects of gene regulatory perturbations to disease risk.

Epigenetics Posters - Thursday

PB2486. Variation in DNA methylation is associated with cognitive function in post-surgery breast cancer patients prior to adjuvant therapy.

Authors:

S. Liu¹, D. Liu², C. M. Bender³, K. I. Erickson⁴, S. M. Sereika³, Y. P. Conley^{3,1}, J. R. Shaffer^{1,5}, D. E. Weeks^{1,6}; ¹Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, ²Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, ³Sch. of Nursing, Univ. of Pittsburgh, Pittsburgh, PA, ⁴Dept. of Psychology, Kenneth P. Dietrich Sch. of Arts and Sci., Univ. of Pittsburgh, Pittsburgh, PA, ⁵Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, PA, ⁶Dept. of Biostatistics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA

Abstract Body:

Background: Compared to age-matched controls, up to 33% of women with breast cancer have poorer cognitive function before they begin adjuvant therapy. However, the underlying biological basis remains unclear. The purpose of this study was to explore the role of DNA methylation (DNAm) on four objective (*i.e.*, concentration, attention, mental flexibility, and executive function) and one subjective (*i.e.*, total scores from the Patient Assessment of Own Functioning Inventory) cognitive function phenotypes in postmenopausal women diagnosed with early-stage hormone-receptor positive breast cancer after breast-conserving surgery and prior to initiating aromatase inhibitor (AI) therapy.

Methods: Genome-wide DNAm data were generated from blood samples in 63 individuals using Infinium MethylationEPIC Beadchips. 700,772 CpGs and 62 samples passed quality control checks and were then used for analysis. An epigenome-wide association study (EWAS) that evaluated CpG-specific associations was conducted for each cognitive domain phenotype adjusting for age, verbal IQ score, and global patterns of DNAm. The dmrff R package was then applied to identify differentially methylated regions (DMR) using EWAS summary statistics.

Results: The samples were primarily in stage 0 and I of breast cancer with a mean age of 64 and a mean verbal IQ score of 112. No CpGs met the genome-wide significance threshold of 9×10^{-8} in EWAS of the four objective cognitive domains, but several significant DMRs with Bonferroni-adjusted p-value less than 0.05 were identified, a few of which were known to be associated with cancer or tumorigenesis. Two DMRs were identified for concentration in/near *RGS12* and *HMG3*. Three DMRs in/near *USP10*, *TRAPPC5*, and *PIK3IP1* were identified for attention. Two DMRs were found in/near *TRAV2* and *MUCL1* for mental flexibility. For subjective cognitive function, higher mean DNAm levels of cg13769506 in the promoter region of the tumor suppressor gene *BNIP3L* were significantly associated with better cognitive function with p-value of 2.8×10^{-9} ; in addition, five DMRs were significant after Bonferroni-adjustment which mapped to or near the genes *UVRAG-DT*, *MAGOH*, *MSRB1*, *ACCN3*, and *CIQL2*.

Conclusions: Our results suggest that epigenetic changes in breast cancer patients may be associated with cognitive function even before receiving endocrine therapy. This study lays the foundation for future research to assess association between DNAm levels and cognitive function in a later time point in context of AI therapy and/or an exercise intervention which may provide insights on the interplay of DNAm, treatment, and exercise on cognitive function in women with breast cancer.

Epigenetics Posters - Wednesday

PB2487. *Variation in miRNAs may contribute to orient future therapeutic strategies in hidradenitis suppurativa; An epigenome-widescreening approach*

Authors:

V. Aaren¹, U. Radhakrishna², D. Jhala³, N. Saiyed², G. Damiani⁴; ¹Andhra Univ., Visakhapatnam, India, ²Oakland Univ. William Beaumont School of Med., Royal Oak, MI, ³Sch. of Sci., Gujarat Univ., Ahmedabad, India, ⁴Istituto Ortopedico Galeazzi, Milan, Italy

Abstract Body:

Background: Hidradenitis suppurativa (HS) is a chronic inflammatory disease, influenced by non-genetic factors that modulate the expression of miRNAs. Currently, no miRNA data are available for HS. Thus, we aimed to identify miRNA gene methylation profiles associated with HS susceptibility. **Objectives:** Identify miRNA gene methylation profiles associated with HS susceptibility. This study examined the whole genome-wide methylation patterns of DNAs from 24 healthy controls and 24 patients with HS using Illumina Infinium MethylationEPIC BeadChip array analysis. The differentially methylated miRNAs were analyzed using Ingenuity Pathway Analysis (IPA) to explore the inversely correlated pathways regulated by miRNAs. **Results:** In the study, 54 and 6 CpGs in HS patients were respectively significantly hypomethylated and hypermethylated compared to controls. Some of these CpGs were found to be critical for skin function, such as mir-29, miR-200, mir-205, mir-548, and mir-132. The miR-200c gene was identified as a vital determinant in regulating skin repair after injury and may contribute to age-associated alterations in wound repair. miR-132 was significantly upregulated during the inflammation phase of wound repair, enhancing the activity of *STAT3* and *ERK* pathways that promote keratinocyte proliferation. **Conclusions:** Many of the microRNA genes that were epigenetically altered are known to be involved in wound healing, inflammation, keratinocyte proliferation, and modulation of wound tissues. This is the first study to analyze methylation profiles of miRNA genes in the HS population. This report highlights the unique role that miRNAs could play in the both diagnosis and treatment of HS.

Epigenetics Posters - Thursday

PB2488. Vitamin c contributes to epigenetic regulation of genes related to diabetic retinopathy in retinal endothelial cells.

Authors:

D. Sant¹, J. Reynolds¹, E. Turner¹, T. Alger¹, N. Islam¹, V. Manukyan¹, A. Sheppert¹, G. Wang²; ¹Noorda Coll. of Osteopathic Med., Provo, UT, ²Univ Miami, Miami, FL

Abstract Body:

Diabetic retinopathy is a complication of diabetes that leads to irreversible vision loss and remains the leading cause of blindness in working-age adults. It is primarily a complication of microvascular endothelial cells in the retina and the breakdown of the inner blood-retinal barrier. Vitamin c passes through the blood-retinal barrier through GLUT1 transporters, which also transport glucose. Glucose acts as a competitive inhibitor to vitamin c, and hyperglycemia leads to a local vitamin c deficiency in the eyes of diabetics. Vitamin c acts as a cofactor for the TET enzymes, and a local deficiency leads to impaired DNA methylation-demethylation dynamics.

Primary, human retinal endothelial cells were cultured in 15.7 mM glucose (283 mg/dL) either in the presence or absence of vitamin c (50 µM) for five days. Immunofluorescent staining showed a robust increase in global hydroxymethylcytosine levels, consistent with previous reports in other cell types. Whole transcriptome sequencing (RNA-seq) was performed, and differential expression was determined using a combination of DESeq2 and edgeR. RNA-seq revealed that 437 genes were found to have upregulated transcription and 308 genes were found to have downregulated transcription.

Pathway analysis highlighted changes in genes related to leukocyte adhesion, tethering and rolling. Inflammation is an early characteristic of diabetic retinopathy that precedes measurable changes in vasculature or vision. In particular, vitamin c treatment caused a reduction in transcription of *SELP*, the gene that codes for P-selectin. GWAS studies have found that mutations in *SELP* are associated with diabetic retinopathy and plasma levels have found an association of P-selectin levels in the plasma and severity of diabetic retinopathy. Vitamin c-induced reduction in *SELP* was confirmed by qPCR in a second batch of cells. Glutathione, an antioxidant stronger than vitamin c, failed to reduce *SELP* transcription, indicating that the reduction is likely due to the epigenetic effects of vitamin c treatment. These data suggest that local vitamin c deficiencies in the eyes of diabetics affect transcription of genes related to leukocyte adhesion and may contribute to the progression of diabetic retinopathy.

Epigenetics Posters - Wednesday

PB2489. Whole genome bisulfite sequencing identifies methylation markers in cord blood linked to maternal hyperglycaemia and childhood metabolic abnormalities

Authors:

Y. Li¹, K. K. Wong², C. H. Tam^{2,3}, C. K. Lim^{2,3}, C. C. Wang⁴, X. L. Yang⁵, A. El-Osta⁶, W. L. Lowe⁷, W. H. Tam⁴, K. Y. Yip^{1,8,9}, R. C. Ma^{2,3,9}; ¹Dept. of Computer Sci. and Engineering, The Chinese Univ. of Hong Kong, Hong Kong, China, ²Dept. of Med. and Therapeutics, The Chinese Univ. of Hong Kong, Hong Kong, China, ³Li Ka Shing Inst. of Hlth.Sci., The Chinese Univ. of Hong Kong, Hong Kong, China, ⁴Dept. of Obstetrics and Gynaecology, The Chinese Univ. of Hong Kong, Hong Kong, China, ⁵Dept. of Epidemiology and Biostatistics, Tianjin Med. Univ., Tianjin, China, ⁶Dept. of Diabetes, Central Clinical Sch., Monash Univ., Melbourne, Australia, ⁷Ctr. for Diabetes and Metabolism, Northwestern Univ., Chicago, IL, ⁸Sanford Burnham Prebys Med. Discovery Inst., La Jolla, CA, ⁹Hong Kong Inst. of Diabetes and Obesity, The Chinese Univ. of Hong Kong, Hong Kong, China

Abstract Body:

Hyperglycaemia affects around 1 in 6 pregnancies globally. Offspring of women with hyperglycaemia in pregnancy have an increased risk of diabetes and obesity in later life and epigenetic mechanisms have been implicated. We performed epigenomic profiling in cord blood to identify methylation changes associated with maternal glycaemic status and later risk of metabolic disease in the offspring. We included 84 mother-offspring pairs (including 27 GDM mothers) from the Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) study Hong Kong field centre. All mothers underwent oral glucose tolerance tests (OGTTs) between 24-32 weeks of gestation. Cord blood was collected at the time of birth. Cord blood DNA was subjected to whole-genome bisulfite sequencing using Illumina HiSeq platform at 150PE. After the removal of low-quality bases and adapter sequences, alignment and methylation calling were performed using gemBS. Cell type compositions were estimated using methylCC with cord blood reference. We used linear regression models to assess the association between cord blood DNA methylation and clinical traits of the pregnant mothers and the offspring at follow-up at age 7. Maternal age at pregnancy, maternal pre-pregnancy BMI, offspring sex, gestational age, the estimated cell type proportions and batch information were added as confounders in the regression models. We identified 13 CpG sites (each with $p < 10^{-5}$, adjusted $p < 0.2$) associated with maternal 2-hour glucose levels during OGTT, located within or close to genes including *GBP3*, *AASS*, *LINC01339*, *GPR137B*, *ERO1B*, *SLC12A2*, *FGF1*, *ZNF516* and *PLCB4*. The gene *FGF1* has been shown to regulate glucose levels while the role of other genes in metabolic programming has not been previously demonstrated. We also found 4 significant CpG sites (each with $p < 10^{-8}$, adjusted $p < 0.05$) whose methylation levels in cord blood were associated with follow-up variables in the offspring at age 7, including insulinogenic index and beta-cell function. The most significant CpG site ($p = 2.06 \times 10^{-13}$, adjusted $p = 3.79 \times 10^{-7}$) associated with insulinogenic index is located within the gene *SFII*. We performed functional enrichment analysis of the top genes associated with antenatal maternal glucose, follow-up insulinogenic index and beta-cell function respectively and pathways related to neural development were identified in all the analyses. In summary, our results suggest that cord blood methylation is not only associated with maternal glucose during pregnancy, but also with metabolic traits in the offspring during follow-up.

Epigenetics Posters - Thursday

PB2490. Whole-methylomics reveals differentially methylated genes in blood associated with Late-onset Alzheimer's disease.

Authors:

C. Breen, K. J. Hogan, A. Madrid, L. A. Papale, L. R. Clark, S. Johnson, S. Asthana, S. Keles, R. Alisch; Univ Wisconsin - Madison, Madison, WI

Abstract Body:

Alzheimer's disease (AD) is the most common neurodegenerative disease in adults. Late-onset sporadic AD (LOAD) is characterized by signs and symptoms of dementia that present after the age of 65. While the full etiologies of LOAD remain unknown, recent evidence suggests that environmental influences increase risk of disease in part through interactions with the epigenome. Epigenome-wide association studies (EWAS) of brain tissue report differential DNA methylation in known and newly recognized LOAD genes, thereby underscoring the utility of EWAS in disclosing novel genes and pathways associated with LOAD pathogenesis. As an alternative to the study of donor brain tissues, investigation of DNA methylation in accessible peripheral tissues offers the opportunity to improve LOAD diagnosis and prognosis. Using a comprehensive whole-methylome sequencing approach and existing banked biofluids with phenotypic data from the Wisconsin Alzheimer's Disease Research Center's (WADRC), we profiled more than 25 million CpGs from male and female LOAD and non-LOAD participants (N = 71) and found ~80,000 differentially methylated loci (DMLs; FDR < 0.05), of which 49% are hypermethylated. Filtering these DMLs to those residing in gene bodies or their corresponding promoter regions (2kb upstream of transcription start site), we identify ~2,600 genes with differentially methylated regions (DMRs) of five or more DMLs, all hypomethylated or hypermethylated associated with dementia, including DMRs in KIF25/KIF25-AS1, MBP, PIEZ02, and DOK6 genes, which have been reported by others to directly participate in LOAD methylation, expression, and pathogenesis. Among the top ten significant gene ontology terms associating with DMR-associated genes are several with known roles in LOAD pathology: axogenesis, regulation of GTPase activity, and regulation of neuron projection development. Together, these data provide evidence that DNA methylation levels in blood can be used to develop novel biomarkers useful in clinical settings and suggest a complex relationship between DNA methylation levels in brain and blood.

Epigenetics Posters - Wednesday

PB2491*. Widespread age-associated changes in the chromatin architecture of skeletal muscle cell type populations.

Authors:

K. Moo¹, R. Albanus², A. Varshney¹, P. Orchard¹, N. Manickam¹, M. Laakso³, J. Tuomilehto⁴, T. A. Lakka³, K. L. Mohlke⁵, M. Boehnke¹, L. Scott¹, H. A. Koistinen⁴, F. S. Collins⁶, S. C. J. Parker¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Washington Univ., St. Louis, MO, ³Univ. of East Finland, Kuopio, Finland, ⁴Univ. of Helsinki, Helsinki, Finland, ⁵Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁶NIH, Bethesda, MD

Abstract Body:

Aging progressively alters the regulatory landscape of the epigenome and is associated with increased risk for multiple diseases including cancer, Alzheimer's, and type 2 diabetes. Genome-wide association studies have identified hundreds of genetic signals associated with these aging-related diseases, and the majority of these signals occur in noncoding regions of the genome. Together, these findings suggest that aging-related epigenomic changes are one component of disease risk. Here we analyze skeletal muscle biopsies from 284 Finnish study participants ranging from age 20 to 79 (mean 60) using single nucleus ATAC-seq to identify patterns in chromatin architecture that are associated with age. To find aging-related open chromatin regions in an unbiased way without assuming a linear relationship we used Autoregressive Integrated Moving Average (ARIMA) models. Of the 13 skeletal muscle cell types we identified, we selected the most abundant six (Type 1, 2a, and 2x muscle fiber, endothelial, mesenchymal stem cells, and smooth muscle) for analysis. We identified 983,155 consensus chromatin accessibility peaks across the 13 cell types and focused our analyses on the 149,751 to 289,430 peaks within each of the six selected cell types. Using a generalized linear model, we adjusted the sample-specific peak accessibility for body mass index and technical covariates such as batch. Then, we used the ARIMA modeling framework and found that between 6.2% and 15.5% of tested peaks had an age-associated pattern. The age-associated peaks were assigned to three broad categories (increasing, decreasing, and dynamic) using unsupervised clustering. Breakpoint analysis of these three categories showed that the chromatin landscape is changing at its greatest magnitude around 60, a result that was similar across the cell types. We are currently performing pathway enrichments to determine the underlying molecular and biological processes associated with these changes. These results highlight age-associated epigenomic differences in six skeletal muscle cell types, giving us insight into cellular dynamics across human lifespan.

Molecular Effects of Genetic Variation Posters - Thursday

PB2492. A Catalogue of Transcriptomes and Associated Genetic Effects on 2,000 Qataris Uncovering the Functional Impact of Middle Eastern Genetic Variation and Identifying Novel Pathways Underlying Human Traits and Diseases

Authors:

Y. Mokrab¹, H. Naeem¹, I. Diboun¹, M. Ghorbani¹, R. Razali¹, M. Hashmi¹, R. Mathew¹, V. Mattei¹, N. James¹, L. Mathew¹, L. Silcock¹, K. Wang¹, M. Kalikiri¹, F. Vempalli¹, M. Tamanni¹, A. Khouly¹, S. Poolat¹, T. Zaid¹, A. Akil¹, S. Lorenz¹, R. Al Ali¹, Q. Consortium², K. Fakhro¹, S. Tomei¹, D. Chaussabel¹, S. Montgomery³; ¹Sidra Med., Doha, Qatar, ²Weill Cornell Med.-Qatar, Dept. of Genetic Med., Doha, Qatar, ³Stanford Univ., Stanford, CA

Abstract Body:

Mapping genetic variation to transcriptome activity at a population level is a robust approach to linking regulatory mechanisms to traits and diseases, allowing both association and causality to be inferred. Here, we present a large study of predominantly healthy individuals from the Qatar Genome Project (QGP), combining whole genome sequencing (WGS) (30x coverage; n=6,216) and whole blood RNA-Seq (20-100 M reads; n=2,127) data to build a comprehensive catalogue of genetic variants regulating various transcription traits. Data were processed using cutting-edge bioinformatics pipelines and comparative analysis revealed significant transcriptome variation among the major genetic subgroups of Arab populations characterized recently (Razali et al. Nature Comms. 2021). Most of this variation was mapped to immune related pathways implying ancestral differences in disease risk and susceptibility to infection between the genetic subgroups. We further investigated the effects of SNPs, Indels, and Structure Variants (SVs) on gene expression, splicing (isoform-level and intron-level), and allelic expression (variant-level and gene-level) in cis, identifying genetic regulatory associations for 13,195 genes at FDR 0.05. These included novel hits for 3.8m gene expression quantitative trait loci (eQTLs), 3m alternate isoform expression quantitative trait loci (isoQTLs), 1.3m splicing quantitative trait loci (sQTLs), and (1.1m variant-level and 3.5m gene-level) allele-specific expression QTL (aseQTLs). We co-localized the QTL signals with a published GWAS on 6,218 subjects from the same QGP cohort, featuring 45 clinically relevant traits. With this analysis, we identified 845 SNPs in 832 genes to be causal for both genetic expression and disease pathology in a set of target genes previously associated with a range of Anthropometric, Electrolytes, Enzymes, Coagulation, Blood Cells, Lipids, other biochemical-related trait loci. The generated data constitutes the largest whole blood multi-QTL resource available to date based on non-imputed data and samples, uniformly-generated and processed in a single center. By revealing novel loci unique to Arab and Middle Eastern populations, this contributes to generating a valuable reference dataset to further characterize under-studied populations with important relevance worldwide. Our QTL and GWAS colocalization analysis revealed underlying molecular mechanisms to allelic heterogeneity and pleiotropic effect on complex traits.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2493. A gain-of-function recurrent missense variant leads to a GABAergic/glutamatergic imbalance in a forebrain organoid model of PACS1 syndrome

Authors:

L. Rylaarsdam, A. Guemez-Gamboa; Northwestern Univ., Chicago, IL

Abstract Body:

PACS1 syndrome is a neurodevelopmental disorder characterized by intellectual disability and craniofacial abnormalities resulting from a *de novo* p.R203W variant in phosphofurin acidic cluster sorting protein 1 (PACS1). PACS1 is known to play roles in both the endosomal pathway and nucleus, but little is known about how this variant affects the developing nervous system and patients have few therapeutic options. Here, we used stem cell-derived forebrain organoids to show that PACS1^(+/R203W) cells share convergent molecular features of autism spectrum disorder (ASD) through a gain-of-function mechanism. These features include an aberrant propensity towards GABAergic differentiation, increased inhibitory synaptic density, and impaired expression of an ASD-specific gene network enriched at synapses. This work is the first to investigate the impact of the p.R203W variant on the developing brain and suggests a therapy that either clears p.R203W PACS1 or targets convergent pathological mechanisms of ASD could be beneficial for patients.

Molecular Effects of Genetic Variation Posters - Thursday

PB2494. A genome-first approach characterizing the frequency and phenotype of potentially pathogenic germline *DGCR8* variants in two large, unselected cohorts

Authors:

J. Kim¹, J. Haley², D. Smelser², DiscovEHR Collaboration, B. Rivera³, W. Foulkes⁴, D. Carey², D. Stewart⁵; ¹Natl. Cancer Inst., Rockville, MD, ²Geisinger Hlth.System, Danville, PA, ³Bellvitge BioMed. Res. Inst., Barcelona, Spain, ⁴McGill Univ. Hlth.Ctr., Montreal, QC, Canada, ⁵Natl. Cancer Inst., ROCKVILLE, MD

Abstract Body:

DGCR8 forms a microprocessor complex with *DROSHA*, where it is required for processing pri-microRNAs. Recent work identified two unrelated families with a germline *DGCR8* missense variant, p.Glu518Lys, segregating with schwannoma/thyroid neoplasia or hyperplasia. We used a genome-first approach to characterize prevalence and phenotype in individuals with germline *DGCR8* variants. We used the Geisinger DiscovEHR study (175K exomes) and UK Biobank (200K exomes). We analyzed predicted loss-of-function (pLOF) variants; missense and splice site variants were classified as “damaging” if 3 of 4 missense predictors (metaSVM, CADD, REVEL, bayesdel) or 2 of 3 splice-site predictors (dbscSNV, spliceAI, spidex) exceeded set thresholds. We queried neoplasms, thyroid phenotypes and procedure codes using a self-reported questionnaire, EHR, and/or cancer and death registries. Controls were selected from the full cohort who do not carry a *DGCR8* pLOF or “damaging” missense variant. We identified 29 individuals (1:6,050) and 16 individuals (1:12,540) with a pLOF variant and 16 individuals (1:12,540) and 17 individuals (1:10,321) with missense variants in the DiscovEHR and UKBiobank, respectively. In both cohorts, we did not see any enrichment of neoplasms with pLOF variants. However, in DiscovEHR, we identified one woman with a papillary thyroid cancer at 27 and an ovarian cystic teratoma at 17 (a tumor reported in the original *DGCR8* pedigree). In the UK Biobank there was a male never-smoker with tonsil cancer at 34. Next, we evaluated thyroid phenotypes in pLOF carriers. We did not find anyone with thyroid phenotypes in UK Biobank; in the DiscovEHR cohort, only hyperthyroidism was found to be significant (p=0.04) vs controls. In both cohorts there was no difference in age at diagnosis for malignancy, but age was significantly younger for thyroid phenotypes in DiscovEHR (40.3 in pLOF vs. 52.4 in non-carriers, p=0.01). Individuals with a missense variant did not show any enrichment or earlier age at diagnosis for malignancy or thyroid phenotypes. However, one female in the UK Biobank with a “damaging” missense variant (p.Pro346Ala) had a cranial nerve neurilemmoma (schwannoma) diagnosed at 40; she had no germline variant in *NF2*. *DGCR8* Glu518Lys was not observed in either cohort. In DiscovEHR, we observed an earlier age of diagnosis for some thyroid phenotypes; lack of thyroid diagnoses in the UK Biobank may reflect differences in medical culture between the US and UK. Our study suggests that germline *DGCR8* variants are not associated with increased risk of cancer, however, the detection of a schwannoma, teratoma and papillary thyroid cancer diagnoses merit follow-up.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2495*. A Homozygous IER3IP1 Mutation Causes Secretory Protein Trafficking Defects in Neural Progenitor Cells

Authors:

L. Ahn¹, K. Chase², A. Peden³, K. Zhang², A. Schaffer¹; ¹Case Western Reserve Univ., Cleveland, OH, ²Mayo Clinic, Dept. of NeuroSci., Jacksonville, FL, ³The Univ. of Sheffield, Sheffield, United Kingdom

Abstract Body:

Microcephaly with simplified gyration, generalized epilepsy, and permanent neonatal diabetes syndrome (MEDS) is a severe autosomal recessive disorder characterized by the aforementioned clinical features. It is caused by deleterious bi-allelic variants in the *immediate early response-3 interacting protein-1* (*IER3IP1*) gene. The role of IER3IP1 in the pathogenesis of MEDS remains elusive. Yos1p, a yeast homolog, has been reported to be involved in the anterograde transport of cargos from the endoplasmic reticulum (ER) to the Golgi via COPII vesicles. Also, recent proteomic data collected from HeLa cells, as well as a genetic knockout screen performed in cerebral organoids, suggest that IER3IP1 may perform a conserved function in protein trafficking. Based on prior research and the clinical presentations of MEDS patients, we hypothesize that human IER3IP1 regulates secretory protein trafficking during neurogenesis. To test this hypothesis, we generated isogenic neural progenitor cells (NPCs) and induced pluripotent stem cells (iPSCs) lines reprogrammed from MEDS patient fibroblasts for phenotypic analysis. To investigate whether IER3IP1^{L78P/L78P} causes structural changes in organelles critical in protein trafficking, we performed immunocytochemistry for Calnexin, a marker for ER structures, as well as electron microscopy imaging of organelles. We saw IER3IP1^{L78P/L78P} NPCs had reduced expression of Calnexin, swollen ER morphology with abundant multilamellar bodies, and an increased number of vacuolized mitochondria upon electron microscopy imaging, compared to IER3IP1 wildtype NPCs. Structural abnormalities in organelles in IER3IP1^{L78P/L78P} NPCs suggested functional defects in protein secretion. Thus, we sought to quantify and compare protein secretion capacity directly in IER3IP1^{L78P/L78P} and control NPCs. We generated a GFP trafficking reporter that utilizes an inducible FKBP aggregation domain. The reporter enables us to measure the rate of GFP-fused protein trafficking in real-time and small molecule-dependent manner. Using this reporter, we found the patient variant caused a defect in protein trafficking and secretion. These phenotypes were ameliorated after correcting the IER3IP1 variant through CRISPR/Cas9 gene-editing of the patient-derived cells, indicating these cellular phenotypes were caused by the IER3IP1^{L78P/L78P} variant. In summary, we demonstrate that IER3IP1^{L78P/L78P} is deleterious and likely leads to MEDS by disrupting ER structure and secretory protein trafficking.

Molecular Effects of Genetic Variation Posters - Thursday

PB2496. A humanized yeast model to identify disease-associated, dominant-negative effects on protein function

Authors:

A. Antonellis¹, R. Meyer-Schuman², S. Marte¹, T. Smith¹, S. Feely³, M. Kennerson⁴, G. Nicholson⁵, M. Shy⁶, K. Koutmou¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Baylor Coll. of Med., Houston, TX, ³Univ Iowa, Iowa City, IA, ⁴ANZAC Res. Inst., Concord, Australia, ⁵Concord Hosp, Concord, Australia, ⁶Univ. of Iowa, Iowa City, IA

Abstract Body:

Aminoacyl-tRNA synthetases (ARSs) are ubiquitously expressed, essential enzymes that translate the genetic code by charging tRNA molecules with cognate amino acids. Heterozygosity for missense variants or small in-frame deletions in five ARS genes causes dominant axonal peripheral neuropathy, a disorder characterized by impaired motor and sensory function in the distal extremities. The majority of neuropathy-associated ARS variants reduce enzyme activity, but do not significantly decrease protein levels. Interestingly, all five implicated ARS genes encode enzymes that function as homodimers. These observations raise the possibility of a dominant-negative effect, where non-functional mutant ARS subunits dimerize with wild-type ARS subunits and reduce overall ARS activity below 50%, which may breach a threshold required for peripheral nerve axons and activate the integrated stress response, as recently shown *in vivo* for neuropathy-associated glycyl-tRNA synthetase (*GARS1*) variants. To test for dominant-negative properties of neuropathy-associated ARS alleles, we developed a humanized yeast assay to co-express pathogenic human alanyl-tRNA synthetase (*AARS1*) mutations with wild-type human *AARS1*. Here, we will present our unpublished data on a series of pathogenic, neuropathy-associated *AARS1* missense variants and show that: (1) pathogenic variants cause a loss-of-function effect; (2) pathogenic variants do not ablate protein-protein interactions with wild-type *AARS1*; (3) pathogenic variants directly impair the function of wild-type *AARS1*; and (4) reducing dimerization between mutant and wild-type *AARS1* rescues the impaired function of wild-type *AARS1*. In sum, these data demonstrate that neuropathy-associated *AARS1* variants exert a dominant-negative effect, which supports a common, loss-of-function mechanism for ARS-mediated dominant peripheral neuropathy.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2497. A lupus-associated variant in IRF7 amplifies IFN- α production.

Authors:

S. Virolainen^{1,2}, K. Dunn¹, C. Forney¹, O. Donmez¹, S. Parameswaran¹, M. Weirauch¹, L. Kottyan¹; ¹Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, ²Univ. of Cincinnati Coll. of Med., Cincinnati, OH

Abstract Body:

Systemic Lupus Erythematosus (SLE) is an autoimmune disease characterized by widespread inflammation and organ destruction that disproportionately affects individuals with two X chromosomes and those of African American ancestry. SLE has high heritability, with genetics contributing to nearly half of disease risk. More than 90 genetic risk loci are implicated in the etiology of SLE, with additive effects of these loci likely contributing to SLE risk through dysregulation of immune cells. Genetic variants within loci encoding multiple Interferon regulatory factors (IRFs) are associated with SLE. IRF proteins are key regulators of the expression of and response to type I interferons (IFNs). While IFNs are critical for antiviral immune responses, dysregulation of IFNs is a major contributor to autoimmunity. A majority of SLE patients exhibit elevated levels of circulating IFN, a feature that correlates with disease activity. The genetic mechanisms contributing to chronically elevated IFN in SLE patients are poorly understood, but have been generally correlated with the SLE risk genotype at the *IRF7*, *IRF5*, and *STAT1* SLE risk loci. In this study, we examine the downstream consequences of an SLE-associated missense variant in IRF7 on IFN secretion.

We focus on rs1131665 (p -value $< 5 \times 10^{-8}$, OR =1.24 in Europeans and OR =1.17 in African Americans). rs1131665 results in an arginine (R, non-risk) to glutamine (Q, risk) amino acid change in the autoinhibitory domain of IRF7. We generated human cell lines expressing both the risk and non-risk IRF7 proteins and assessed differential IFN production at both the mRNA and protein levels before and in response to toll-like receptor-7 (TLR-7) stimulation. Strikingly, genotype-dependence was evident prior to TLR-7 induction, with cells expressing the risk allele showing higher baseline IRF7 activity at the mRNA and protein levels as assessed by RNA-seq and ELISA, respectively. Following TLR-7 stimulation, bulk RNA-seq analysis revealed ~20 differentially expressed genes between cells expressing the risk and non-risk constructs. A majority of these genes are involved in the IFN pathway and were upregulated in cells expressing the risk construct. IFN- α ELISA on cell supernatant showed similar heightened (~2 fold) IFN- α secretion at the protein level in cells expressing the IRF7 risk variant compared to those expressing the non-risk variant.

We conclude that the SLE-associated variant within IRF7 results in heightened IRF7 activity as measured by IFN output at the mRNA and protein levels. This genotype dependent activity can be independent of TLR-7 stimulation but is also observed with TLR-7 stimulation.

Molecular Effects of Genetic Variation Posters - Thursday

PB2498. A Novel De novo Likely Pathogenic Variant of *WFS1* gene in a Pakistani Child

Authors:

M. Hanif¹, S. A. Ahmed², M. N. Ibrahim¹, S. J. Raza¹; ¹NICH, KARACHI, Pakistan, ²Kaiser Permanente, Oakland, CA

Abstract Body:

Abstract Background: Wolfram Syndrome is a rare autosomal recessive genetic disorder caused by mutations in *WFS1* gene, located on chromosome 4p16.1. This gene encodes for wolframin membrane-embedded protein of endoplasmic reticulum. Pathogenic genetic alterations in *WFS1* gene usually result in endoplasmic reticulum (ER)-stress leading to neurodegeneration pancreatic β -cell dysfunction and apoptosis. Wolfram syndrome is associated with juvenile onset diabetes mellitus, optic atrophy, sensorineural deafness, and diabetes insipidus. This syndrome is also known as DIDMOAD (Diabetes Insipidus, Diabetes Mellitus, Optic Atrophy, and Deafness). Though Wolfram syndrome is generally considered an autosomal recessive disorder, heterozygous pathogenic variants have been noted to cause Wolfram like syndrome (congenital progressive hearing impairment, diabetes mellitus, and optic atrophy). **Objective:** The aim of this study is to highlight a novel de-novo likely pathogenic variant of *WFS1* in a Pakistani child liable to cause Wolfram like syndrome. **Case Discussion:** We report a novel de novo likely pathogenic variant in a Pakistani child in *WFS1* gene who presented with clinical features consistent with wolfram like syndrome (MedGen UID: 481988). This seven-year-old boy initially presented to our endocrine clinic for management of uncontrolled type 1 diabetes mellitus despite of being on glargine and aspart insulin his HbA1c was high (14%) though other investigations were normal. A physical exam noted a short stature (-4.55 SDS for age 7 years), otherwise, he appeared normal. Past history was significant for sensorineural deafness in his first year of life. He also had bilateral cataract surgery at the age of five years. Family history is only significant for type 2 diabetes on his paternal side. Based on these clinical findings diabetes with extra-pancreatic features was suspected and genetic testing was ordered for a multi-gene panel which included the *WFS1* gene. The original report noted c.2586G>T (p.Lys862Asn) as a variant of uncertain significance but subsequent parental testing led to the reclassification of the variant as a likely pathogenic variant. **Conclusion:** Based on this new diagnosis an eye exam was requested that showed bilateral moderate optic atrophy which further supported this diagnosis our report highlights the rare possibility of a heterozygous variant causing Wolfram like syndrome and identifies a novel likely pathogenic variant in this gene which is supported by clinical features consistent with Wolfram like syndrome.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2499. A rare genetic variant in the cleavage site of *prepro-orexin* is associated with idiopathic hypersomnia

Authors:

T. Miyagawa¹, S. Tanaka², M. Shimada³, N. Sakai⁴, S. Nishino⁴, K. Mishima⁵, K. Tokunaga³, M. Honda¹; ¹Tokyo Metropolitan Inst. of Med. Sci., Tokyo, Japan, ²Kansai Med. Univ., Osaka, Japan, ³Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, ⁴Stanford Univ., Stanford, CA, ⁵Akita Univ. Graduate Sch. of Med., Akita, Japan

Abstract Body:

Idiopathic hypersomnia (IH) is a rare, heterogeneous sleep disorder characterized by excessive daytime sleepiness. The exact prevalence of IH remains unknown due to the absence of epidemiological studies. A high proportion of IH patients (26.9-39.1%) reported a family member with excessive daytime sleepiness. In contrast to narcolepsy type 1, which is a well-defined type of central disorders of hypersomnolence, the etiology of IH is poorly understood. No susceptibility loci associated with IH have been clearly identified, despite the tendency for familial aggregation of IH. We performed a variation screening of *prepro-orexin/hypocretin* and *orexin receptors* genes, and an association study for IH in a Japanese population, with replication (598 IH patients and 9826 controls). We identified a rare missense variant (g.42184347T>C; p.Lys68Arg; rs537376938) in the cleavage site of *prepro-orexin* that was associated with IH (minor allele frequency of 1.67% in cases versus 0.32% in controls, $P = 2.7 \times 10^{-8}$, odds ratio = 5.36). All patients with this mutation were heterozygous carriers. Subjective sleepiness as evaluated with the Epworth Sleepiness Scale scores in unmedicated conditions and arousal index (polysomnography) were nominally higher in the orexin mutation-positive patients compared with the orexin mutation-negative patients ($P = 0.049$ and $P = 0.046$, respectively), suggesting that mutation-positive IH patients suffered more sleepiness and sleep instability. No significant association between narcolepsy (type 1 and type 2) and rs537376938 was observed. Two forms of orexin (orexin-A and -B) are generated by cleavage of one precursor peptide, prepro-orexin. A cleavage assay was performed using proprotein convertase subtilisin/kexin (PCSK) type 1 and PCSK2 to examine the differences in cleavage efficiency between wild-type (Gly-Lys-Arg; GKR) and mutant (Gly-Arg-Arg; GRR) peptides at the cleavage site of prepro-orexin. In both PCSK1 and PCSK2 assays, the mutant peptide was less processed than the wild-type peptide. We also confirmed that the prepro-orexin peptides themselves transmitted less signaling through orexin receptors than mature orexin-A and orexin-B peptides. These results indicate that a subgroup of IH is associated with decreased orexin signaling, which is believed to be a hallmark of narcolepsy type 1.

Molecular Effects of Genetic Variation Posters - Thursday

PB2500. A risk variant for Barrett's esophagus and esophageal adenocarcinoma at chr8p23.1 affects enhancer activity and implicates multiple gene targets

Authors:

M. Buas¹, M. Ali², J. Chen¹, X. Wang³, J. Dai³, T. Vaughan³, G. Casey²; ¹Mem. Sloan Kettering Cancer Ctr., New York, NY, ²Univ. of Virginia, Charlottesville, VA, ³Roswell Park Comprehensive Cancer Ctr., Buffalo, NY, ⁴Fred Hutchinson Cancer Res. Ctr., Seattle, WA

Abstract Body:

Nineteen genetic susceptibility loci for esophageal adenocarcinoma (EAC) and its precursor Barrett's esophagus (BE) have been identified through genome-wide association studies (GWAS). Clinical translation of such discoveries, however, has been hindered by the slow pace of discovery of functional/causal variants and gene targets at these loci. We previously developed a systematic informatics pipeline to prioritize candidate functional variants using functional potential scores, applied the pipeline to select high-scoring BE/EAC risk loci, and validated a functional variant at chr19p13.11 (rs10423674). Here, we selected two additional prioritized loci for experimental interrogation: chr3p13/rs1522552 and chr8p23.1/rs55896564. Candidate enhancer regions encompassing these variants were evaluated using luciferase reporter assays in two EAC cell lines. One of the two regions tested exhibited allele-specific enhancer activity - 8p23.1/rs55896564. CRISPR-mediated deletion of the putative enhancer in EAC cell lines correlated with reduced expression of three candidate gene targets: B lymphocyte kinase (*BLK*), nei like DNA glycosylase 2 (*NEIL2*), and cathepsin B (*CTSB*). Expression quantitative trait locus (eQTL) mapping in normal esophagus and stomach revealed strong associations between the BE/EAC risk allele at rs55896564 (G) and lower expression of *CTSB*, a protease gene implicated in epithelial wound repair. These results further support the utility of functional potential scores for GWAS variant prioritization, and provide the first experimental evidence of a functional variant and risk enhancer at the 8p23.1 GWAS locus. Identification of *CTSB*, *BLK*, and *NEIL2* as candidate gene targets suggests that altered expression of these genes may underlie the genetic risk association at 8p23.1 with BE/EAC.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2501. A single-nucleus transcriptome-wide association study implicates novel genes in depression pathogenesis

Authors:

L. Zeng; Columbia Univ. Irving Med. Ctr., New York, NY

Abstract Body:

Depression is a common psychiatric illness and global public health problem. However, our limited understanding of the biological basis of depression has hindered the development of novel treatments and interventions. To identify new candidate genes for therapeutic development, we integrated human single-nucleus RNA sequencing (snucRNAseq) data from the dorsolateral prefrontal cortex (N=424) in 7 cell types and 81 cell subtypes with depression genome-wide association study (GWAS) results (N=500,199). Namely, we performed a transcriptome-wide association study of depression followed by Mendelian randomization. We identified 67 causal genes in depression that have a role in specific neocortical cell subtypes snucRNAseq; 51/67 of them were novel compared to previous studies. Importantly, we demonstrate that, depression TWAS genes showed a cell type specific pattern, with the greatest enrichment being in neurons and astrocyte. Compared to lower genetic correlated traits (e.g., BMI) with depression, higher correlated traits (e.g., schizophrenia) have more common TWAS genes with depression. In parallel, we performed differential gene expression analysis in relation to depression in 92 cortical cell types, and we found that genes such as *CCDC6*, *MADD*, *TAOK3* and *MEF2A* are associated with depression in specific cell types. These two analyses illustrate the utility of large snucRNAseq data to uncover both genes whose expression is altered in specific cell subtypes in the context of depression and to enhance the interpretation of well-powered GWAS so that we can prioritize specific susceptibility genes for further analysis.

Molecular Effects of Genetic Variation Posters - Thursday

PB2502. A splicing variant found in the human myostatin gene encodes an isoform that inhibits myostatin

Authors:

K. Maeta, M. Farea, H. Nishio, M. Matsuo; Faculty of Rehabilitation, Kobe Gakuin Univ., Kobe, Japan

Abstract Body:

[Background] MSTN gene encodes myostatin that inhibits myocyte proliferation and is conserved across species. Although this gene has only three exons, splicing variants (SVs) have been identified in sheep and birds. Surprisingly, the isoform encoded by MSTN SV inhibited myostatin and suppressed muscle atrophy. Therefore, MSTN SV has attracted much attention as a myostatin inhibitory molecule. However, there is no report of MSTN SV in humans. In this study, by expanding the analysis of MSTN mRNA from the coding region to the 3' UTR, we succeeded in cloning MSTN SV in humans for the first time in the world. Furthermore, this MSTN SV encoded an isoform with myostatin inhibitory activity. [Methods] MSTN SV was identified by RT-PCR of MSTN mRNA from rhabdomyosarcoma (RMS) cells. Expression of MSTN SV was analyzed by RT-PCR amplification of myoblast- and skeletal muscle-derived RNA. Expressed MSTN SV and their myostatin inhibitory activity were evaluated by western blotting and SMAD-dependent luciferase reporter, respectively. [Results] PCR amplification of the coding region of MSTN mRNA amplified a single band of MSTN mRNA. However, when the PCR amplified region was expanded deep into the 3' UTR, a small amplification product was obtained in addition to the normal product. This product was a deletion of approximately 1,000 bases in the 5' end of exon 3. This SV was activated atypical splice acceptor site consisting of two bases of TG in the 3' UTR (MSTN-b). MSTN-b was amplified in myoblasts but not in skeletal muscle. Expression of MSTN-b resulted in the detection of a 40 kDa band that reacted with a myostatin N-terminal recognition antibody, and the same band was also detected in RMS cells. From the above, we concluded that the protein encoded by MSTN-b is a novel myostatin isoform (myostatin-b). Although myostatin-b is 15 amino acids shorter than the prodomain of myostatin, its expression inhibited myostatin signaling in RMS cells. However, it did not inhibit the signal elicited by recombinant GDF11. Thus, we conclude that myostatin-b is a novel myostatin inhibitor specific for myostatin. [Discussion] MSTN-b is thought to be previously undiscovered because it utilizes an atypical splice acceptor located deep in the 3'UTR. Since myostatin-b has myostatin-specific inhibitory activity, it is expected to be clinically applicable as an innovative myostatin inhibitor for the prevention and treatment of muscle atrophy.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2503. A structural variant of the C-terminal prion-like domain of TDP-43 causes vacuolar muscle degeneration.

Authors:

P. Ervilha Pereira¹, N. Schuermans¹, A. Meylemans², P. LeBlanc¹, L. Versluys¹, E. Debackere¹, O. Vanakker², S. Janssens², J. Baets³, K. Verhoeven⁴, M. Lammens⁵, S. Symoens², B. De Paepe², J. De Bleecker², E. Bogaert¹, B. Dermaut¹; ¹Ghent Univ., Gent, Belgium, ²Ghent Univ. Hosp., Gent, Belgium, ³Univ. of Antwerp, Wilrijk, Antwerp, Belgium, ⁴Sint-Jan Hosp. Bruges, Bruges, Belgium, ⁵Antwerp Univ. Hosp., Antwerp, Belgium

Abstract Body:

Neuronal TDP-43-positive inclusions are a hallmark lesion found in frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS), present in 45% and 97% of cases, respectively. Missense mutations in the C-terminal prion-like domain (PrLD) of TDP-43 are the main type of disease-causing mutations reported for ALS. This domain mediates self-interaction, and mutations of this region are thought to lead to aggregate formation. However, the scope of TDP-43 proteinopathies goes beyond neural tissues, with reports showing TDP-43 pathology in vacuolar myopathies. Nevertheless, genetic evidence for a primary role for TDP-43 in myopathies is still amiss. Here we identified a multigenerational family with an autosomal dominant rimmed vacuole myopathy. Whole exome sequencing and genome-wide linkage analysis mapped the disease to an 11bp deletion in TARDBP (LOD-score of 3.6), causing a frameshift mutation in the C-terminal domain (CTD) of TDP-43 (TDP-43^{p.Trp385IlefsTer10}). This constitutes a novel type among described TDP-43 mutations, which are predominantly missense mutations. Patient-derived muscle biopsies revealed the presence of p62/TDP-43-positive sarcoplasmic inclusions and nuclear depletion of TDP-43. Additionally, we verified higher numbers of autophagosomes and a transcriptomic signature indicative of reduced mitochondrial and lipid metabolism, alongside a switch in sarcomeric protein isoforms suggesting increased muscle regeneration. Together with these observations, functional assays in *D. melanogaster* showed that TDP-43^{p.Trp385IlefsTer10} retains normal function but has reduced toxic gain-of-function properties. By studying this unique variant of TDP-43 it is our goal to clarify the importance of the CTD of TDP-43 and how its remodeling can affect the formation of aggregates. Furthermore, these results genetically link TDP-43 to vacuolar muscle degeneration for the first time. This not only highlights the importance of the PrLD in pathological conditions in a tissue-specific manner, but it also expands the implications of TDP-43 proteinopathies, from a nearly neuronal-exclusive context into a broader spectrum encompassing myopathies as well.

Molecular Effects of Genetic Variation Posters - Thursday

PB2504*. A uniquely prevalent deleterious variant within *POMC* gene identified in Estonian Biobank cohort

Authors:

E. Abner, N. Taba, T. Nikopensius, T. Esko; Univ. of Tartu, Tartu, Estonia

Abstract Body:

Genome wide association studies based within specific population groups provide an excellent opportunity for pinpointing rare and high impact genomic variants in already well studied traits, such as body-mass index (BMI).

We performed an exploratory GWAS in Estonian biobank (EstBB) to describe BMI-affecting variants (185,429 participants, BMI as continuous trait). Our analysis identified 178 significant loci ($p < 5 \times 10^{-8}$; at least ± 1 Mb apart), of which a majority have previously been reported to associate with obesity-related traits. To pinpoint functionally relevant SNVs, we characterized individual variants by their consequence on protein-coding genes, in correlation to their CADD score and beta coefficient.

Among the most significant functional hits, a stop-gain variant rs202127120 (*POMC*:Glu206X) within the pro-opiomelanocortin (*POMC*) gene, was identified. This variant is present in 0.87% of EstBB participants, displaying to our knowledge a dramatically higher allele frequency in Estonia than in other global populations. The *POMC*:Glu206X variant results in an early truncation of the *POMC* prohormone and results in lower β -MSH and β -endorphin levels in the hypothalamic-pituitary-adrenal axis, thus resulting in an inadequate leptin-melanocortin pathway triggering.

POMC:Glu206X variant carriers display significantly higher BMI among all biobank participants ($+0.85 \text{ kg/m}^2$; $p = 9.2 \times 10^{-13}$), higher overall weight ($+2.70 \text{ kg}$; $p = 3.8 \times 10^{-13}$), waist circumference ($+1.69 \text{ cm}$; $p = 1.9 \times 10^{-5}$) and higher BMI among 18-39 year old participants ($+1.01 \text{ kg/m}^2$; $p = 1.5 \times 10^{-8}$). A PheWAS utilizing data from medical electronic health records confirmed the enrichment of obesity related diagnoses among the *POMC*:Glu206X variant carriers (Top hit: ICD-10 category E66 "Overweight and obesity", OR 1.42, $p = 8.3 \times 10^{-5}$). Although the variant was originally imputed based on Estonian specific genome reference panels, we reconfirmed the presence of *POMC*:Glu206X via Sanger sequencing and confirmed 183 heterozygous and 13 homozygous carriers (from total 198 EstBB participants). We currently are in process of verifying the carrier status of additional predicted $n=1,579$ biobank participants and will be measuring quantitatively the effect *POMC*:Glu206X has on *POMC*-derived peptides in plasma.

Here, we identified a remarkably prevalent stop-gain variant in *POMC* gene, which previously has been considered highly deleterious. We demonstrate that *POMC*:Glu206X significantly associates with risk of obesity and we will next be proposing clinical intervention trials to ameliorate the effect *POMC*:Glu206X has on the local population.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2505. *A2m1l*-Knockout Mouse as a Potential Model of Chronic Otitis Media

Authors:

C. Elling¹, S. hirsch¹, R. Santos-Cortez²; ¹Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ²Univ. of Colorado, Dept of Otolaryngology, Aurora, CO

Abstract Body:

Inflammation and infection of the middle ear, known as otitis media (OM), is a leading cause of hearing loss and the most frequently diagnosed disease in children worldwide. Traditionally, mouse models for OM rely on inducing acute infection through inoculation of the middle ear, e.g. with the human otopathogen non-typeable *Haemophilus influenzae* (NTHi), with very few genetic models with spontaneous or chronic OM. *A2ML1* variants, including loss-of-function variants, are associated with susceptibility to OM in humans, but no animal model has been reported for OM. Here, we report our middle ear findings in a mouse line with a CRISPR-induced knockout (KO) of *A2m1l*. Mice were X-rayed prior to harvest to determine if there are craniofacial or skeletal abnormalities. Tissue from mouse middle ears, as well as other upper respiratory mucosal tissues, were harvested. The harvested middle ear bullae were examined under microscope for phenotypic indications of OM. RNA samples isolated from middle ear tissue were assayed for expression of genes correlated with *A2ML1* expression in humans. There were no significant craniofacial differences by genotype, however skeletal abnormalities that are concordant with phenotypes observed in *A2ML1*-related Noonan-like syndrome in humans were identified. Otomicroscopic findings in mice heterozygous for the *A2m1l*-KO are frequently indicative of otitis media, with tympanic membrane perforations or thickening, as well as cases of middle ear effusion or fluid. Gene expression studies are in progress and will be reported. Thus far, our preliminary results in this *A2m1l*-KO mouse line indicate spontaneous occurrence of OM in these mice without the need for NTHi inoculation.

Molecular Effects of Genetic Variation Posters - Thursday

PB2506*. Aberrant splicing prediction across human tissues.

Authors:

N. Wagner¹, M. Çelik², F. Hölzlwimmer¹, V. Yezpe¹, C. Mertes³, H. Prokisch⁴, J. Gagneur¹; ¹Technical Univ. of Munich, Garching bei München, Germany, ²Univ. of California Irvine, Irvine, CA, ³Technical Univ. of Munich, Garching, Germany, ⁴Helmholtz Zentrum Muenchen, Neuherberg, Germany

Abstract Body:

Aberrant splicing is a major cause of genetic disorders but its direct detection in transcriptomes is limited to clinically accessible tissues such as skin or body fluids. While DNA-based machine learning models allow prioritizing rare variants for affecting splicing, their performance on predicting tissue-specific aberrant splicing remains unassessed. Here, we generated the first aberrant splicing benchmark dataset, by calling aberrant splicing events on 16,213 RNA-seq samples from 49 human tissues from the GTEx dataset, comprising 8.8 million rare variants in paired genotype data from 946 individuals. At 20% recall, state-of-the-art DNA-based models SpliceAI and MMSplice cap at 10% precision. We observed that many false predictions originated from inaccurate genome annotations. We constructed a tissue-specific splicing map (SpliceMap) by mapping and quantifying tissue-specific splice site usage transcriptome-wide and modeling isoform competition. Using SpliceMap together with SpliceAI and MMSplice in a combined model that we call AbSplice increased precision by three-fold at the same recall. These results replicated in two independent rare disease cohorts. Moreover, the GTEx dataset consists of post-mortem collected RNA-seq samples across a vast variety of tissues and thereby offers the unique opportunity to evaluate to what extent aberrant splicing in an accessible tissue reflects aberrant splicing of another tissue of interest. Integrating RNA-sequencing data of clinically accessible tissues brought precision to 60%. We provide precomputed AbSplice scores [1] for all possible single-nucleotide variants genome-wide and publicly available software [2] to score variants including indels directly from VCF files. Altogether, our results substantially contribute to non-coding loss-of-function variant identification and to genetic diagnostics design and analytics.

[1] Zenodo: [10.5281/zenodo.6408331](https://zenodo.org/record/6408331)

[2] <https://github.com/gagneurlab/absplICE>

Molecular Effects of Genetic Variation Posters - Wednesday

PB2507. Abnormal cell fate allocation in embryogenesis caused by *Zic3* loss-of-function

Authors:

A. Haaning, A. Phatak, H. M. Bellchambers, M. B. Padua, S. M. Ware; Herman B Wells Ctr. for Pediatric Res., Indiana Univ. Sch. of Med., Indianapolis, IN

Abstract Body:

Pathogenic variants in *ZIC3* cause X-linked heterotaxy, a laterality disorder caused by abnormal left-right (L-R) patterning. Laterality defects include congenital heart defects (CHD), but some CHD associated with *ZIC3* variants are isolated and/or not specifically related to laterality defects, such as ventricular non-compaction. Furthermore, the spectrum of CHD in *Zic3* null mice, which recapitulates laterality defects and CHD observed in humans with *ZIC3* variants, is broader and more severe than in another heterotaxy mouse model resulting from *Foxj1* loss-of-function. Whereas *Foxj1* is specifically expressed in and affects functionality of the L-R organizer, *Zic3* is broadly expressed during gastrulation and influences planar cell polarity (PCP), as well as cell fate transitions and pluripotency. We hypothesize that some CHD resulting from *Zic3* loss-of-function stem from abnormalities in cell fate allocation. To test this hypothesis, we used scRNA-seq analysis to investigate whether *Zic3* loss-of-function leads to abnormal cell lineage allocation in E8 embryos. We observed that proportions of several cell types in *Zic3* null embryos differed significantly from wildtype, and, notably, cardiac progenitor cell numbers were increased in *Zic3* nulls. Pseudo-bulk analysis identified genes that were differentially expressed in *Zic3* nulls. *Hmgal1* was downregulated in *Zic3* nulls in 11 of 15 cell type clusters and whole embryo pseudo-bulk datasets. HMGA proteins function as epigenetic regulators of transcription by binding chromatin at AT-rich sites, causing changes in chromatin conformation that impact transcription factor binding. *Hmgal1* is highly expressed in pluripotent cells and its silencing results in downregulation of pluripotency genes. Broad and consistent downregulation of *Hmgal1* in *Zic3* nulls suggests that *ZIC3* may have a role in epigenetic regulation of cell differentiation. Of the genes upregulated in *Zic3* nulls, three encode transcription factors in the Early Growth Response (EGR) family: EGR2, EGR3, and EGR4. EGR proteins respond to mitogenic signals in a variety of cells and impact cell growth and proliferation. Results of this study suggest that *ZIC3* is a regulator of cell differentiation, affecting cell lineage allocation, signifying a potential novel cell-fate mechanism underlying CHD.

Molecular Effects of Genetic Variation Posters - Thursday

PB2508. Abundance of immunoprotein CD99 affected by loss of chromosome Y

Authors:

J. Mattisson¹, M. Danielsson¹, M. Hammond¹, H. Davies¹, C. Gallant¹, J. Nordlund², A. Raine², M. Edén³, M. Ingelsson³, J. Dumanski^{1,4}, J. Halvardson¹, L. Forsberg^{1,5}; ¹Dept. of Immunology, Genetics and Pathology and Sci. for Life Lab., Uppsala Univ., Uppsala, Sweden, ²Dept. of Med. Sci. and Sci. for Life Lab., Uppsala Univ., Uppsala, Sweden, ³Dept. of Publ. Hlth. and Caring Sci. / Geriatrics, Uppsala Univ., Uppsala, Sweden, ⁴Faculty of Pharmacy and 3P Med. Lab., Intl. Res. Agendas Programme, Med. Univ. of Gdańsk, Gdańsk, Poland, ⁵The Beijer Lab., Uppsala Univ., Uppsala, Sweden

Abstract Body:

Aging men have an increasing risk of somatically losing the Y chromosome. This aneuploidy, known as mosaic loss of chromosome Y (LOY), is the most common somatic mutation in leukocytes of elderly men. LOY is also associated with increased all-cause mortality and disease risk, including Alzheimer's disease, autoimmune disease, cardiovascular disease and cancer. Most studies have thus far investigated the clinical associations of LOY, with the functional impacts of Y loss on a cellular level remaining largely unexplored. Here, we investigated CD99, a cell surface protein essential for a normal immune functionality of leukocytes. The *CD99* gene is located in the pseudoautosomal region of chromosome X and Y, and has previously been shown downregulated in association with Y loss. However, as transcription does not necessarily translate directly into protein levels, especially for cell surface proteins, we sought to investigate whether LOY also affects the cell surface abundance of CD99. This had to be done while accounting for cell type since many transcriptional effects of LOY have been reported to be cell type specific. We therefore used CITE-seq, a single cell technology that simultaneously quantifies RNA and protein levels, applying monoclonal antibodies to target the cell surface features of interest. Peripheral blood mononuclear cells (PBMCs) sampled from four elderly men were sequenced, targeting CD99 as well as CD4, CD8, CD14, CD16, CD19 and CD56. The six additional proteins are cell type markers commonly used to classify leukocytes found in PBMCs. Here, they allowed us to validate the performance of CITE-seq and also worked as controls that should be unaffected by LOY. We found an overall lower cell surface abundance of CD99 in cells affected by LOY. This CD99 decrease was significant in all of the studied cell types, with the largest effect found in B lymphocytes, corresponding to a loss of about 27% in CD99 cell surface abundance. In contrast, the abundance of the cell type markers, used as controls, were not significantly affected by LOY. Considering the critical role that CD99 plays in facilitating the immune response of leukocytes, its decreased cell surface abundance from LOY could influence disease vulnerability in affected men. This study is therefore the first to suggest a possible mechanism between LOY and disease on a protein level.

Molecular Effects of Genetic Variation Posters - Thursday

PB2510. All-but-one conditional analysis of eQTL isolation in peripheral blood sheds light onto causal variant prioritization.

Authors:

M. Brown¹, E. Greenwood¹, B. Zeng², G. Gibson³; ¹Georgia Inst. of Technology, Atlanta, GA, ²Mount Sinai, New York, NY, ³Georgia Tech, Atlanta, GA

Abstract Body:

Expression quantitative trait locus (eQTL) detection has become increasingly important for understanding how non-coding variants contribute to disease susceptibility and complex traits. The major challenge in eQTL fine-mapping and causal variant discovery relate to the impact of linkage disequilibrium on signals due to one or multiple functional variants that lie within a credible interval. We contrast eQTL fine-mapping using the all-but-one approach, which conditions each signal on all others detected in an interval, with results from forward stepwise conditional analysis as well as Bayesian localization method, all applied to the CAGE cohorts of microarray-based peripheral blood gene expression in 2,138 European-ancestry human adults. All-but-one conditioning significantly modifies effect-size estimates for 51% of 2,351 eQTL peaks, but has only a modest effect on credible interval size and location. On the other hand, both conditioning approaches result in unexpectedly low overlap with Bayesian credible intervals, with just 57% peak concordance and between 50% and 70% SNP sharing, leading us to caution the assumption that any one localization method is superior to another. We also cross referenced our results with ATAC-seq data, cell-type specific eQTL, and ABC-enhancers, leading to proposal of a 5-tier approach to further reduce credible interval sizes and prioritize likely causal variants for all known inflammatory bowel disease risk loci active in immune cells.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2511. Allele-specific expression in blood cells during aging

Authors:

M. Harwood^{1,2}, E. Bader^{1,2}, M. Agbessi¹, V. Bruat¹, M-J. Fave¹, P. Awadalla^{1,2}; ¹Ontario Inst. for Cancer Res., Toronto, ON, Canada, ²Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Gene expression changes during aging are associated with a decline in physical and cognitive abilities. Allele-specific expression (ASE) is the preferential expression of one of two alleles, and has been demonstrated to change based on increases in genetic and environmental variation. ASE can be caused by regulatory variation, post-transcriptional modifications, or environmental exposures, and has the potential to alter phenotype and disease risk. Although ASE is important in genetic regulation, it remains unknown how ASE contributes to variation in aging processes. Here, we use a population cohort, The Canadian Partnership for Tomorrow's Health, to evaluate ASE changes in blood cells during aging. Blood cell composition and functions are impacted by the aging process and some individuals retain healthy blood parameters as they age. Using age and a complete blood count risk score, we utilize whole genome sequencing and single-cell RNA sequencing from blood of 374 individuals who are i) young, healthy (n=90), ii) young, unhealthy (n=88), iii) aged, healthy (n=101), or iv) aged, unhealthy (n=95). We test for ASE at over 6,000 genomic loci in 9 blood cell types. We show an increase in ASE as individuals age, particularly in CD8+ T cells, which are important for killing cancerous or virally infected cells in the adaptive immune system. We identify four genes (RPS17, RPL13, HLA-B, HLA-DRB1) with consistent ASE between young and aged individuals, whereas >50 genes demonstrate differential ASE across age groups. We further classified ASE variants into common ASE (observed in >10 individuals), versus stochastic ASE (only observed in 1-2 individuals). Using gene set enrichment analyses, we identify that genes with common ASE are involved in genetic regulation, such as regulation of the immune system or differentiation, whereas genes with stochastic ASE are involved in cellular response, such as to stress or stimuli. We observe an increase in stochastic ASE in aged and young unhealthy individuals, which was particularly prominent in B cells. B cells are important for antibody production for response to viruses and bacteria, thus more stochastic ASE may be indicative of dysregulation in these older and unhealthy individuals. We also observe more stochastic ASE in CD4+ T cells in aged individuals, further suggesting a dysregulation in immune related cells during aging. Our results suggest that there is a reduction of genetic regulation through ASE during aging that may alter the immune response in unhealthy agers. This research identifies mechanisms for how genetic variation impacts aging processes, which can help identify negative effects of aging.

Molecular Effects of Genetic Variation Posters - Thursday

PB2512. Amplicon sequencing-based noninvasive fetal genotyping for RHD-Positive D antigen-negative alleles

Authors:

A. Hori^{1,2}, A. Sasaki³, K. Takahashi⁴, H. Ogata-Kawata¹, K. Taniguchi¹, O. Migita^{1,5}, A. Kawashima⁶, A. Okamoto⁴, A. Sekizawa⁶, H. Sago³, F. Takada⁷, K. Hata^{1,8}, K. Nakabayashi¹; ¹Natl. Res. Inst. for Child Hlth. and Dev., Tokyo, Japan, ²Nippon Med. Sch. Musashikosugi Hosp., Kanagawa, Japan, ³Natl. Ctr. for Child Hlth. and Dev., Tokyo, Japan, ⁴The Jikei Univ. Sch. of Med., Tokyo, Japan, ⁵Faculty of Med., Univ. of Tsukuba, Ibaraki, Japan, ⁶Showa Univ. Sch. of Med., Tokyo, Japan, ⁷Kitasato Univ. Graduate Sch. of Med. Sci., Kanagawa, Japan, ⁸Gunma Univ. Graduate Sch. of Med., Gunma, Japan

Abstract Body:

Fetal RHD genotyping using maternal cell-free DNA prevents unnecessary anti-D administration when RHD-negative women carry an RhD-negative fetus. In East Asian countries including Japan, fetal RHD genotyping is not performed for the pregnancies of Rh-negative women, and all cases are managed as RhD incompatible pregnancies. In Europeans, the frequency of the RHD gene deletion types among the RhD-negative alleles is more than 99%. However, in East Asians, about 25% of the RhD-negative alleles are non-deletion types. An alternative method to the qualitative PCR-based fetal RHD genotyping, which is widespread in Europe, was needed for East Asians. We have recently developed an amplicon-based noninvasive fetal genotyping method that distinguishes the wild-type RHD allele not only from the RHD-negative D antigen-negative allele (the RHD deletion allele), but also from RHD-positive D antigen-negative alleles. This method requires PCR amplification from four genomic intervals, upstream and downstream Rhesus boxes, RHD exon 9, and RHCE exon 9. Although two regions are co-amplified with one primer pair, each of the sequence reads obtained by the amplicon-sequencing using MiSeq can be accurately mapped to its origin because of one or two base differences between two regions. To distinguish the true absence of a fetal RHD-positive allele in maternal cell-free DNA and the false negative result due to a low fetal fraction (FF), it is important to determine FF for each of maternal cell-free DNA samples. Therefore, we tested whether simultaneous determination of the fetal RHD genotype and FF is possible by multiplex PCR using primers for RHD genotyping and a panel of multiplexed insertion/deletion polymorphisms for 35 loci, followed by amplicon-sequencing. In seven pregnancy cases of RhD-negative women carrying an RhD-positive fetus, we successfully detected a fetal RHD-positive allele (0.8 % ~ 8.6 %) and also determined a FF (3.8 % ~ 16.9 %) in all cases. In two pregnancy cases of RhD-negative women carrying an RhD-negative fetus, we detected the read counts for an RHD-positive allele below the background level determined using control samples and the FFs of 9.9 % and 7.8 %. In the latter cases, the FF data helped us assume the true absence of fetal RHD-positive allele in maternal cell-free DNA. Our fetal RHD genotyping method offers the first opportunity for East Asia countries to introduce a genotyping service for RhD-negative pregnant women and represents a model for other nonwhite countries to establish a genotyping strategy customized to the RHD-positive D antigen-negative alleles prevalent in each country.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2513. An African-specific Alzheimer disease-associated *ABCA7* frameshift deletion results in altered microglia functionality

Authors:

D. Dykxhoorn¹, B. DeRosa², J. Laverde-Paz², L. Adams², T. Starks³, J. Vance¹, M. Cuccaro⁴, J. Haines⁵, G. Byrd⁶, M. Pericak-Vance², H. Cukier²; ¹Univ. of Miami, Miami, FL, ²Univ. of Miami Miller Sch. of Med., Miami, FL, ³Wake Forest Sch. of Med., Winston-Salem, NC, ⁴John P. Hussman Inst. for Human Genomics, Miami, FL, ⁵Case Western Reserve Univ, Cleveland, OH, ⁶Wake Forest Univ., Winston-Salem, NC

Abstract Body:

The *ATP Binding Cassette Subfamily A Member 7 (ABCA7)* gene encodes for an ATP-binding cassette lipid transporter that has been implicated as a risk factor for Alzheimer's disease (AD) across different populations. However, the risk effect of *ABCA7* variants in African Americans (AAs) is stronger than in other populations, such as non-Hispanic whites (NHWs). We previously identified a 44 base pair deletion in *ABCA7* that is significantly associated with AD in AAs (frequency in cases=15.2%, controls=9.74%, $p=1.414 \times 10^{-5}$) and virtually absent in NHWs. The deletion is predicted to produce a frameshift mutation resulting in a truncated protein (p.Arg578Alafs). To further understand the impact of this *ABCA7* variant on AD pathogenesis, we developed induced pluripotent stem cell (iPSC) lines from two unrelated AA individuals with AD who are heterozygous for the deletion, as well as age-matched AA controls. These lines were validated for pluripotency, genomic stability and sequencing confirmation of the *ABCA7* deletion in the AD iPSC lines. The iPSC lines were differentiated into microglia, as recent studies have implicated a key role for *ABCA7* in amyloid- β (A β) clearance and cellular stress responses in the brain. Reverse transcription PCR analysis of cases demonstrate that a stable RNA transcript is expressed from the *ABCA7* deletion allele. Our preliminary results show that while patient-derived microglia have normal rates of phagocytosis, they are impaired in the uptake and clearance of fibrillar A β . Furthermore, when exposed to the proinflammatory stimulus lipopolysaccharide, patient-derived microglia have decreased cytokine responses. This *ABCA7* deletion is an ancestry-specific alteration in AD that may reduce the ability of microglia to clear A β from the brain and impair responsiveness to proinflammatory signals. The generation of isogenic pairs of deletion bearing and CRISPR-corrected iPSC lines will help to comprehensively examine the contribution of this *ABCA7* deletion to AD pathology in AAs and refine the phenotypic analyses.

Molecular Effects of Genetic Variation Posters - Thursday

PB2514. An Alzheimer's disease risk variant in *TTC3* modifies the growth and transcriptional profile of iPSC-derived forebrain neurons

Authors:

H. Cukier, C. Duarte, J. Laverde-Paz, M. Del Mar Muniz, S. Simon, D. Van Booven, A. Miyares, J. Vance, M. Pericak-Vance, A. Griswold, D. Dykxhoorn; Univ. of Miami, Miami, FL

Abstract Body:

We identified a rare, nonsynonymous variant in the *tetratricopeptide repeat domain 3 (TTC3)* gene that segregated in all 11 Alzheimer disease (AD) individuals in a non-Hispanic white late onset Alzheimer disease (LOAD) family (Kohli, et al, 2016). This missense alteration, rs377155188 (p.S1038C), is predicted to be deleterious and extremely rare in the gnomAD database (allele frequency=3.231x10⁻⁵). Studies have reported that cortical *TTC3* expression is reduced in LOAD patients and negatively correlated with AD neuropathology. To understand the mechanism by which this *TTC3* variant may contribute to LOAD risk, CRISPR-based genome edited was performed to introduce this variant into induced pluripotent stem cells (iPSCs) from an unaffected individual. This resulted in a pair of isogenic iPSC lines which either lacked or were homozygous for the *TTC3* variant. These cells were validated for pluripotency, genomic stability, and potential off target effects from the CRISPR procedure (Laverde-Paz, et al, 2021). These isogenic lines were differentiated to forebrain neurons to examine cellular and transcriptional consequences of this variant. Quantitative PCR analysis demonstrated that *TTC3* levels were decreased in edited compared to unedited iPSCs, as well as differentiated neurons. Cellular motility was assessed in developing neuronal progenitor cells (NPCs) using a wound healing assay. The NPCs bearing the *TTC3* variant were quicker to migrate into the scratch but showed more disorganized movement than that of the parental iPSC-derived neurons. Since other groups have shown evidence that modulation of *TTC3* affects neurite growth, morphological measures of axon formation were assessed in differentiating neuronal cultures using the Incucyte Zoom. The *TTC3* variant bearing neurons had increased neurite outgrowth and branching. This effect of the *TTC3* variant on axon outgrowth phenocopies results found in rat hippocampal neurons in which *TTC3* expression has been silenced. RNA-seq of day 70 neurons identified 1,259 genes that were differentially expressed (FDR < 0.05) between isogenic lines. This included known AD genes (*BACE1*) and genes in AD GWAS loci (*ADAMTS1*, *MAF*, *ZKSCAN*). KEGG pathway analysis identified the PI3K-Akt signaling pathway, which *TTC3* has been previously implicated in, as well as the axon guidance pathway, the GABAergic synapse pathway, and the Wnt signaling pathway as differentially enriched between the parental and *TTC3* variant bearing cells. Combined, these results suggest that the *TTC3* p.S1038C variant acts through a loss of function mechanism.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2515. An integrated machine learning and functional analysis approach for resolution of variants of uncertain significance (VUS) in *PCDH19*.

Authors:

J. Calhoun¹, H. Mefford², S. Schnell³, V. Aguiar-Pulido⁴, M. Ross⁵, J. Parent⁶, L. Isom⁶, EpiMVP Consortium, G. Carvill¹; ¹Northwestern Univ., Chicago, IL, ²St. Jude Children's Res. Hosp., Memphis, TN, ³Univ. of Notre Dame, Notre Dame, IN, ⁴Univ. of Miami, Miami, FL, ⁵Weill Cornell Med., New York, NY, ⁶Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Variants of uncertain significance (VUS) pose a significant challenge for genetic diagnosis of epilepsy, in that these variants can be classified as neither pathogenic nor benign. In order to address this challenge, we established a highly integrated Epilepsy Multiplatform Variant Predictor (EpiMVP) Center Without Walls to develop precise, single-gene *in silico* pathogenicity prediction tools (EpiPred) for the most common epilepsy-associated genes. We are using supervised and unsupervised machine learning algorithms to discriminate pathogenic from benign variants. This approach will be bolstered by functional characterization of a subset of variants in multiple cellular and animal models. Functional data will then be incorporated into the machine learning model to improve classification performance in an iterative fashion. Herein we report our investigation of *PCDH19*, a protocadherin gene family member implicated in *PCDH19*-Epilepsy, that causes a female-limited epilepsy, except in the instances of mosaic males. We first curated a robust training set for the development of our classifier. This consisted of known benign or likely benign (BLB) missense *PCDH19* variants from the general population (gnomAD; n=129) and known pathogenic or likely pathogenic (PLP) missense *PCDH19* variants from multiple sources, including industry partners, Clinvar, and the literature (n=90). Variants were annotated with multiple features (n=39) using *in silico* measures related to evolutionary conservation, protein structure and stability, among others. Our approach employed standard unsupervised and supervised machine learning algorithms to develop an optimal single-gene classifier. Our classifier outputs a score ranging from PLP (high) to BLB (low) and was used to score *PCDH19* VUSs (n=267) collated from ClinVar and industry partners. Our classifier is successful at discriminating *PCDH19* missense BLB and PLP variants ($ROC_{AUC} > 0.9$). We selected representative variants for functional modeling in multiple cellular and animal models. The first iteration of our EpiPred classifier can distinguish between BLB and PLP variants with high specificity. After functional validation in the EpiMVP cellular and animal models, we will improve the predictive power of EpiPred and release this model to the wider scientific and epilepsy community for interpretation of VUS. Our overarching goal is to develop gene-specific classifiers for additional epilepsy-associated genes with an emphasis on non-ion channel genes with the highest volume of VUS and PLP variants identified in routine clinical genetic testing.

Molecular Effects of Genetic Variation Posters - Thursday

PB2516*. An updated genetic atlas of the plasma proteome: Findings from the UK Biobank Pharma Proteomics Project (UKB-PPP)

Authors:

B. Sun¹, J. Chiou², C. Benner³, Y-H. Hsu⁴, T. Richardson⁵, P. Surendran⁶, A. Mahajan⁷, C. Robins⁸, S. Vasquez-Grinnell⁹, L. Hou¹⁰, E. Kvikstad⁹, L. Ward¹¹, C. Willis¹², O. Burren¹³, N. Telis¹⁴, S. Welsh¹⁵, L. Burkitt-Gray¹⁵, UKB-PPP Consortium, E. Smith¹⁶, S. Szalma¹⁷, M. Black¹⁸, S. Petrovski¹⁹, J. Howson⁵, H. Kang²⁰, L. Mitnaul²¹, B. Gibson²², M. Miller², C. Whelan²³; ¹Biogen, Cambridge, MA, ²Pfizer, Cambridge, MA, ³Genentech, South San Francisco, CA, ⁴Amgen, Cambridge, MA, ⁵Novo Nordisk Res. Ctr. Oxford, Oxford, United Kingdom, ⁶GlaxoSmithKline, Stevenage, United Kingdom, ⁷Genentech, San Francisco, CA, ⁸GSK, Atlanta, GA, ⁹Bristol Myers Squibb, Princeton, NJ, ¹⁰Janssen Res. & Dev., LLC, Spring House, PA, ¹¹Alnylam, Cambridge, MA, ¹²Alnylam Pharmaceuticals, Cambridge, MA, ¹³Discovery Sci., R&D, AstraZeneca, Royston, United Kingdom, ¹⁴Calico Life Sci., South San Francisco, CA, ¹⁵UK Biobank, Stockport, United Kingdom, ¹⁶Takeda California, Inc, SAN DIEGO, CA, ¹⁷Takeda, San Diego, CA, ¹⁸Janssen Res. & Dev. (Johnson & Johnson), Lansdale, PA, ¹⁹AstraZeneca, Cambridge, United Kingdom, ²⁰Regeneron Genetics Ctr., Tarrytown, NY, ²¹Regeneron, Piscataway, NJ, ²²Amgen, South San Francisco, CA, ²³Biogen Inc., Cambridge, MA

Abstract Body:

Genetic analyses of population biobanks are increasingly employed as tools for drug target discovery and validation. High-throughput, population-scale proteomics holds potential to bridge the gap between the human genome and human diseases, improving genetics-guided drug development. Here, we describe results from the first phase of the UK Biobank Pharma Proteomics Project (UKB-PPP); a public-private collaboration between the UK Biobank (UKB) and 13 biopharmaceutical companies profiling concentrations of an initial 1,463 plasma proteins across 54,306 UKB participants using the multiplex, antibody-based Olink™ Proximity Extension Assay. The study identifies 10,248 genetic associations with 1,377 proteins, including 1,163 protein quantitative trait loci (pQTLs) acting in *cis*- and 9,085 acting in *trans*. 85% of pQTLs are not previously described. We highlight the utility of these pQTL data for disease risk etiology and drug development through specific deep-dive examples, including: (i) validating and enhancing the Mendelian randomization derived relationships between PCSK9 concentration and lipid concentrations, cardiovascular disease and stroke; (ii) disentangling the contributions of lung-expressed proteins towards COVID-19 susceptibility; (iii) identifying a BAG3 missense variant affecting cardiac small heat shock protein complexing (BAG3-HSPB6), with downstream effects on blood biomarkers of heart failure (NT-proBNP and proBNP); (iv) underlining examples of complex network *trans* QTLs linking protein-protein interactions and biological pathways, including cytokine and complement pathways; and (v) highlighting long-range epistasis between ABO blood group (chr 9) and FUT2 secretor status (chr 19) on protein levels, with conserved expression in gastrointestinal tissues. We also make empirical estimates of how pQTL detection scales with sample size and number of proteins measured, providing context for current and future proteogenomic studies. Our private-public collaboration highlights the value of precompetitive investments in population biobanks, providing the scientific community with an open-access resource of unprecedented depth and breadth. Future UKB-PPP initiatives will further expand proteomic coverage, integrate additional longitudinal samples, and pilot other affinity- and mass spectrometry-based technologies, aiming to achieve more complete coverage of the plasma proteome and its genetic regulation at population scale.

Molecular Effects of Genetic Variation Posters - Thursday

PB2517. Antibody repertoire heavy chain gene usage is explained by common genetic variants in the immunoglobulin heavy chain locus.

Authors:

O. Rodriguez¹, Y. Safonova², C. Silver¹, K. Shields¹, W. Gibson¹, J. Kos¹, D. Tieri¹, H. Ke³, K. Jackson⁴, S. Boyd⁴, M. Smith¹, W. Marasco³, C. Watson¹; ¹Univ. of Louisville Sch. of Med., Louisville, KY, ²Johns Hopkins Univ., Baltimore, MD, ³Harvard Med. Sch., Boston, MA, ⁴Stanford Univ. Sch. of Med., Stanford, CA

Abstract Body:

Antibodies (Abs) are critical components of the adaptive immune system. The heavy chains of Abs are encoded by >100 variable (V), diversity (D), and joining (J) gene segments within the immunoglobulin heavy chain (IGH) locus. IGH is among the most structurally complex genomic regions, containing extensive haplotype diversity, with elevated numbers of single nucleotide variants (SNVs) and structural variants (SVs). Mounting evidence shows that IGH variants influence the expressed Ab repertoire, with potential consequences on Ab function in human health. However, due to the complexity of IGH, which has hindered the use of standard high-throughput sequencing and genotyping approaches, genetic variants contributing to Ab repertoire variation have not been fully characterized. To address this, we conducted the first genetic association study to identify IGH variants that impact expressed Ab repertoire signatures. Utilizing our novel targeted long read sequencing approach, we generated paired IGH haplotype and expressed Ab repertoire sequencing data in 154 healthy adults, the largest dataset of its kind. Variants identified included 27 SV alleles from 8 large (> 9Kb) SVs, ranging from 9.5 to 284 Kb, 966 indels and 71 small SVs (< 9 Kb), and 7,980 common SNVs, 5,057 (64%) of which were missing from or contained no allele frequency data in dbSNP. Our ability to characterize the full spectrum of genetic variants in IGH for the first time demonstrated the impact of accurate long reads for resolving complex loci. Our analysis revealed IGH genetic variants strongly influence the composition of the expressed Ab repertoire through both coding and non-coding SNPs, and large SVs. In total, 3,464 variants were associated with the usage of 58 (73%) IGH genes. This included genetic effects on either single genes, or multiple genes, demonstrating coordinated gene regulation. Importantly, associated SNVs overlapped GWAS variants, and were enriched in regulatory regions, including CTCF binding sites, a transcription factor involved in V(D)J recombination. These data provide the first robust links between IGH polymorphisms and the Ab repertoire, dispelling the long held belief that Ab diversity arises solely through stochastic processes. These datasets will (1) allow for the discovery of novel genomic factors and molecular mechanisms influencing Ab repertoire development and diversity, (2) facilitate better characterization of the Ab response in disease, and (3) inform the design of improved therapeutics, vaccines and implementation strategies.

Molecular Effects of Genetic Variation Posters - Thursday

PB2518. *APOE* Genotype Contributes to Distinct Profiles in Human iPSC Astrocytes and Neurons

Authors:

S. Clayton¹, R. Panitch², Y. You³, T. Ikezu³, L. Farrer⁴, G. Jun⁵; ¹Boston Univ., Brighton, MA, ²Boston Univ., potomac, MA, ³Mayo Clinic Florida, Jacksonville, FL, ⁴Boston Univ Sch Med, Boston, MA, ⁵Boston Univ., Boston, MA

Abstract Body:

The APOE gene is the major risk factor for Alzheimer disease (AD) risk, the $\epsilon 4$ allele increases while the $\epsilon 2$ allele decreases risk for AD. However, differential expression profiles between $\epsilon 2$ and $\epsilon 4$ alleles in neurons and astrocytes have not been well-characterized. We previously developed human isogenic APOE allele-specific induced pluripotent stem cells (iPSC)-derived excitatory neurons (iNeurons) and astrocytes (iAstrocytes) in a co-culture system (Jun, 2022). We conducted RNA-sequencing in 6 isogenic iNeurons and iAstrocytes from different APOE genotypes: APOE KO, $\epsilon 2/\epsilon 2$ (APOE2), $\epsilon 3/\epsilon 3$ (APOE3), and $\epsilon 4/\epsilon 4$ (APOE4). We conducted differential expression analysis using previously identified AD genes as well as transcriptome-wide data between APOE2 and APOE4 iPSCs separately in iNeurons and iAstrocytes. We conducted pathway enrichment analysis with significantly differentially expressed genes (DEG) with false discovery rate adjusted p (adjP) <0.05 between APOE2 and APOE4 iPSCs. APOE expression was highest in APOE2 iAstrocytes followed by APOE3 and APOE4 (log₂ of APOE2:APOE3:APOE4=9.28:8.81:8.50; p <0.05), while APOE expression in iNeurons was inconsistent (log₂ of APOE2:APOE3:APOE4=4.65:5.28:4.07; 0.05 $<p<0.2$). We also observed differentially expressed genes between APOE2 and APOE4 for the established AD genes with adjP <0.05 , including JAZF1, PRKD3, and ICA1 in iNeurons (best gene, JAZF1: log₂ fold change=-1.38, adjP=2.42E-8) and MAF, DOCK1, and C4A in iAstrocytes (best gene, MAF: log₂ fold change=1.06, adjP=1.81E-4). In transcriptome-wide analysis, we identified significantly DEGs between APOE2 and APOE4 genotypes including BMPER, SRY, and REEP5 in iNeurons whereas FEZF1, PCSK9, and NTRK1 in iAstrocytes with adjP <0.05 . Pathways significantly downregulated in APOE4 compared to APOE2 in iAstrocytes included pathways related to response to amyloid-beta, extracellular matrix structure and organization, and nervous system development (enrichment adjP <0.05). In iNeurons, pathways significantly upregulated in APOE4 include immune-related pathways, cellular senescence, and systemic lupus erythematosus (enrichment adjP <0.05). These results demonstrate that APOE genotype contributes to distinct expression profiles in iAstrocytes and iNeurons.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2519. Assessing the contribution of variants with functional effects on post-transcriptional regulation to genetics of psychiatric disease

Authors:

J. Gu, X. He; Univ. of Chicago, Chicago, IL

Abstract Body:

Identifying causal variants for complex traits remains challenging, as most GWAS identified variants are located in non-coding regions, without clear functions. A major challenge is thus to annotate functional effects of variants. Existing work along this line includes various experimental and computational approaches such as enhancer mapping, transcription factor binding site profiling and expression QTL studies. These efforts have been largely focused on putative transcriptional effects of variants. Numerous studies, however, have supported the importance of post-transcriptional regulatory (PTR) processes in development and diseases, such as RNA splicing, N⁶-methyladenosine (m6A) modification, and alternative polyadenylation. Nevertheless, the importance of variants with possible PTR effects in genetics of complex traits remains largely unexplored. This is particularly the case for m6A modification, which is a key process regulating RNA stability and translation. There have been very few studies directly linking m6A modification to human genetic variations.

To fill these gaps, we annotated GWAS SNPs with putative PTR effects using two sources of data. We collected m6A profiles of human fetal tissues from a recent study and computed scores for splicing effects using two deep-learning based prediction tools, Spidex and SpliceAI. Applying Stratified-LD score regression to Schizophrenia (SCZ) GWAS, we found that m6A sites from fetal brains show ~9-fold enrichment with disease heritability. The predicted splice variants, on the other hand, show no significant enrichments. We estimate that transcriptional features, defined by open chromatin regions from neuronal cells explain about 24.9% of SCZ heritability, while fetal m6A sites explain about 4.5% heritability. These results thus suggest that m6A features make a substantial contribution to disease genetics. The lack of enrichment of splicing features, may reflect the limitation of current splicing prediction tools to accurately predict splicing effects of variants.

Molecular Effects of Genetic Variation Posters - Thursday

PB2520. Assessing the genetic control of whole blood RNA editing and its role in modulating risk of common diseases.

Authors:

D. Stacey¹, E. Aiton¹, C. Acciari¹, A. Nath^{1,2}, E. Persyn¹, J. Marten¹, T. Vanderstichele³, K. Walter³, K. Kundu³, D. Roberts⁴, E. Di Angelantonio¹, J. Danesh¹, A. Butterworth¹, N. Soranzo³, E. E. Davenport³, M. Inouye^{1,2}, D. S. Paul^{1,5}; ¹Univ. of Cambridge, Cambridge, United Kingdom, ²Baker Heart and Diabetes Inst., Melbourne, Australia, ³Wellcome Trust Sanger Inst., Hinxton, United Kingdom, ⁴Univ. of Oxford, Oxford, United Kingdom, ⁵AstraZeneca, Cambridge, United Kingdom

Abstract Body:

RNA editing is a naturally occurring mechanism by which nucleotides in RNA sequence are chemically modified. This leads to nucleotide changes in RNA that potentially impact gene function, but that are not detectable by DNA sequencing. RNA editing is a vastly under-explored area, with major gaps in knowledge concerning (i) the functional effects of non-coding RNA edit sites and (ii) the molecular chain of events by which RNA edit sites might impact disease.

Using linked RNA-sequence and whole-genome sequence data from 2,918 whole blood samples from healthy participants of the INTERVAL study, we have called a 'high-confidence' set of 2,575 RNA edit sites. In accordance with previously published datasets, these high-confidence sites are enriched in Alu elements, they reside primarily in 3' UTRs, and are depleted of recoding (i.e., protein-altering) events. About 20% are currently not annotated in REDIportal, the largest database cataloguing known RNA edit sites. Furthermore, in support of the validity of our RNA edit dataset, we found that higher whole blood mRNA levels of *ADARI* - which encodes the key enzyme involved in catalysing RNA editing in blood cells - are significantly ($p < 2 \times 10^{-16}$) associated with higher overall editing levels.

To assess the genetic control of RNA editing, we have performed preliminary edit QTL analyses using matrix eQTL, with the high-confidence edit sites as outcome variables. We have uncovered significant ($p < 1.29 \times 10^{-12}$) associations at 3,415 independent loci. Approximately one third of these associations relate to edit sites at the immunoglobulin locus on chromosome 2, suggesting a genetically controlled role for RNA editing in modulating immunoglobulins produced by B cells. A pheWAS indicated that of the 3,415 independent loci, 269 are associated with at least one disease or trait at genome-wide significance. We aim to assess the extent to which these loci might play a causal role in disease by performing statistical colocalization and Mendelian randomization analyses. Finally, we aim to leverage the multiomic data available in the INTERVAL study to better define the causal chain of events leading to disease risk. We anticipate our findings will enhance our understanding of the genetic regulation of RNA editing and the role it plays in disease.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2521. Assessment of genotype-phenotype correlation of human *TEX11* variants associated with non-obstructive azoospermia in mouse

Authors:

J. Hardy¹, C. Doungkamchan¹, N. Pollock², J. Kuong², K. Tran¹, Y. Sheng², T. Jaffe³, M. Olszewska⁴, M. Kurpisz⁴, F. Tuttelmann⁵, M. Brieno-Enriquez¹, K. Orwig¹, A. Yatsenko¹; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²Univ. of Pittsburgh Med. Ctr., Pittsburgh, PA, ³West Virginia Univ., Morgantown, WV, ⁴Polish Academy of Sci., Poznan, Poland, ⁵Univ. of Munster, Munster, Germany

Abstract Body:

Introduction Approximately 10% of infertile men suffer from azoospermia or complete absence of mature spermatozoa in the ejaculate. To date, as much as 50% of all male infertility cases are estimated to be the result of genetic abnormalities. Variants in *TEX11*, a testis specific gene involved in meiosis are posited to account for ~2% of all non-obstructive (w/o physical impairment to the reproductive tract) azoospermia (NOA) cases. A 214kb deletion in the 224kb mouse *Tex11* orthologue was shown to cause early meiotic arrest yet pathogenicity and functional consequence of variants associated with human male infertility have yet to be validated. Here we utilized a similar transgenic mouse model approach to assess causality of *TEX11* point mutations and indels associated with NOA in humans.

Methods CRISPR/Cas9 technology was used to introduce 3 likely pathogenic *TEX11* variants, one splice site in the meiosis specific SPO22 domain, one frameshift leading to a premature stop in the 2nd structural tetratricopeptide repeat (TPR) motif, and one missense in the 3rd TPR into the homologous mouse gene. Fertility status, testis morphology, seminiferous tubule histology, immunohistochemistry, in situ hybridization, and prophase I analysis were assessed using standard and advanced molecular techniques.

Results In 12-week old, BL6/D2 F1 mice, each of the 3 variants resulted in testicular phenotypes that at least partially mimicked the human condition. The frameshift variant which caused infertility due to arrest in pachynema most strongly resembled the human phenotype of complete meiotic arrest though mouse spermatocytes were able to progress to metaphase. Both point mutations did not result in infertility at this age/generation though defects in spermatocyte chromosome segregation similar to the knockout were noted. By the F3 generation, mice harboring the splice site variant had a statistically reduced epididymal sperm count and testis weight began trending toward significance (p=0.053402).

Conclusion Human *TEX11* shares 74% sequence identity with the mouse orthologue. The unique 26% could plausibly account for the disparity in spermatozoa production observed with human *TEX11* variants versus the mouse complement. Additionally, the 15% of the total genome that is not shared could allow for a compensatory mechanism in mouse. Alternatively, *TEX11* induced NOA may be a progressive disorder with increasing severity due to age and/or transgenerational effect. Long term analysis of germ cell dynamics in mouse may provide additional insight into the consequence and mechanism of identified variants. This work was supported by P50 HD096723 to KEO and ANY and T32 HD087194 to JJH.

Molecular Effects of Genetic Variation Posters - Thursday

PB2522. Association of GWAS and candidate gene loci of dopaminergic system with major depression, schizophrenia and bipolar disorder in the Pakistani population

Authors:

A. Hashmi¹, R. Dharejo², R. Taj³, M. Ajmal¹, Z. Agha⁴, R. Qamar⁵, M. Azam⁶; ¹Comsats Univ. islamabad, Islamabad, Pakistan, ²WAPDA Administrative Staff Coll., Lahore, Pakistan, ³Pakistan Inst. of Med. Sci., Islamabad, Pakistan, ⁴COMSATS Univ. ISLAMABAD, ISLAMABAD, Pakistan, ⁵Sci. and Technology Sector, ICESCO,, Rabat, Morocco, ⁶COMSATS Univ. Islamabad, Islamabad, Pakistan

Abstract Body:

Background: The dopaminergic pathways control neural signals that modulate mood and behaviour along and have a vital role in the aetiology of major depression (MDD), schizophrenia (SHZ) and bipolar disorder (BD). Genome-wide association studies (GWAS) have reported several dopaminergic pathway's and other genetic loci's association with these disorders however, no such comprehensive data was available regarding the Pakistani population. **Aim:** the present study was conducted to analyse the GWAS and candidate gene loci of the dopaminergic and cognitive system genes in MDD, SHZ, and BD, in the Pakistani population. **Methods:** A total of 1237 subjects [MDD n=479; BD n=222; SHZ n=146; and controls n=390], were screened for eleven genetic variants through polymerase chain reaction (PCR) techniques. Univariate followed by multivariate logistic regression analysis were applied to determine the genetic association. **Results:** Significant risk associations were observed for rs4532 and rs1799732 with MDD; and rs1006737 and rs2238056 with BD. However, after applying multiple test corrections rs4532 and rs1799732 association did not remain significant for MDD. Moreover, a protective association was found for three variants; DRD4-120bp, rs10033951 and rs2388334 in the current cohort. **Conclusion:** The current study revealed the risk association of SNPs rs1006737 and rs2238056 with BD and the protective effect of the DRD4-120bp variant in MDD and BD, of rs2388334 in BD and of rs10033951 in MDD, BD, and SHZ in the current Pakistani cohort. Thus, the study is valuable in understanding the genetic basis of MDD, BD and SHZ in the Pakistani population, which may pave the way for future functional studies.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2523. Association of MTHFR C677T polymorphism with obesity in women: A case-control study and meta-analysis

Authors:

M. Irfan, R. Rani; Pir Mehr Ali Shah, Arid Agriculture Univ., Rawalpindi, Pakistan

Abstract Body:

Methylenetetrahydrofolate reductase (MTHFR) is an enzyme that stimulates and modulates folate metabolism in the body. The main and specific function of MTHFR is to convert 5, 10-Methylenetetrahydrofolate into 5-Methylenetetrahydrofolate (5-MTHF), which is active folate. 5-MTHF is a key component in the synthesis of nucleotides for the production of RNA and DNA by methylation reactions; production of S-Adenosylmethionine (SAM); methylation of DNA, proteins, neurotransmitters (NTs), & phospholipids; and also the conversion of homocysteine into methionine by remethylation reaction. Many genetic variations in the MTHFR enzyme are recognized and reported which alter the folate metabolism hence methylation and overall health. Numerous studies have shown that low folate levels lead to an increased body mass index. Therefore, in the present study was hypothesized that the genetic polymorphisms of the MTHFR gene might be related to obesity resulting in decreased levels of active folate. The objective of this study was to determine the association of a specific MTHFR C677T polymorphism with obesity in obese women. Blood samples (5ml) along with the data on sociodemographic, physical health and lifestyle aspects were also collected. DNA was extracted from the blood samples by using a standard phenol-chloroform method and stored at -20°C. The specific primers were designed and optimized for the gene. The PCR amplified products were digested by a specific restriction enzyme (Hinf I) to determine polymorphisms. The digested product was electrophoresed on agarose gel with ethidium bromide staining which was then visualized through ultraviolet transillumination. The fragments of wild type and mutated DNA were obtained on the gel. The allele frequency of the C to T polymorphism was determined by counting alleles through electrophoresis gel analysis. Chi-square analysis was used to determine the Hardy-Weinberg equilibrium of the alleles in the population. The association of the polymorphism with obesity was determined by logistic regression analysis adjusting the effects of confounding factors i.e. age, socioeconomic factors and lifestyle. A p-value of < 0.05 was considered statistically significant. We found no association of MTHFR C677T with obesity in women. Furthermore, the meta-analysis also confirmed our result of no association between obesity in women and MTHFR C677T.

Molecular Effects of Genetic Variation Posters - Thursday

PB2524. Bayesian fine-mapping of sex-specific gene regulation during neurodevelopment shows link to neurodevelopmental disorders

Authors:

J. Lalli¹, S. Pochareddy², D. Liang², R. Kovner², J-Y. An³, N. Sestan², S. Sanders⁴, D. Werling¹; ¹Univ. of Wisconsin-Madison, Madison, WI, ²Yale Univ., New Haven, CT, ³Korea Univ., Seoul, Korea, Republic of, ⁴Univ. of California - San Francisco, San Francisco, CA

Abstract Body:

Autism spectrum disorder (ASD) is diagnosed more commonly in males than females, and multiple lines of evidence suggest that individuals with the same autosomal genetic background are more likely to be diagnosed with ASD if they are male. While it is possible that female sex has a neuroprotective effect against ASD during early neurodevelopment, our understanding of this phenomenon is hampered by limited knowledge of the effect of sex on gene regulation during neurodevelopment.

To address this problem, we utilized a previously published dataset of bulk RNA-seq and whole genome DNA sequences from the dorsolateral prefrontal cortices (DLPFCs) of 81 human fetuses with gestational ages of 14 to 21 weeks, when sex hormone levels peak during development. To determine if a SNP was significantly associated with a nearby gene's expression level, we compared the use of FDR correction, permutation analysis, and Bonferroni correction using the of effective tests performed per gene as determined by eigenvector decomposition. We ultimately applied a previously published Bayesian fine-mapping technique to all SNPs in cis with expressed genes in our dataset.

We identified 335 credible sets of SNPs that affect the expression of 297 nearby genes in a sex-specific manner. Gene ontology analysis suggests that these genes are significantly associated with neurodevelopmental disorders, and are associated with E2F1 and EGR1 regulatory pathways - critical for neuroprogenitor proliferation and brain plasticity, respectively. Our work suggests that genes important for neurodevelopment are differentially regulated in male and female DLPFCs during a critical period, providing a potential mechanism for the sex-biased prevalence of some neurodevelopmental disorders.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2525. Body mass index as an environmental context in functional genomic annotations.

Authors:

R. Signer¹, A. de Pins¹, H. Young¹, A. Cote², C. Seah¹, J. Johnson¹, L. Huckins¹; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²Icahn Sch. of Med., New York, NY

Abstract Body:

Background Genome-wide association studies (GWAS) have succeeded in identifying numerous disease risk loci while expression quantitative trait loci (eQTL) studies have annotated tissue-specific functional effects across the genome. Despite advances in recruitment and annotation, significant challenges remain translating estimates learned from the population back to the individual. Body mass index (BMI) represents a physiologic state with psychiatric and non-psychiatric disorder associations and is a source of variability amongst cases and controls. Additionally, BMI plays a critical and poorly understood role in eating disorder diagnosis, treatment, and genomic models. Here, we explore the dynamic nature of eQTLs in the context of BMI, test for associations with phenotypes including anorexia nervosa (AN), and define the utility of including physiologic measures in functional genomics to advance precision medicine.

Methods We used matched whole genome- and transcriptome-sequencing from 46 tissues in the Genotype-Tissue Expression (GTEx) project (Including all tissues with > 100 samples; N_{total}=950) to identify BMI-dynamic eQTL and eGenes. We first considered a continuous BMI and tested for a BMI*SNP interaction effect. BMI-dynamic eGenes had at least one SNP with a significant interaction after applying a Bonferroni correction for N_{SNPs} in the cis-eGene window. We next considered categorical BMI and called eQTL within BMI quintile. Finally, we imputed gene expression from psychiatric, immune, and other BMI-associated disorder summary statistics using S-prediXcan and tested for overlap with BMI-dynamic eGenes.

Results We identified 9,750 BMI-dynamic eGenes across all tissues (39.9% of all eGenes tested). Tissues with the most BMI-dynamic eGenes included four brain regions: cortex, caudate, nucleus accumbens, putamen and four digestive tissues: small intestine, salivary gland, transverse colon, and gastroesophageal junction. Minor salivary gland emerged as a BMI-responsive immune organ, with eGenes enriched for innate immune response ($p=1.50 \times 10^{-4}$) and interferon signaling ($p=2.01 \times 10^{-4}$). BMI-dynamic eGene-tissue pairs overlap with significant AN-associated eGenes at loci including *MGMT*-Lymphocytes. BMI-responsive eGenes were enriched in ovary ($p=0.0020$) amongst nominally significant AN-associated eGenes.

Conclusion Our results show the dynamic nature of eQTL across the BMI spectrum and demonstrate the relevance of BMI in characterizing functional loci within psychiatry. These data support the inclusion of individual environmental measures in complex disorder risk models to enhance precision of effect estimates.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2526. Calibration set for high throughput functional assay for *RYR1* Variants.

Authors:

J. Johnston¹, A. Douglass¹, R. T. Dirksen², K. Stowell³, L. G. Biesecker¹; ¹NHGRI, NIH, Bethesda, MD, ²Univ. of Rochester Med. Sch., Rochester, NY, ³Massey Univ, Palmerston North, New Zealand

Abstract Body:

The ClinGen malignant hyperthermia susceptibility (MHS) variant curation expert panel specified the ACMG/AMP criteria for *RYR1*-related MHS and classified 335 *RYR1* variants. Sixty-five percent of these variants (218) were classified as variants of uncertain significance (VUS) due to limited evidence supporting pathogenicity. Prioritization of functional studies is needed to resolve the large number of VUS classifications and allow for appropriate risk assessment. However, Brnich et al. (2020) concluded that functional studies often do not have adequate controls to support using the data in the ACMG classification scheme. For supporting strength, according to the Bayesian model of combining the ACMG criteria (Tavtigian et al. 2018), the evidence should support an odds of pathogenicity of 2.08:1 and for moderate strength an odds of pathogenicity of 4.3:1, based on likelihood ratios derived from experimental datasets. *RYR1* is a calcium release channel and numerous functional studies are available to measure the release of calcium in response to agonist. However, few benign variants have been analyzed, limiting the strength of the functional evidence. In this work, we set out to design a validation set of *RYR1* variants that will allow calibration of *RYR1* functional assays. The calibration set comprises 17 *RYR1* variants classified as benign and six *RYR1* variants classified as pathogenic (5) or likely pathogenic (1). Variants were introduced into the full length human *RYR1* cDNA using standard subcloning and site directed mutagenesis. All variants were verified for sequence integrity. Stable expressing cell lines were generated in HEK293 cells using the pcDNA5/FRT mammalian expression vector (ThermoFisher Scientific). Caffeine concentration-response curves will be determined using the Flexstation 3 Plate Reader (Molecular Devices) with the goal of validating this assay for use in ACMG pathogenicity classification. These variants will also be used to develop a large-scale functional assay for *RYR1*. In summary, we have prioritized functional studies in *RYR1* to resolve the large number of VUS classifications. This validation set will support calibration of functional assays using objective data, contributing to the transition of variant classification from an art to a science.

Molecular Effects of Genetic Variation Posters - Thursday

PB2527. CD4+ T cell gene expression influences COVID-19 severity and susceptibility: results from single cell sequencing and Mendelian randomization

Authors:

J. Willett^{1,2}, T. Lu^{1,2}, S. Yoshiji^{1,2}, T. Nakanishi^{1,2}, G. Butler-Laporte^{3,2}, S. Zhou¹, Y. Farjoun⁴, B. Richards^{1,2}; ¹McGill Univ., Montreal, QC, Canada, ²Lady Davis Inst., Montreal, QC, Canada, ³McGill Univ., Montréal, QC, Canada, ⁴Lady Davis Inst., JGH, Montreal, QC, Canada

Abstract Body:

Background: Severe COVID-19 is due, in part, to immune hyperstimulation. The dynamics of this immune response are poorly understood. Using cis-expression quantitative loci (cis-eQTL) from single-cell RNA (sc-RNA) sequencing of CD4+ T-cells before and after experimental stimulation, we hypothesized colocalization of these dynamic cis-eQTLs with COVID-19 outcomes would identify targets for disease prevention and treatment. We assessed transcripts detectable before and during infection and in cells relevant regardless of one's previous exposure to antigen (naïve and memory cells).

Methods: We colocalized q-value corrected ($q < 0.1$) cis-eQTLs from CD4+ sc-RNA T-cell data from 119 healthy, British ancestry individuals against 87 loci in European-stratified Host Genetics Initiative (HGI) release 7 summary statistics for COVID-19 outcomes (very severe, defined as death or requiring ICU care, severe requiring hospitalization, or susceptibility). We identified estimated causal genes among pre- (CD4+ naïve and memory) and post-Leiden algorithm clustered (T naïve, T effector memory, T central memory, T regulatory) cell populations, defined by strong colocalization (posterior probability ≥ 0.8) and passing sensitivity testing. We conducted Mendelian randomization (MR) with sensitivity and directionality testing for strong colocalization examples.

Results: T regulatory cell expression did not colocalize for any outcome. Three genes colocalized to all outcomes (NAPSA, CAT, RALGDS). NAPSA colocalized in whole-blood bulk-RNA sequenced GTEx v8 EUR-stratified data, corresponding to a noninfectious state. It was not expressed in sc-RNA CD4+ cells pre-stimulation. NAPSA expression in T effector memory cells 40 hours post-stimulation, per MR, influenced all outcomes (95% confidence interval odds ratio (OR) very severe 1.18 [1.10, 1.27], severe 1.15 [1.10, 1.20], susceptibility 1.04 [1.03, 1.06]). Pre-stimulation CD4+ naïve catalase expression influenced very severe and severe COVID-19 (OR 1.21 [1.08, 1.40] and 1.13 [1.01, 1.28]). RALGDS expression 16 hours post-stimulation in T effector memory cells was protective for all outcomes (OR very severe 0.78 [0.71, 0.87], severe 0.87 [0.79, 0.96], susceptibility 0.88 [0.86, 0.90]). Another gene, RAB2A, influenced severe COVID-19 in CD4+ memory cells 40 hours post-stimulation (OR 1.67 [1.38, 2.01]). NAPSA expression can be inhibited by pepstatin A, CAT can be inhibited by fomepizole and cannabidiol, and RALGDS can be stimulated by STAT3 signalling.

Conclusions: These results implicate specific cell types, cell states and transcripts that influence COVID-19 severity and susceptibility.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2528. Cellular consequences of Hirschsprung disease-associated *RET* variants studied utilizing CRISPR-mediated genome engineering.

Authors:

L. Fries, A. Chakravarti, S. Chatterjee; NYU Grossman Sch. of Med., New York, NY

Abstract Body:

Genome sequencing studies have led to vast catalogs of disease-associated genes and variants, but biological studies of how they mediate disease are lagging. This is in part owing to a lack of targeted functional screens that can quantify the effects of variants on specific cellular processes underlying the disorders.

Hirschsprung disease (HSCR), a developmental disorder, arises due to the inability of enteric neural crest derived cells to proliferate and migrate to form a functional enteric nervous system. Despite extensive gene and variant heterogeneity, ~50% patients harbor coding and enhancer variants at the receptor tyrosine kinase gene *RET*. Therefore this study focuses on understanding how *RET* variants lead to aganglionosis. CRISPR-based prime editing was used to engineer multiple HSCR-associated rare coding *RET* variants in the neuroblastoma cell line CHP-212. Three tyrosine variants, with varying degrees of predicted pathogenicity, spanning the entire protein were made: Y204X, Y791F, and Y981X. Quantitative functional screens to measure proliferation and migration in stable heterozygous mutant lines demonstrated that all variants decreased ability of cells to perform these functions, reflective of a HSCR phenotype; with Y204X, Y791F and Y981X leading to 60%, 45%, and 30% ($P < 0.001$ for all) reduced migration, respectively. Thus, earlier protein truncation is associated with a more severe cellular effect, though missense mutations can be as severe as nonsense changes.

Additionally, the effect of the HSCR associated common enhancer variant rs2435357 in *RET* intron 1 was studied by engineering mice via CRISPR based deletion of the homologous region. This mutant was crossed to create a compound mutant carrying two deletions, in the *Ret* gene and the enhancer, modeling HSCR patients with both coding and non-coding variants. An allelic series was created, demonstrating mice carrying homozygous enhancer deletion showed a 20% ($P < 0.001$) reduction in *Ret* expression in the embryonic gut at E14.5 compared to the wildtype, 50% ($P < 0.001$) reduction in heterozygous *Ret* knockout, and a 60% ($P < 0.001$) reduction in the compound mutant. Single cell RNA-sequencing was completed on these mutants in conjunction with immunohistochemistry to examine specific neuronal cell types altered.

These functional screens demonstrate that coding and noncoding variants have additive effects on gene expression leading to variable cellular phenotypes that don't necessarily parallel predicted pathogenic risk scores. These assays provide a route to functionally quantify pathogenicity of variants discovered in patients of HSCR and other neurodevelopmental motility disorders.

Molecular Effects of Genetic Variation Posters - Thursday

PB2529. Characterization of a leaky splice mutation in *AIRE*, associated with milder phenotype and late debut of autoimmune polyendocrine syndrome type 1.

Authors:

A. Berger¹, B. E. Oftedal¹, E. Bratland^{1,2}, S. Johansson^{2,1}, E. S. Husebye^{1,2}; ¹Univ. of Bergen, Bergen, Norway, ²Haukeland Univ. Hosp., Bergen, Norway

Abstract Body:

Autoimmune Polyendocrine Syndrome type 1 (APS-1) is a rare monogenic autoimmune disorder caused by mutations in the gene *AIRE*. *AIRE* or the Autoimmune Regulator acts as a transcriptional regulator in the thymus, facilitating the expression of thousands of genes important for the process of negative T-cell selection and immunological tolerance against self. Most APS-1 causative mutations in *AIRE* result in a quite severe phenotype with the three main manifestations being autoimmune Addison's disease, hypoparathyroidism and chronic mucocutaneous candidiasis, although many patients also exhibit a variety of other autoimmune manifestations. However, not all mutations lead to a severe phenotype which is illustrated by a cohort of four patients with a mutation in the canonical donor splice site of exon 7, c.879+1G>A, exhibiting a milder phenotype with relatively late debut. Utilizing a C57BL/6 crispr knock-in mouse model we investigate the transcriptional impact of c.879+1G>A with bulk RNA sequencing of FACS sorted medullary thymic epithelial (mTEC) cells, with cells from wildtype mice and mice homozygous for the nonsense mutation p.C313X as controls. We confirm the exon skipping effect of c.879+1G>A without frameshift of the open reading frame, leading to a shortened *AIRE* without the amino acids 267 to 293, encoded by exon7. However, this splicing effect may be leaky and while substantially reduced, there are mRNA reads within exon 7 which may be evidence of some full-length *AIRE* expression. On the transcriptional level the effect of c.879+1G>A is less severe than that exhibited by p.C313X showing an intermediate pattern compared to wildtype *AIRE* with higher relative and absolute expression of *AIRE*-regulated genes compared to p.C313X and fewer significantly downregulated genes (among wildtype *AIRE* regulated genes) in total. This transcriptional effect may either be caused by the aforementioned leaky expression of some full-length *AIRE* or by some residual effect of *AIRE* lacking amino acids 267-293. We confirm that the milder phenotype exhibited by APS-1 patients with the c.879+1G>A mutation in *AIRE* correlates to a less severe effect on the transcriptional level in the mTEC cells of the thymus of mice carrying the corresponding mutation, compared to other known deleterious *AIRE* mutations.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2530. Characterization of haplotype diversity in the human immunoglobulin heavy chain constant region using long-read sequencing

Authors:

U. Jana¹, O. Rodriguez², A. Dorgham¹, W. Gibson¹, K. Shields¹, C. Silver³, M. Emery³, G. Deikus³, R. Sebra⁴, E. Eichler⁵, M. Smith³, C. Watson³; ¹Univ. of Louisville, Louisville, KY, ²Univ. of Louisville Sch. of Med., Louisville, KY, ³Icahn Sch. of Med. at Mount Sinai, New York, NY, ⁵Univ of Washington, Seattle, WA

Abstract Body:

The heavy chain constant region of antibodies (Ab) mediates immune function by playing a pivotal role in neutralization and elimination of antigens through various effector functions. Critically, these functions are mediated by different Ab isotypes (IgM, IgG, IgA, IgE and IgD) each encoded by genes within the immunoglobulin heavy chain constant (IGHC) locus located near the telomeric end of chromosome 14. Despite previous studies highlighting extreme allelic and structural variation within human IGH, our knowledge of haplotype diversity remains limited. This stems from technical barriers that have hindered the use of standard next-generation sequencing for resolving these complex loci, stressing the need for improved genomic resources and tools for more effectively characterizing IGH diversity. Here, we leveraged targeted long-read sequencing (Pacific Biosciences) of genomic DNA and fosmid clones to generate high quality haploid resolved assemblies spanning the IGH region for 7 individuals of different African, East Asian and European descent to effectively capture IGH diversity. From these assemblies, we identify novel SNVs and large SVs (n=12) ranging in size up to 100 Kb and include duplications, deletions and inversions. These SVs are resolved for the first time at nucleotide resolution, leading to the discovery of novel IG genes and alleles (n=70). In addition, we use these assemblies to benchmark our high-throughput targeted long-read sequencing protocol and bioinformatics pipeline, revealing that our novel approach can generate accurate assemblies and genotype callsets, outperforming alternative methods. These data provide a more complete understanding of IGH haplotype diversity and provide a set of vetted and curated reference assemblies for this locus. The availability of these haplotype resources will facilitate the design of novel high-throughput approaches for more accurately interrogating IGH polymorphisms, further improving the characterization of IGH gene diversity at the genome and population-level. This will be critical for delineating the roles of germline variation in Ab effector functions in disease and clinical phenotypes.

Molecular Effects of Genetic Variation Posters - Thursday

PB2531. Characterization of mutants in *SOS1* in an endothelial cell model of lymphatic anomaly

Authors:

L. Matsuoka¹, M. March², D. Li³, S. Sheppard⁴, R. Pellegrino⁵, J. Garifallou², C. Vaccaro⁶, C. Kao⁷, H. Hakonarson³; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Children s Hosp. of Philadelphia, Philadelphia, PA, ³Children's Hosp. of Philadelphia, Philadelphia, PA, ⁴Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., Rockville, MD, ⁵The Children s Hosp. of Philadelphia, Philadelphia, PA, ⁶The Children s Hosp. of Philadelphia, Philadelphia, PA, ⁷Ctr. for Applied Genomics, Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Mutations in *SOS1* are known causes of conditions that involve anomalies in the development of the lymphatic system, most notably Noonan syndrome. As part of our ongoing efforts to functionally characterize mutations that contribute to vascular and lymphatic anomalies, we have undertaken efforts to examine the effects of *SOS1* mutations in primary human dermal lymphatic endothelial cells (HDLECs). A known pathogenic mutation (E846K) and a variant of unknown significance were expressed in HDLECs via retroviral transduction, and the effects on intracellular signaling were examined via western blotting of relevant substrates. Additionally, the ability of the variant of unknown significance to activate Ras was determined through an *in vitro* Ras activation assay. The functional consequences of expression of *SOS1* mutants in HDLECs was determined through an established spheroid sprouting assay, previously used to describe the effects of *ARAF*, *KRAS*, and *NRAS* mutations. The role of *SOS1* as a guanine nucleotide exchange factor for RAS family GTPases places it in a position where it may regulate signaling through MAP kinase, PI-3-kinase, and mTOR pathways. In addition to the information we derived from the biochemical studies, inhibitors of each of the indicated pathways were tested for their ability to reverse the effects of *SOS1* mutants in the functional spheroid sprouting assay. The results of these experiments elucidate the role of *SOS1* specifically in lymphatic endothelial cells, inform as to the appropriate pathways to target therapeutically, and reveal the effects of a variant of unknown significance in *SOS1*.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2532. Characterizing a novel missense variant in the methionine salvage gene *MRI1* that conveys protection for type 2 diabetes.

Authors:

K. Bandesh, M. Traurig, L. Baier; Phoenix Epidemiology and Clinical Res. Branch, Natl. Inst. of Diabetes and Digestive and Kidney Diseases, NIH, Phoenix, AZ

Abstract Body:

Methionine is an essential amino acid which is strictly recycled in the body. Dietary methionine is needed for cellular processes including methylation and DNA repair, but it may also influence metabolic disease. Dietary restriction of methionine prevents hyperglycemia and improves hepatic insulin response in mice; in humans higher blood methionine levels associate with increased type 2 diabetes (T2D). In a prior genome-wide association study for T2D among 7,200 American Indians, an Arg149Cys (rs551664067) in the methionine salvage enzyme *MRI1* was the strongest missense association; individuals carrying the Cys allele had a lower T2D prevalence ($P=1.2 \times 10^{-6}$, Odds ratio= 0.43) and a later age of T2D onset ($P=1.6 \times 10^{-5}$; carriers 51.2 yrs, noncarriers 46.2 yrs). The Arg149Cys is very rare in all ethnic groups except Americans Indians (Cys allele frequency in TOPMED cohort= 0.00001 vs 0.02 in American Indians). *MRI1*, which has never been implicated in T2D, is expressed in all human tissues and recycles methionine from a primary byproduct by the universally conserved methionine salvage pathway. The Arg149Cys resides in a conserved protein domain necessary for *MRI1* salvage activity. We assessed the stabilities of the two *MRI1* proteins by predicting the change in free energy (ΔG); the variant Cys149 protein was less stable ($\Delta G=-0.74$ kcal/mol) with reduced solvent accessibility (69%, Cys149 vs 75%, Arg149). Using the available crystal structure of *MRI1* (PDB ID: 4LDQ) as a template, we generated the structure of Cys149 protein by homology modeling; superimposition of the two proteins revealed close structural similarity (RMSD= 1.77Å). However, the Cys149 protein revealed a smaller, more compact ligand-binding pocket than the Arg149 (38 and 76 residues respectively). We investigated the binding of both forms of *MRI1* to its ligand, Methylthioribose-1-P, by molecular docking and found that both forms of *MRI1* have catalytic activity. However, the Cys149 protein had an improved ligand binding ($\Delta G=-6.1$ kcal/mol, Cys149; -5.8 kcal/mol, Arg149) where the ligand uniquely formed a salt-bridge with residue Arg98 leading to a higher enzymatic activity. We are validating the observed ligand-binding energetics using ^1H -Nuclear Magnetic Resonance (NMR) spectroscopy; our preliminary results confirm that the Cys149 protein is an active enzyme. To verify differential ligand-binding, we are studying the catalytic turnover of the two proteins over a reaction time of 24-hours by comparing the peak intensities of the ligand and the product obtained in a real-time NMR setting. Our work provides insight into novel metabolic process that confer protection for T2D.

Molecular Effects of Genetic Variation Posters - Thursday

PB2533. Characterizing the effects of a novel *RDH12* donor splice site variant using *in vivo*, *in vitro*, and *in silico* methods

Authors:

M-Y. Chung^{1,2}, Y-C. Chen³, S-J. Chen³; ¹Dept. of Life Sci. and Inst. of Genome Sci., Natl. Yang Ming Chiao Tung Univ., Taipei, Taiwan, ²Dept. of Med. Res., Taipei Veterans Gen. Hosp., Taipei, Taiwan, ³Dept. of Ophthalmology, Taipei Veterans Gen. Hosp., Taipei, Taiwan

Abstract Body:

Retinol dehydrogenase 12 (*RDH12*), a member of the dual-function retinoid dehydrogenase/reductase family, catalyzes the conversion of 11- *cis*-retinal to vitamin A in the rod and cone photoreceptors. *RDH12* has been implicated in a number of autosomal recessive retinal dystrophies, including early onset Leber congenital amaurosis and adolescent-onset macular dystrophy. Objective: Whole exome sequencing revealed a proband affected with Leber congenital amaurosis were a compound heterozygote of a known nonsense mutation, c.187C>T (p.Arg62Ter) in exon 4, and a novel variant at donor splice site of intron 6, c.448+4_448+7del (clinical significance unknown and MAF=1.32x10⁻⁵, based on gnomAD v.3.1.2). We characterized this novel *RDH12* donor splice site variant using *in vivo*, *in vitro* and *in silico* methods. Methods: The effect of the variant was initially predicted by four bioinformatics tools. In addition, we analyzed transcripts isolated from peripheral blood leukocytes of the proband and parents by RT-PCR and sequencing as well as minigene assay to quantify the ratio between the wild-type and mutant transcripts resulting from this variant. Results: The results revealed that two out of the four bioinformatic tools score the variant below the threshold for a typical human donor splice site, and that, from both *in vivo* and *in vitro* studies, about 85% of the *RDH12* transcript carrying the variant had skipped the exon 6 and the translated product would be in frame deletion of 35 amino acids. Conclusion: Based on the nature and relative amounts of aberrant and wild-type transcripts, this variant is very likely pathogenic.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2534*. Chromatin conformation during CD4+ T cell activation implicates putative autoimmune disease candidate genes

Authors:

M. Pahl¹, P. Sharma¹, R. M. Thomas¹, Z. Thompson¹, J. Pippin¹, C. Su¹, S. F. A. Grant², A. D. Wells³; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Children's Hosp. of Philadelphia/Univ. of Pennsylvania, Philadelphia, PA, ³Children's Hosp. of Philadelphia and Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

The majority of genetic variants identified by genome-wide association studies (GWAS), including for autoimmune traits, are located in non-coding regions and contribute to disease risk by modulating cis-regulatory element (cRE) activity to influence gene expression. To determine how these variants contribute to disease risk, it is essential to define both the cREs where causal variants reside and their target effector genes. As cRE function is dynamic across tissue types and cellular states, characterizing the dynamic epigenetic status of cREs across several physiological processes should provide additional insight into disease-relevant variants. Here we investigated 822 signals from 15 autoimmune GWAS in the context of cREs active in naïve CD4+ T cells during TCR-CD28 costimulation. CD4+ T cells coordinate adaptive immune responses, and the mechanisms that contribute to gene regulation in this cell type are crucial for balancing effective immunity with the maintenance of self-tolerance. To characterize how dynamic changes in gene expression correlate with dynamic vs. stable cRE activity, we performed RNA-seq, ATAC-seq, and HiC to capture the gene expression, chromatin accessibility, and chromatin conformation across three phases (0hr, 8hr, 24hr) of naïve CD4+ T cells activation. We found enrichment of transcription factor motifs for *KLF2*, *FOXO3*, *JUN*, and *IKZF1* at cREs in contact with promoters that may influence gene expression to restrain vs. promote CD4+ T cell activation. We further integrated CD4+ T cell cREs with autoimmune GWAS signals and found that most of the autoimmune traits enriched in cREs connected to early differential genes, while SLE enriched for genes dynamic at later timepoints. We finally prioritized 423 signals and identified 1,836 (1,200 protein-coding) genes contacted with these signals with a median distance of 103kb. These putative disease-relevant genes are enriched for expression of quantitative trait loci (eQTLs) and genes previously implicated in CD4+ T cell proliferation and function by high-throughput screens and other studies, suggesting that autoimmune diseases may disproportionately influence processes involved in CD4+ T cell proliferation and activation. A subset of these genes can be prioritized based on potential pharmacological evidence, which in turn should prove useful for repurposing efforts.

Molecular Effects of Genetic Variation Posters - Thursday

PB2535. Clonal hematopoiesis of indeterminate potential and kidney function decline in the general population

Authors:

C. Vlasschaert¹, B. Kestenbaum², A. Bick³, P. Natarajan⁴, N. Franceschini⁵, A. Kottgen⁶, M. Lanktree⁷, M. Rauh¹, C. Robinson-Cohen³; ¹Queen's Univ., Kingston, ON, Canada, ²Univ. of Washington, Seattle, WA, ³Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁴Broad Inst., Cambridge, MA, ⁵Univ North Carolina at Chapel Hill, Chapel Hill, NC, ⁶Univ Hosp Freiburg, Freiburg, Germany, ⁷McMaster Univ., Hamilton, ON, Canada

Abstract Body:

Introduction: Clonal hematopoiesis of indeterminate potential (CHIP) is a common age-related process whereby somatic mutations in select driver genes lead to the biased production of leukocytes that exhibit dysregulated inflammation. CHIP affects at least 10% of individuals aged 65 and older and has been associated with increased mortality and multi-system morbidity, including the progression of chronic kidney disease. We tested the hypothesis that CHIP is associated with kidney function decline in the general population.

Study Design: Cohort Study

Setting and Participants: 11,941 individuals from three community-based cohorts in the TOPMed Consortium

Exposure: CHIP status from blood DNA-derived whole genome or exome sequences

Outcome: Incident 30% decline in estimated glomerular filtration rate (eGFR) after CHIP determination.

Analytical approach: Cox proportional hazards models for 30% decline endpoint. Study-specific estimates were combined using fixed-effects meta-analysis.

Results: Median baseline eGFR was 87 ml/min/1.73m² and median baseline age was 63 years. Prevalence of CHIP was 5.0%, 7.8% and 12.3% in persons 50-60, 60-70 and >70 years old, respectively. Over a median follow-up of 9 years, 2020 kidney function decline events occurred, with an event rate of 2.0 per 100-person-years among CHIP carriers and 1.8 per 100-person-years among those without CHIP. In meta-analysis adjusting for age, age², sex, baseline eGFR, self-reported race and diabetes status, CHIP was associated with a 22% higher risk of kidney function decline (95% CI: 5% to 41% higher; p-value=0.008).

Conclusions: We detected an association between CHIP and kidney function decline in three general population cohorts. Further studies are needed to investigate this novel condition.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2536. Clustering of missense mutations follows a non-random distribution across the human genome and correlates with specific classes of genes

Authors:

C. Rivolta^{1,2,3}, M. Quinodoz^{1,2,3}, V. G. Peter^{1,2,3,4}, K. Cisarova⁵, B. Royer-Bertrand⁵, P. Stenson⁶, D. Cooper⁶, S. Unger⁵, A. Superti-Furga⁵; ¹Inst. of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland, ²Univ. of Basel, Basel, Switzerland, ³Univ. of Leicester, Leicester, United Kingdom, ⁴Lausanne Univ. Hosp. (CHUV), Lausanne, Switzerland, ⁵Univ. of Lausanne and Lausanne Univ. Hosp. (CHUV), Lausanne, Switzerland, ⁶Cardiff Univ., Cardiff, United Kingdom

Abstract Body:

Using a machine learning approach, we analyzed the distribution of missense variants observed in genetic conditions and cancer, and discovered that within-gene clustering of mutations is widespread across the human genome, significantly deviating from a random distribution in as many as 46% of all well-characterized disease genes from ClinVar. We took advantage from these findings to develop a pathogenicity predictor, MutScore, which integrates qualitative features of DNA substitutions with this new additional information on clustering. In its first four months of activity, MutScore's website received more than 2,000 queries from the scientific community. This prompted us to assess its performance on additional, more recent ClinVar submissions, as well as to expand our research on variant clustering. With a global area under the curve (AUC) value of more than 0.95 for sensitivity vs. specificity, MutScore showed an even better predictive power on these novel variants than previously reported.

In addition, we investigated further the biological relevance of degrees of mutational clusters, by assessing whether the within-gene distribution of pathogenic variants would correlate with specific classes of genes. We found that genes with high clustering of mutations within small regions were significantly more likely to encode for proteins (1) having DNA-binding properties, (2) being transcription factors, and (3) localizing in the nucleus, whereas enzymes were significantly underrepresented in this category. Conversely, proteins encoded by genes with no relevant mutational clustering were enriched in enzymes and depleted in transcription factors, at a statistically significant level. These findings very likely reflect the molecular mechanisms shaping pathogenicity at the protein level for both genetic diseases and cancer. For instance, DNA-binding domains that characterize transcription factors and other regulatory proteins are in general relatively small (20-100 aa residues), and therefore pathogenic variants would tend to be restricted to such domains, leading in turn to dense mutational clusters at the DNA level. Conversely, functional domains of enzymes are larger, giving rise to stretches of medium clustering.

In conclusion, our analyses confirm MutScore's high predictive power and identify a new and significant link between degrees of mutational clustering and protein function, across the whole human genome.

Molecular Effects of Genetic Variation Posters - Thursday

PB2537. COL9A1 and FLNB: Variants corresponding to complex clinical phenotype.

Authors:

J. Vázquez¹, M. Santos², C. Velez³, G. Serrano⁴, S. Carlo⁵, F. Vélez-Bartolomei⁶, A. Cornier⁷; ¹Genetic Diagnostic Group, Bayamón, PR, ²Univ. of Puerto Rico Med. Sci. Campus, Sch. of Publ. Hlth., Calle Antosnsanti #1607 San Juan, Puerto Rico, ³PUCPR, Ponce, PR, Puerto Rico, ⁴Genetic Diagnostic Group/Ponce Hlth.Sci. Univ., San Juan, PR, ⁵Ponce Hlth.Sci. Univ., Cabo Rojo, PR, ⁶San Jorge Children's Hosp./Genetic Diagnostic Group, San Juan, PR, ⁷San Jorge Children s Hosp./ Ponce Hlth.Sci. Univ., San Juan, PR

Abstract Body:

We are presenting the case of an 18-year-old male that was referred to genetics by cardiology due to skin striae and to rule out a connective tissue disorder (CTD). During the physical examination there was evidence of pronounced joint hyperlaxity, striae on the back and on the lower extremities, arachnodactyly and chest asymmetry. The patient also presented with scoliosis that was documented by spine X-rays and mitral valve prolapse that was documented by an echocardiogram. Due to clinical symptoms such as tenderness in the lower extremities, recurrent headaches and evidence of Marfanoid habitus a comprehensive connective tissue disorder gene panel was requested. The results of the CTD gene panel indicated that the patient is heterozygous for variants in two genes: *COL9A1* and *FLNB*. The *COL9A1* gene codes for a specific part of the type IX collagen protein. A sequence variant *c.1636G>A* was recorded in this patient. This nucleotide change is predicted to cause an amino acid substitution designated as *p.Asp546Asn*. Pathogenic variants of the *COL9A1* gene has been associated with autosomal recessive Stickler syndrome IV and autosomal dominant multiple epiphyseal dysplasia 6. The *FLNB* gene is responsible for providing the blueprint to produce a protein called filamin B. This protein is important because it provides a scaffold that binds to actin in order to form a network that is known as the cytoskeleton of the cell. Filamin B is of major importance because it plays a role in skeleton development before birth and it is a protein expressed by chondrocytes. A sequence variant *c.4447T>C* in the *FLNB* gene was documented in this patient. This change is predicted to result in an amino acid substitution *p.Phe1483Leu*. Pathogenic variants of *FLNB* are associated with autosomal recessive spondylocarpotarsal synostosis syndrome, autosomal dominant ateloteggenesis type 1 and type 3, Boomerang dysplasia and Larsen syndrome. The *COL9A1* and the *FLNB* variants have never been reported in the scientific literature. We are presenting a case of a patient with a genetic variant in the *COL9A1* gene that has not been associated before to Stickler syndrome type 4 and may be responsible of the patient's phenotype. *FLNB* variant may be responsible for some of this patient's phenotype, resembling an atypical presentation of Larsen syndrome which has a wide range of phenotypic variability. Both genes have a dominant effect on phenotype and these variants have not been reported in the literature before. Functional proteomic studies would be useful in order to define the degree of involvement of these proteins, and this would further characterize the molecular pathophysiology of these genetic gene variants.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2538. Colocalization of blood cell traits GWAS associations and variation in *PU.1* genomic occupancy prioritizes causal noncoding regulatory variants

Authors:

R. Jeong^{1,2}, M. L. Bulyk^{1,2}; ¹Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, ²Harvard Univ., Cambridge, MA

Abstract Body:

Although genome-wide association studies (GWAS) uncovered numerous trait-associated loci across the human genome, most are located in noncoding regions and are often hard to fine-map to individual causal variants because of widespread linkage disequilibrium (LD). To address this challenge, we present a strategy utilizing transcription factor (TF) binding quantitative trait loci (bQTLs) for colocalization analysis to identify genetic associations potentially mediated by TF occupancy variation and to pinpoint likely causal variants. To demonstrate, we sought to identify blood cell trait GWAS loci that may be mediated by *PU.1*, a hematopoietic master regulator. We analyzed available *PU.1* ChIP-seq data in 49 genotyped lymphoblastoid cell lines (LCLs), detected genetic variants associated with *PU.1* occupancy levels, and tested colocalization with overlapping blood cell trait GWAS signals. 69 significantly colocalized blood cell trait associated loci are potentially driven by *PU.1* occupancy variation. We identified *PU.1* motif-altering variants at 55 of them, which we nominate as the likely shared causal variants for the colocalized blood cell trait associations. We also demonstrate the regulatory impact of two such variants, using available reporter assay results. The ability to hone in on variants altering a particular TF motif facilitated pinpointing a likely causal regulatory variant at a locus, as well as highlighting the likely direct regulatory impact of the variant. This work motivates the use of TF bQTL data to elucidate GWAS associations to various complex traits, where relevant TFs are known.

Molecular Effects of Genetic Variation Posters - Thursday

PB2539. Combining human genetics and functional genomics identifies regulators of the lysosomal lipid BMP.

Authors:

D. Leto, S. V. Andrews, S. S. Davis, A. Rana, R. Ghosh, T. Logan, R. D. Maciuca, J. Suh, T. Sandmann, G. Di Paolo, A. G. Henry; Denali Therapeutics, South San Francisco, CA

Abstract Body:

Background: The phospholipid bis(monoacylglycerol)phosphate (BMP) is highly enriched in the endo-lysosomal system, where it facilitates cholesterol egress and sphingolipid degradation. BMP dysregulation is observed in many lysosomal storage disorders, although it is unclear whether this lipid is a driver or marker of disease. Elevated levels of BMP are also observed in urine from carriers of the Parkinson's Disease (PD) risk variant G2019S in LRRK2, a protein implicated in the regulation of lysosomal function. Early-stage clinical data has shown that LRRK2 inhibition reduces urine BMP in a dose-dependent manner. Despite its function in maintaining lysosomal health and its connection to disease, the genes that control BMP metabolism remain poorly understood.

Methods: We used two complementary approaches to identify genetic regulators of BMP: (1) we performed GWAS of four BMP species using WGS and urine BMP measurements from the Parkinson's Progression Markers Initiative (PPMI) and the Accelerating Medicines Partnerships - Parkinson's Disease (AMP-PD) (N = 634 samples). We nominated causal genes via eQTL and chromatin interaction data and examined overlap with regions identified in PD GWAS of European samples. (2) We conducted a targeted CRISPR-knockout (KO) screen in mouse bone marrow-derived macrophages (BMDMs). We assessed the effects of deleting 129 genes, nominated either as causal genes in the BMP GWAS or selected through literature curation, on the intracellular levels of five different BMP species via LC/MS.

Results: In the BMP GWAS study, the LRRK2 G2019S variant was a genome-wide significant ($p < 5E-8$) quantitative trait locus for all four BMP species. Gene set enrichment analysis of putative causal genes in loci reaching a suggestive association threshold ($p < 1E-5$) with BMP levels implicated endocytosis and phospholipid transport pathways. Two loci overlapped with regions identified in PD GWAS, including the ANK2/CAMK2D/LARP7 region on chromosome 4. The BMP CRISPR-KO screen identified gene knockouts that affected the total abundance or composition of the intracellular BMP pool ($Z\text{-score} \geq 2$). In addition to validating known modifiers of intracellular BMP, the screen also identified genes that have not previously been implicated in BMP regulation, including causal genes nominated in the BMP GWAS.

Conclusion: Combining human genetics and functional genomics approaches is an effective method for identifying and prioritizing putative regulators of clinical biomarkers.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2540. Comparative titin gene frequencies and distributions across four population cohorts.

Authors:

P-S. Lai, R. Lim, J. Koo, J. Rao, G. Tan; Natl. Univ. of Singapore, Singapore, Singapore, Singapore

Abstract Body:

Mutations in the titin gene (*TTN*) are known to cause a wide spectrum of genetic diseases, ranging from cardiomyopathies to skeletal muscle diseases. *TTN* gene, spanning 363 exons, encodes for titin, the largest human protein which spans four domains, namely I-, A- and M-bands, and Z-disk. The increasing use of next-gen sequencing (NGS) has led to a surge in discovery of numerous *TTN* variants, many of which are rare and unique. In this study, we investigated the frequency and spectrum of *TTN* variants in a cohort of 60 local neuromuscular (NMD) patients and compared these against variants in public domains from three other populations, namely cardiomyopathies, titinopathies and healthy individuals. Variants in the NMD cohort were identified from standard NGS pipelines using whole genomes sequenced at an average 30X coverage. Pathogenicity classification was performed according to the American College of Medical Genetics and Genomics (ACMG) guidelines. A total of 15,757 *TTN* variants was observed from the 60 NMD cases, with more than half being novel or rare in the general population. Most of the variants (90.60%) were non-synonymous, of which 1.05% of these were truncating (TTNtv). Missense variants make up the majority of the exonic variants (59.39%) followed by synonymous variants (32.74%) and non-frameshift indels (4.57%). Rare or novel variants accounted for 698 out of 934 unique variants in coding regions. Among the 38 *TTN* variants curated as clinically pathogenic in the 60 NMD cases, 25 were novel. In terms of frequencies and distributions among the titin domains, titin I-band and M-band had the highest TTNtv densities at 0.525 and 0.547 variants/kb. In contrast, TTNtv variants were enriched in the A and M-bands more than the I-band or Z-disk for healthy, titinopathies and cardiomyopathies cohorts. Overall, all the disease cohorts had significantly higher TTNtv densities compared to the general population ($p = 0.0478$). These results highlight the challenge in interpreting the clinical significance of *TTN* variants especially with the frequent presence of rare and unique variants across all populations. The high percentage of novel and rare TTNtv variants identified in this study also indicate the need for further understanding on the distribution of these variants to help in the identification of pathogenic variants.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2541. Comparison of pedigree-based graph workflow to more traditional workflows for rare candidate variant analysis

Authors:

B. Pusey¹, J. Fu¹, W. N. Gahl², D. R. Adams³; ¹Natl Human Genome Res. Inst., Bethesda, MD, ²NHGRI (NIH), Bethesda, MD, ³NIH, Bethesda, MD

Abstract Body:

The NIH Undiagnosed Diseases Program (UDP), a subgroup of the Undiagnosed Diseases Network (UDN), was established with the goal of conducting research for patients with significant illnesses that remain undiagnosed despite extensive medical evaluation. Genome sequencing approaches became available around the same time that the UDP started and are important tools for making diagnoses in some cases. Given the increasing prevalence of exome and genome sequencing in standard medical care, many current UDP applicants have been evaluated with this technology before they apply. As a result, a major focus of UDP genomic analyses is the assessment and prioritization of variants of unknown significance, often in genes not currently associated with human disease. This work often requires an “agnostic” approach, maximizing the sensitivity and scope of variant analysis. A recently described FASTQ to filtered-variant alignment pipeline (VG-Pipeline, Markello et al, 2022, PMID 35483961) was developed in part with UDP cases. We present a descriptive head-to-head comparison of the VG-Pedigree analysis pipeline with our prior GATK-derived pipelines. The comparison cohort contains approximately 200 nuclear families with a wide variety of clinical presentations. We conclude by highlighting the strengths and weakness of each approach along with considerations for future work.

Molecular Effects of Genetic Variation Posters - Thursday

PB2542. Comprehensive detection of trans-regulatory signal from scRNA-seq and perturbation datasets

Authors:

N. Babushkin, X. Liu; Univ. of Chicago, Chicago, IL

Abstract Body:

Single cell RNA-sequencing (scRNA-seq) coupled with effective CRISPR/Cas9-mediated perturbations offers the ability to functionally map regulatory elements of gene regulation, including trans-gene regulation. Recently, several studies (Xie 2019 Cell Rep; Gasperini 2019 Cell and Morris 2021 bioRxiv) generated large scale scRNA-seq perturbation datasets to study cis-regulatory effects. However, most focused on detecting cis-regulatory effects, probably because of the computational and statistical challenges of detecting trans-signals, including low statistical power and high false positive rates. Short RNA-sequencing reads mapped to multiple homologous locations on the genome lead to more than 75% of the false positive trans-signals in bulk RNA-seq data. Similarly, we found that around 70% of trans-signals in scRNA-seq perturbation data are likely false positives due to an analogous multi-mapping issue. However, using the solution for bulk RNA-seq, which involves removing suspicious trans-signals that are cross-mappable, is overly stringent in scRNA-seq data and results in substantial loss of true signal. In this study, we carefully addressed the multi-mapping issue in order to reduce false trans-signals while preserving power in scRNA-seq perturbation data. More specifically, we thoroughly removed scRNA-seq reads mapped to genomic regions of low mappability, such as repeat regions, before quantifying gene expression in each cell. Using this approach, we processed a collection of scRNA-seq and perturbation datasets, targeting more than 6,500 regulatory elements and profiling expression in more than 370,000 cells. We then applied the state-of-the-art association method SCEPTRE (Barry 2021 Genome Biology) to the datasets to detect trans-signals. Our work will produce an unprecedented, comprehensive and high-quality set of trans-regulatory signals from scRNA-seq and perturbation datasets.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2543. Comprehensive signal identification highlights the contribution of protein QTLs to complex trait genetics

Authors:

J. Chiou, H. Kim, UK Biobank Pharma Proteomics Project, FinnGen, E. Fauman, M. Miller; Pfizer, Cambridge, MA

Abstract Body:

Large-scale protein quantitative trait loci (pQTL) mapping presents opportunities for unraveling biological mechanisms underlying the genetic risk of complex diseases. However, comprehensive signal identification, especially at the oft-ignored major histocompatibility complex (MHC) and X chromosome, is necessary to leverage the full potential of pQTL datasets. In this study, we identified 28,867 conditionally independent pQTL signals of 1,424 plasma proteins measured across 35,306 individuals from the UK Biobank Pharma Proteomics Project using SuSiE regression. Cis regions were more likely to contain multiple independent signals (cis: 88.1%, trans: 17.5% of regions), contained more signals on average (cis: 5.7, trans: 1.3 signals), and had better resolved fine-mapping (cis: 9.5, trans: 43.1 credible set variants). In the extended MHC locus, we observed 1,265 independent signals, 78.3% of which mapped outside of the MHC class I/II genes. Joint tagging between two or more causal variants boosted the significance of non-causal variants in 3.3% of regions, where the index variant from the marginal association was not identified in any of the conditional credible sets. Strong primary signals masked the effects of 5.6% of independent signals, which had attenuated significance in the marginal association. We highlight the contribution of independent pQTL signals to complex trait genetics through specific examples on the X chromosome, including (1) a causal relationship between increased circulating CD40 ligand levels and risk for autoimmune diseases using Mendelian randomization (OR=1.05, 95% CI=1.01-1.08) and (2) colocalization between a single (out of 6 distinct signals) non-primary cis pQTL signal that reduces ACE2 levels and a rare variant (rs190509934) that decreases COVID-19 susceptibility. We are currently broadening the scope of our independent pQTL signal to GWAS integration and expect to report high-level findings from these analyses. These results underscore the importance of modeling all variants in associated regions for accurate signal identification and demonstrate how pQTLs can support therapeutic hypotheses of disease risk mechanisms in understudied genomic regions.

Molecular Effects of Genetic Variation Posters - Thursday

PB2544*. Conditionally distinct adipose eQTL signals in 2,256 individuals identify hundreds of colocalized genes for cardiometabolic traits

Authors:

S. Brotman¹, L. Guan², J. El-Sayed Moustafa³, A. Jackson⁴, K. Broadaway¹, D. Wang³, K. Currin¹, A. Roberts³, C. Raulerson¹, M. Erdos⁵, N. Narisu⁵, H. Stringham⁴, M. Boehnke⁴, F. Collins⁵, P. Pajukanta⁶, M. Laakso⁷, K. Mohlke¹, K. Small³, L. Scott⁴; ¹Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, ²Dept. of Computational Med. & Bioinformatics, Univ. of Michigan, Ann Arbor, MI, ³Dept. of Twin Res. and Genetic Epidemiology, King's Coll. London, London, United Kingdom, ⁴Dept. of Biostatistics and Ctr. for Statistical Genetics, Sch. of Publ. Hlth., Ann Arbor, MI, ⁵Natl. Human Genome Res. Inst., NIH, Bethesda, MD, ⁶Dept. of Human Genetics and Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ⁷Inst. of Clinical Med., Kuopio Univ. Hosp., Univ. of Eastern Finland, Kuopio, Finland

Abstract Body:

GWAS have implicated thousands of loci in disease, yet most underlying genes remain unknown. To detect genetic variants associated with both disease risk and gene expression level, we colocalized GWAS signals with expression quantitative trait loci (eQTL) signals in subcutaneous adipose tissue from 2,256 individuals in 5 studies. While most previous eQTL studies focused on primary eQTL signals, we used conditional analysis to detect non-primary signals. Of 28,346 genes tested in ≥ 2 studies, 17,969 (63%) had ≥ 1 significant eQTL ($P \leq 1e-06$); fewer genes (8,113-11,254) were detected in individual studies. Of the 17,969 genes, 50% had ≥ 2 eQTL signals and 21% had ≥ 3 signals. Among genes with ≥ 4 eQTL signals, non-primary eQTL had lower median effect sizes: 0.38, 0.27, 0.21, and 0.19 for 1st, 2nd, 3rd and 4th-10th signals, respectively. Additionally, the distances of lead eQTL variants from their transcription start sites were a median of 31, 39, 48 and 68 kb for 1st, 2nd, 3rd and 4th-10th signals, respectively. Adipose eQTL variants for the 1st-4th signals were enriched in both promoters and enhancers ($P < 2.5e-3$), however the levels of enrichment (particularly for promoters) were successively lower with increasing order of signal discovery. Gene expression levels in TwinsUK showed a median heritability estimate of 0.19 for eGenes and 0.07 for genes without an eQTL; heritability increased with the number of eQTL signals per gene, independent of gene expression level. Preliminary analyses using proportions of 4 cell types in TwinsUK showed 0.42% of 1st signals had an interaction with adipocyte proportion, compared to 0.35, 0.27 and 0.33% of 2nd, 3rd, and 4th signals, respectively. We compared the adipose eQTL to a 10-fold larger study in blood ($n = 31,684$): among 24,666 significant adipose gene-signal pairs tested in eQTLGen, only 15,622 (63%) were had a $P \leq 1e-06$, highlighting the value of disease-relevant tissues. Finally, we used *coloc* ($PP4 \geq 0.5$) and LD ($r^2 \geq 0.5$) to identify adipose eQTL colocalizing with GWAS loci from 3 cardiometabolic traits. We identified 158 colocalized signals for T2D, 186 for WHR, and 237 for WHRadjBMI, corresponding to 150, 176, and 224 genes, respectively. Of 581 total colocalized GWAS-eQTL signals, 31% were from non-primary eQTL and would have been missed if distinct signals were not identified. Colocalizations include allelic series where ≥ 2 GWAS signals colocalized with ≥ 2 eQTL signals for the same gene, further supporting the role of these genes in disease. Together, we found that half of genes with an eQTL have at least 2 signals and that GWAS colocalization with non-primary signals greatly expanded the number of candidate genes for common traits.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2545. Congenital Myasthenic Syndrome Due to *DOCK7* Gene Mutation Case Study Report

Authors:

A. Lebron Ilarraza^{1,2}, G. Serrano Rodriguez³, J. Pascual³, G. Arroyo Figueroa^{1,2}, S. Carlo^{3,1}, F. Velez¹, N. Arciniegas³, A. Cornier⁴; ¹San Jorge Children's Hosp., San Juan, PR, ²Univ. of Puerto Rico Piedras Campus, San Juan, PR, ³Ponce Hlth.Sci. Univ., Ponce, PR, ⁴San Jorge Children s Hosp./ Ponce Hlth.Sci. Univ., San Juan, PR

Abstract Body:

Congenital myasthenic syndrome (CMS) are rare genetic conditions resulting from a defect at the junction where your nerve stimulates muscle activity. To date, few cases of CMS due to the *DOCK7* gene mutation have been documented, but it must remain within the differential diagnosis for patients with muscular weakness and other neurological problems. We found a missense change in Exon 14 that replaced valine with alanine at codon 539 of the *DOCK7* protein, which has not been reported and may be deleterious. A 20-year-old male presents himself with a history of generalized muscular weakness, asthma, fatigue, ptosis, intestinal dysmotility, dysphagia (upper dysmotility), and other conditions like non-epileptic events, intense headaches, and multiple osseous fractures. The intriguing facts about this case are that, aside from the low frequency, the manifestation of the symptoms can be non-specific, and it may imitate other pathologies. This case could represent a novel way of identifying individuals with congenital myasthenia that could have passed unnoticed or misdiagnosed due to *DOCK7* gene mutations.

Molecular Effects of Genetic Variation Posters - Thursday

PB2546. Connecting rare variation to extremes of plasma protein levels

Authors:

X. Xie¹, T. Li², C. Benner¹, J. Irudayanathan¹, B. van de Geijn¹, R. Pendergrass¹, A. Mahajan¹, A. Battle³, M. McCarthy¹; ¹Genentech, South San Francisco, CA, ²Johns Hopkins Sch. of Med., Baltimore, MD, ³Johns Hopkins Univ., Baltimore, MD

Abstract Body:

The emergence of large scale high-throughput proteomic data enables integrative analysis of the relationships between genetic variation and plasma proteins: these can provide insights into regulatory mechanisms and biomarkers relevant to disease. While genome-wide association studies (GWAS) on plasma proteomics have identified many common variants associated with plasma protein levels, rare variants, which are collectively abundant in the human genome, are inefficiently tagged in GWAS. Here, we focus on characterizing the relationship between rare variants and extreme plasma protein levels. We used measures of 1463 plasma proteins generated, using the OLINK platform, on 54306 UK Biobank participants (another 1500 proteins will be available shortly). In the pilot analyses with 16991 baseline UKB participants, we assessed the prevalence of rare coding variants (MAF<0.1%) amongst protein level outliers (defined as individuals with expression values >3SD from the average expression level of the cognate protein). We adjusted the protein levels by known covariates, such as sex, age, and study sites, plus 100 PEER factors to account for hidden confounders. Amongst individuals who were outliers for a given protein, we found a marked excess of rare coding variants in the cognate gene, consistent with our hypothesis that rare coding variants underlie extremes of protein abundance. We found that, compared to non-outliers, outliers are 2.31 times more likely to be rare variant carriers ($p=8.51e-88$): this enrichment increased to 2.49 ($p=1.08e-42$) and 2.52 ($p=1.90e-21$) when more stringent thresholds (>4SD and >5SD) were used to define outliers. The magnitude of this enrichment signal decreased as the number of PEER factors was reduced (from OR=2.31 with 100, to 2.09 with 50 and 1.83), confirming PEER factors' effectiveness in capturing the hidden confounding variables in measured proteomics. Other mechanisms that may lead to extreme protein levels, including artefacts related to differential antibody binding, will be assessed in further analysis. Access to whole-genome sequence data from UKB will allow us to extend these analyses to include rare variants in non-coding space. We are developing a hierarchical Bayesian model, Watershed, to synthesize genomic annotations and observed outlier signals across proteins in trans to prioritize rare variants and interrogate their effects on disease risk given UKB's extensive phenotype data. The ultimate goal is to establish the value of plasma protein profiling as a valuable tool to sequence-based diagnostics, which can improve interpretation of rare variants and identification of causal genes for complex traits.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2547*. Consequences of loss of function burden effects on the human plasma proteome in 54,306 UK Biobank participants

Authors:

C. Whelan¹, S. Loomis², N. Okugawa¹, H. McLaughlin¹, T. Trinh¹, H. Runz³, S. Engle¹, A. Chen¹, B. Sun²; ¹Biogen Inc., Boston, MA, ²Biogen, Cambridge, MA, ³Biogen Inc., Cambridge, MA

Abstract Body:

Proteins are the functional modules underlying biological processes and the main targets of drugs. To date, studies of genetic determinants of plasma protein levels have primarily focused on single variant associations, most of which are non-coding. Here, we directly assessed the effects of gene-based protein truncating variant (PTV) burden (19,209 genes; PTV MAF <1%) on 1,463 plasma proteins measured in 54,306 individuals in the UK Biobank using the Olink Explore assay. We identified 1,642 significant PTV associations with circulating proteins ($p < 2.6 \times 10^{-6}$), including 584 genes associated with the encoded protein (*cis*) and 1,058 associations with genes other than for the encoded proteins (*trans*). For 38% (223) of *cis* gene PTV-burden associations, we did not find any associations within the corresponding gene in genome-wide single variant pQTL analyses, demonstrating that PTV burden can help refine candidate genes at single-variant pQTL loci. 99% of *cis* PTV effects led to decreased protein levels (mean $\text{abs}(\beta) = 1.67$); *trans* PTV effects were on average weaker (mean $\text{abs}(\beta) = 0.78$, $p = 6.6 \times 10^{-153}$) and more evenly distributed in direction (40% decreased, 60% increased protein levels). *Trans* effects highlighted pleiotropic loss of function effects at genes that may act as master regulators of protein pathways and networks. For example, PTV burden at *TET2* was associated with perturbations of several immune pathway proteins, including decreased concentrations of CCL5, CLC, FLT3LG, KIR2DL3, KIR3DL1 and LCN2, and increased levels of CD1C, CD200R1, CD207, CD33, CLEC4C, FCGR2A, FLT3, LILRB2 and TNFSF13B. We also found *IGF2R* PTV burden leads to decreased IGF2R levels and downstream increases in multiple cathepsin levels, reflecting the role of IGF2R in lysosomal transport of cathepsins. Additionally, we illustrated how PTV gene burden can mimic effects of inhibitory drugs, in cases such as PCSK9 and ANGPTL3. Finally, we experimentally validated two putative loss-of-function variants at *APP*, revealing reduced plasma protein and corresponding, genotype-dependent reductions in iPSC mRNA and protein levels for two stop gained variants at exon 7. Our study highlights how PTV burden can be linked to protein levels to help implicate specific genes in driving proteome expression, highlight associations and networks missed at the single-variant level, and directly elucidate physiological implications of inhibitory drugs on disease.

Molecular Effects of Genetic Variation Posters - Thursday

PB2548. CRISPR perturbation screens identify predicted regulatory sequences and genes within GWAS loci associated with erythrocyte density traits

Authors:

N. Brosseau¹, T. Pincez², K. Sin Lo¹, M. Beaudoin³, G. Lettre⁴; ¹Montreal Heart Institute, Montréal, QC, Canada, ²Université de Montréal, Montréal, QC, Canada, ³Montreal Heart Inst, Montreal, QC, Canada, ⁴Université de Montréal, Montreal, QC, Canada

Abstract Body:

GWAS have identified hundreds of loci robustly associated with red blood cell (RBC) traits such as mean corpuscular volume (MCV) and mean corpuscular hemoglobin concentration (MCHC). These traits influence complications in several hematological diseases such as sickle cell disease and malaria, and a better understanding of the genetic regulation of these traits could provide new drug targets for these pathologies. Most of the sentinel SNPs identified are in non-coding regions of the genome. This suggests these genomic regions have a regulatory effect on RBC phenotypes. In this study, we created a library of sgRNA targeting the sentinel SNPs identified by GWAS and their strong linkage disequilibrium variant proxies, as well as potential regulated genes based on proximity and other genes of interest. We used an immortalized CD34⁺ erythroid-like cell line (HUDEP-2) to assess the effect of the sgRNAs on cell density. Three different cell lines stably expressing either Cas9 (CRISPR KO), dCas9-KRAB (CRISPRi), and dCas9-VP64 (CRISPRa) were transduced with our sgRNA library. Cells were then separated according to their density using a Percoll gradient as density combines the effect of changes of MCV and MCHC. Cells were collected at each fraction of the gradient, and the abundance of each sgRNA was assessed using Illumina next-generation DNA sequencing. The expected results are that the sgRNAs that increase the density of the cells will be more abundant in the lower fractions, and sgRNA decreasing the density of the cell will be less abundant in the lower fractions compared to safe sgRNAs. Using this method, we were able to identify (at a false discovery rate <10%) 21 genomic targets with CRISPR KO, 22 with CRISPRi, and about 10 times more with CRISPRa. After validation and further characterization, these findings should provide a better understanding of the genetic regulation of erythrocyte volume.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2549. CRISPR/Cas9-mediated knockout and cDNA rescue using sgRNAs that target exon-intron junctions in *Drosophila* enable functional analysis of pathogenic variants in a tissue-specific manner.

Authors:

W. Yoon, M. Chilian, K. Vargas Parra, A. Sandoval, J. Ramirez; Oklahoma Med. Res. Fndn., Oklahoma City, OK

Abstract Body:

Functional studies of genetic variants in undiagnosed patients in humans require model organisms such as *Drosophila*. Studies of disease-associated variants in *Drosophila* often need to be performed in a tissue-specific manner, especially in cases where human genes have multiple paralogs, one of which is expressed in a specific tissue, while *Drosophila* single orthologs are expressed ubiquitously. OxoGlutrate DeHydrogenase-Like (OGDHL) is one such gene and is mainly expressed in the human brain. In contrast, dOgdh, the *Drosophila* single ortholog, is expressed ubiquitously. In this study, we take CRISPR/Cas9 and Gal4/UAS approaches to achieve tissue-specific knockout in parallel with the rescue of the knockout by cDNA expression in *Drosophila*. We demonstrated that neuronal-specific expression of Cas9 with the sgRNA that targets the exon-intron junction of dOgdh (sgRNA_{dOgdh}) leads to knockout of the endogenous dOgdh gene, but not UAS-dOgdh cDNA transgene. This approach enabled us to determine the pathogenicity of nine missense variants in OGDHL identified from individuals with neurological phenotypes. We anticipate that this new approach will be widely used to study the tissue-specific roles of numerous human disease-associated variants as well as gene function studies in *Drosophila*.

Molecular Effects of Genetic Variation Posters - Thursday

PB2550. Cryptic AS-NMD elements harbor relevant variants in probands with early onset genetic disease.

Authors:

S. Felker¹, J. M. J. Lawlor¹, D. R. Latner¹, M. L. Thompson¹, S. M. Hiatt¹, C. R. Finnila¹, K. M. Bowling², Z. T. Bonnstetter¹, W. V. Kelley¹, A. C. Hurst³, M. A. Kelly⁴, G. Nakouzi⁴, G. M. Cooper¹; ¹HudsonAlpha Inst. For Biotechnology, Huntsville, AL, ²Washington Univ. Sch. of Med., Saint Louis, MO, ³Univ. of Alabama in Birmingham, Birmingham, AL, ⁴HudsonAlpha Clinical Services Lab., Huntsville, AL

Abstract Body:

Families of children with early-onset genetic disease often undertake arduous diagnostic odysseys that can take a considerable emotional and financial toll. Despite advancements in genomic medicine and research, rates of causal variant discovery in these probands remain low. One contributor to this deficiency is that variants in poorly annotated functional genomic elements are often overlooked in conventional analysis. Examples of this are “poison exons” which are non-coding, noncanonical exons in transcripts that are alternatively spliced and result in nonsense-mediated decay (AS-NMD) via the inclusion of a premature termination codon. Such AS-NMD transcripts are difficult to capture and annotate because they may only be expressed at an early developmental or embryonic timepoint. As variants within these cryptic regions have been previously associated with disease, we conducted a genome-wide screen for variants in AS-NMD elements in probands without a molecular diagnosis. Variants from genomes and exomes of 4001 probands with early-onset genetic disease were intersected with elements curated from published studies of differential AS-NMD. The allele frequency, conservation, and CADD score of resulting variants were examined with regard to patient phenotype and mode of inheritance. Variant pathogenicity was determined using ACMG/AMP Standards and Guidelines. This analysis has resulted in likely pathogenic and VUS variants in probands with no previously returned results from conventional analysis. Two novel VUS variants were identified in highly conserved regions of intron 20 of *SCN1A*. This intron contains the well-described 20N poison exon within which several variants associated with Dravet Syndrome and Dravet-like epilepsy phenotype have been found. Another variant was found within the poison exon of *SNRPB*, a gene associated with cerebro-costo-mandibular syndrome (CCMS) which causes abnormal skeletal morphologies and cardiac defects consistent with the proband’s phenotype. This variant had been determined causative for CCMS in earlier publications from other cohorts, so there was sufficient evidence to return as likely pathogenic. This AS-NMD directed variant investigation has resulted in returnable variants in probands who had no molecular diagnosis. The inclusion of cryptic AS-NMD elements in variant analysis pipelines would result in few additional variants per proband, with minimal, if any, added labor from scientists. In conclusion, this investigation has shown that incorporation of AS-NMD elements in genomic analysis would improve diagnostic yield in rare disease.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2551. CYP2C19 Polymorphism in Ischemic Heart Disease Patients Taking Clopidogrel after Percutaneous Coronary Intervention in Egypt

Authors:

M. El-Zawahri; MUST, Giza, Egypt

Abstract Body:

CYP2C19 Polymorphism in Ischemic Heart Disease Patients Taking Clopidogrel after Percutaneous Coronary Intervention in Egypt El-Zawahry M¹ Shawky A² Sabit H^{3†} Baraka K^{4†}

¹Department of Medical Biotechnology, College of Biotechnology, Misr University for Science and Technology, ²Department of Cardiology, College of Medicine, Helwan University, Cairo, Egypt³Department of Environmental Biotechnology, College of Biotechnology, Misr University for Science and Technology, P. O. Box 77, Giza, Egypt⁴Department of Cardiology, College of Medicine, Minia University, Egypt P. O. Box 77, Giza, Egypt†Corresponding author:

hussein.sabit@must.edu.eg. **Abstract Background:** Cardiovascular diseases (CVDs) are considered a leading cause of death worldwide. Allelic variation in the CYP2C19 gene leads to a dysfunctional enzyme, and patients with this loss-of-function allele will have an impaired clopidogrel metabolism, which eventually results in major adverse cardiovascular events (MACE). Ischemic heart disease patients (n=102) who underwent percutaneous cardiac intervention (PCI) followed by clopidogrel were enrolled in the present study. **Methods:** The genetic variations in the CYP2C19 gene were identified using the TaqMan chemistry-based qPCR technique. Patients were followed up for one year to monitor MACE, and the correlations between the allelic variations in CYP2C19 and MACE were recorded. **Results:** During the follow-up, we reported 64 patients without MACE (29 with unstable angina (UA), 8 with myocardial infarction (MI), one patient with non-STEMI, and one patient with ischemic dilated cardiomyopathy (IDC)). Genotyping of CYP2C19 in the patients underwent PCI and treated with clopidogrel revealed that 50 patients (49%) were normal metabolizers for clopidogrel with genotype CYP2C19*1/*1 and 52 patients (51%) were abnormal metabolizers, with genotypes CYP2C19*1/*2 (n=15), CYP2C19*1/*3 (n=1), CYP2C19*1/*17 (n=35), and CYP2C19*2/*17 (n=1). Demographic data indicated that age and residency were significantly associated with abnormal clopidogrel metabolism. Moreover, diabetes, hypertension, and cigarette smoking were significantly associated with the abnormal metabolism of clopidogrel. These data shed light on the inter-ethnic variation in metabolizing clopidogrel based on the CYP2C19 allelic distribution. **Conclusion:** This study, along with other studies that address genotype variation of clopidogrel-metabolizing enzymes, might pave the way for further understanding of the pharmacogenetic background of CVD-related drugs. **Keywords:** Cardiovascular; PCI; Clopidogrel; Egyptian; CYP2C19; Genotyping

Molecular Effects of Genetic Variation Posters - Thursday

PB2552*. Defining the effects of noncoding genetic variation on human regulatory element activity

Authors:

K. Strouse, G. Johnson, K. Siklenka, A. Barrera, W. Majoros, A. Allen, T. Reddy; Duke Univ., Durham, NC

Abstract Body:

Millions of noncoding genetic variants are associated with common human traits and disease. As a step towards understanding how noncoding variants contribute to disease, we are generating a population-scale atlas of the effects of 40 million genetic variants on gene regulatory element activity. To do so, we are assaying regulatory activity genome-wide across 300 individuals from diverse populations using the high-throughput reporter assay STARR-seq. We have developed a pooling approach to whole-genome STARR-seq allowing for comprehensive assaying of regulatory variation across five individuals at once. With those measurements, we will estimate the effects of >40 million variants using a Bayesian model of allele-specific STARR-seq signal. To maximize the number of genetic variants we assay, we have prioritized the genomes of individuals with African ancestry, but still include representation from diverse ancestries.

We have completed genome-wide STARR-seq assays for the first five human genomes in our study. Our assay libraries each contain four billion unique DNA fragments, covering the human genome at 300x per individual. We are already able to detect hundreds of sites with allele-specific regulatory activity in that data using our BIRD model. In total, we will assay all common, most rare, and millions of personal noncoding variants. We expect the insights gained from this work will make the identification of causal genetic variants and mechanisms more routine, allowing researchers to prioritize genes as potential therapeutic targets.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2553*. Deletion mapping of regulatory elements for *GATA3* reveals a distal T helper 2 cell enhancer involved in allergic diseases.

Authors:

G. McVicker¹, H. V. Chen¹, P. C. Fiaux², A. Sen¹, I. Luthra³, A. J. Ho¹, A. R. Chen¹, K. Guruvayurappan¹, M. H. Lorenzini¹, C. O'Connor¹; ¹Salk Inst. for Biological Studies, La Jolla, CA, ²Univ. of California San Diego, La Jolla, CA, ³Univ. of British Columbia, Vancouver, CA, Canada

Abstract Body:

The *GATA3* gene is essential for T cell differentiation and is surrounded by risk variants for immune traits. Interpretation of these variants is challenging because the regulatory landscape of *GATA3* is complex with dozens of potential enhancers spread across a large topological associating domain (TAD). Furthermore, gene expression quantitative trait locus (eQTL) studies provide limited evidence for the function of GWAS variants at this locus potentially due to small effect sizes and limited power. To address these issues, we performed a tiling deletion screen of a 2 MB region surrounding the gene in Jurkat T cells to identify 23 candidate regulatory elements for *GATA3*. Using small deletions in primary T helper 2 (Th2) cells, we validated the function of five of these elements, two of which contain risk variants for asthma and allergic diseases. We fine-mapped genome-wide association study (GWAS) signals in a distal regulatory element, 1 Mb downstream, to identify 14 candidate causal variants. Small deletions spanning candidate rs725861 decrease *GATA3* expression in Th2 cells suggesting a causal mechanism for this variant in allergic diseases. Our study demonstrates the power of integrating GWAS signals with deletion mapping and identifies critical regulatory sequences for *GATA3*.

Molecular Effects of Genetic Variation Posters - Thursday

PB2554. Detection of rare Synaptogyrin 3 (*SYNGR3*) gene missense mutations in patients with schizophrenia.

Authors:

M-C. Cheng, L-Y. Ping, K-H. Lin, T-M. Hu, S-H. Hsu, H-Y. Tsai; Yuli Branch, Taipei Veterans Gen. Hosp., Hualient Country, Taiwan

Abstract Body:

Objective: Schizophrenia is a severe psychiatric illness experienced by approximately 1% of individuals worldwide and has a debilitating impact on perception, cognition, and social function. The heritability of schizophrenia is estimated at 80% on average. Synaptic vesicle-related genes significantly regulate synaptic transmission and play a critical role in various psychiatric diseases. Synaptogyrin 3 (*SYNGR3*) is a transmembrane protein of presynaptic vesicle that warrant genetic and functional analysis for the pathogenesis of schizophrenia. **Methods:** For pathogenic mutation identification, we sequenced the exonic regions of the *SYNGR3* gene in 516 unrelated patients with schizophrenia from Taiwan. We analyzed the *SYNGR3* protein function of the identified missense mutants via immunoblotting and post-translational modification analysis. **Results:** We identified four missense mutations, and *in silico* analysis indicated that these were rare, damaging, or pathological based on putative protein function. The functional studies demonstrated that the *SYNGR3*^{p.Ala81Pro} mutant as a loss-of-function mutant in HEK-293 cells. We also showed that the two missense mutations (*SYNGR3*^{p.Ser190Arg} and *SYNGR3*^{p.Ser210Asn}) abolish the potential serine protein kinase C phosphorylation sites in the C-terminal cytoplasmic tail and one missense mutation (*SYNGR3*^{p.Pro197Thr}) create a potential threonine phosphorylation site. **Conclusion:** The results suggest that the *SYNGR3* gene harbors rare and functional disrupting mutations in some patients with schizophrenia, supporting contributing rare coding variants to the genetic architecture of schizophrenia. Of particular interest is three missense mutations (*SYNGR3*^{p.Ser190Arg}, *SYNGR3*^{p.Pro197Thr}, and *SYNGR3*^{p.Ser210Asn}) might be involved in the regulation of protein kinase C (PKC) function. Thus, more studies using a proper neuronal model are required to characterize the association between schizophrenia-associated *SYNGR3* mutants and PKC changes.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2555. Determining the rate of unintended on-target and off-target effects induced by CRISPR-Cas gene editing in iPSC clones

Authors:

C. Shum¹, Y. Han¹, B. Thiruvahindrapuram¹, Z. Wang¹, B. Zhang¹, M. Sundberg², E. Buttermore², N. Makhortova², C. Chen², M. Sahin², S. Scherer¹; ¹The Hosp. for Sick Children, Toronto, ON, Canada, ²Boston Children's Hosp., Boston, MA

Abstract Body:

CRISPR-Cas usage for genome editing can be hindered by undesired genomic changes, though the prevalence and types of these unintended edits remain poorly understood. To address this, we investigated unintended effects in induced pluripotent stem cells (iPSCs) after homology-directed repair (HDR) of disease-associated mutations using whole genome sequencing (WGS). We developed Off-Flow, a bioinformatic workflow that generates a comprehensive list of predicted off-target sites from 4 prediction programs (CHOPCHOP, Cas-Offinder, CRISPRitz, CRISPR-Offinder) based on the specific guide RNA and Cas enzyme used. We analyzed WGS data from 15 clones generated after CRISPR-Cas9 or -Cas12a editing of 6 separate genomic regions within 3 genes (*KCNQ2*, *ASH1L*, *GNAQ*). In 5 regions, CRISPR-Cas9/Cas12a was used to repair a disease-associated mutation in patient iPSC lines. In the remaining region, CRISPR-Cas9 was used to introduce a disease-associated mutation in a control iPSC line. Cell line-specific variants were intersected with predicted off-target sites from Off-Flow to assess the impact of the gene editing process. We identified a substitution and mono-allelic deletions at or near the target loci in 3 clones. A 3-bp substitution resulting in an amino acid change (p.T234L) was observed at the target locus in an edited clone after CRISPR-Cas9 repair of the target mutation *KCNQ2* c.766G>T, p.G256W. A 1,857bp mono-allelic deletion (chr20:62,068,445-62,070,301, hg19) was observed near the target locus in an edited clone after CRISPR-Cas9 repair of *KCNQ2* (c.821C>T, p.T274M). A 329bp (chr9:80412512-80412841) mono-allelic deletion was observed at the target locus after CRISPR-Cas9 introduction of *GNAQ* c.548G>A, p.R183Q. We identified 20 editing-induced missense variants at other genomic loci in 10 edited and unedited or HDR-failed clones. One such variant, *CACNA1E* c.G669T, p.Q223H, was observed in an edited clone after CRISPR-Cas9 editing of *KCNQ2* c.875_877delTCCinsCCT, p.L292_L293delinsPF. This clinically relevant gene is associated with epileptic encephalopathy and predicted to be likely damaging, thus may have functional consequences. On-target substitution and deletions had escaped standard PCR and Sanger sequencing analysis, while missense variants had escaped Off-Flow. In summary, we identified on-target substitution and deletions and off-target missense variants following CRISPR-Cas9/Cas12a HDR in iPSCs. We propose a standard for quality control of iPSC clones that have undergone gene editing by rigorous analysis of WGS data or the use of multiple edited and control unedited clones to facilitate the integrity of iPSC-based studies.

Molecular Effects of Genetic Variation Posters - Thursday

PB2556. Development of a multiplexed assay of variant effects for the DNA repair gene *MUTYH*

Authors:

S. Hemker, B. Burugula, G. Clark, S. Jayakody, J. Kitzman; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Disruption of *MUTYH* causes MYH-associated polyposis, a recessive condition for which up to 1:50 individuals are carriers, depending upon ancestry group. Nearly 800 missense mutations in *MUTYH* have been reported in ClinVar, 95% of which are classified as variants of uncertain significance (VUS), reflecting the difficulty of classifying missense variants, which can range in effect from wildtype-like to complete functional impairment. To resolve these reported VUS, and as-yet unseen patient variants in *MUTYH*, we are developing a multiplexed assay of variant effects (MAVE) to predict pathogenicity of all *MUTYH* missense variants. The normal function of *MUTYH* is to repair A:8oxoG lesions, a product of oxidative damage. We have developed an optimized cell-based reporter assay for *MUTYH* variant function, in which proper A:8oxoG to C:G repair initiated by *MUTYH* yields fluorescent protein expression. To perform this assay, we generated a CRISPR-induced *MUTYH* knockout human cell line and then introduced different *MUTYH* variants: a wildtype (WT) copy, a truncating nonsense variant (Q363X), and two pathogenic missense variants - one previously demonstrated to have partial loss of function (G368D) and one with nearly total loss of function (Y151C). The Q363X variant exhibited practically no repair of the 8oxoG lesion, while the Y151C and G368D variants were able to repair the lesion at approximately 5% and 30% of WT levels, respectively. Thus, this assay has the dynamic range to distinguish between various levels of functional impairment, and is amenable to high-throughput screens such as deep mutational scans. Using this reporter as a functional read-out, we will complement *MUTYH* knockout cells with libraries comprising every *MUTYH* missense mutation (n=9,899). The pathogenicity predictions of these assays will be 1) compared to clinically identified variants from ClinVar, 2) refined based upon evolutionary conservation, and 3) used to identify structural mechanisms underlying hotspots of functional constraint by mapping data to 3D structure.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2557. Development of a yQTL Discovery Pipeline Applicable for Both Unrelated and Related Individuals

Authors:

M. Li¹, Z. Song², A. Gurinovich³, P. Sebastiani³, S. Monti¹; ¹Boston Univ., Boston, MA, ²Boston Univ. Sch. of Publ. Hlth., Boston, MA, ³Tufts Med. Ctr., Boston, MA

Abstract Body:

Quantitative trait loci (QTL) are DNA sequence variants, such as SNPs, that influence the level of a quantitative trait, for example, gene expression. QTL discovery analysis consists of multiple steps, including genome-wide principal component analysis (PCA), genome-wide association test, as well as downstream analyses such as plotting and result annotations. In order to facilitate and automate the process, we developed yQTL discovery pipeline tool, which is agnostic to the nature of the dependent variable (y) to be modeled. In the genome-wide association step, the pipeline supports two different analysis modalities: i) one using standard linear models based on the use of the R package *matrixQTL*, which is optimized to process hundreds of phenotypes at once and yields QTL results at a high speed, but does not support the incorporation of family structure information, thus adequate only when unrelated subjects are analyzed; ii) one based on the R package *GENESIS* that supports the estimation of genetic relationship matrix (GRM) and include it in linear mixed effect models, thus able to analyze related subjects. Both modalities include the genome-wide PCA and the incorporation of user-specified covariates. Through the adoption of the workflow management tool Nextflow, the pipeline parallelizes the analysis steps. We have tested the pipeline using proteomics and metabolomics data from the New England Centenarian Study and publicly available multi-omics datasets.

Molecular Effects of Genetic Variation Posters - Thursday

PB2558. Discovery and finemapping of eGFR loci in metaanalysis GWAS of 80K African ancestry individuals

Authors:

C. Kintu¹, O. Soremekun², M. Richard³, J. Daudi³, C. Tinashe⁴, S. Fatumo⁵; ¹Dept. of Immunology and Molecular Biology, Sch. of BioMed. Sci., Makerere Univ. Coll. of Hlth.Sci., Kampala, Uganda, ²The African Computational Genomics (TACG) Res. Group, Kampala, Uganda, ³Makerere Univ., Kampala, Uganda, ⁴MRC/Wits Dev.al Pathways for Hlth.Res. Unit, Dept. of Paediatrics, Faculty of Hlth.Sci., Univ. of the Witwatersrand, Witwatersrand, South Africa, ⁵Dept. of Non-Communicable Disease Epidemiology, London Sch. of Hygiene and Tropical Med. (LSHTM), London, United Kingdom

Abstract Body:

Background: Chronic kidney disease is defined by reduced glomerular filtration rate. Genome-wide association studies of over 1 million individuals have already identified 424 loci harboring common variants that are associated with estimated glomerular filtration rate (eGFR). These studies, however, only analyzed data of European ancestry. **Methods:** To enhance the power to identify eGFR associated loci African ancestry (AA) individuals, we conducted a meta-analysis of creatinine-based eGFR in 80,027 AA participants from three global population-based cohorts: The Chronic Kidney Disease Genetics consortium (CKDGen, n = 16474 cases), UK Biobank (UKBB, n = 6217 cases), and the Million Veteran Program (MVP, n = 57336). A fixed-effect inverse-variance weighted meta-analysis was performed using METAL. For each lead variant, we computed a 99% credible set through finemapping and further performed functional analysis to map variants to tissues of importance for kidney function. **Results:** Of the 799 single nucleotide polymorphism (SNPs) that reached genome wide significant threshold of ($p < 5 \times 10^{-8}$), 17 SNPs were identified as the lead SNPs. Further downstream analysis showed that rs3798156 and rs2486272 mapping to genes *SLC22A2* and *GATM* have not been previously reported to be associated with eGFR. Bulk tissue gene expression analysis revealed that these two genes are highly differentially expressed in the kidney-medulla and kidney-cortex. Fine mapping of rs3798156 and rs2486272 showed that there were 4 and 31 credible sets with posterior probability of 71% and 36% respectively. **Conclusion:** These results provide insight into the biological mechanisms of kidney function by identifying two novel variants potentially influencing metabolic functions of the kidney unique to individuals of African ancestry.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2559. Disease-specific analysis improves prioritization of non-coding genetic variants

Authors:

Q. Liang¹, D. Kostka²; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²Univ of Pittsburgh, Pittsburgh, PA

Abstract Body:

There is increasing interest in non-protein-coding genetic variants, and in their role as genetic causes of human disease. In-silico variant prioritization methods quantify a variant's severity, and commonly used instances are single (i.e., organism-wide) scores, as well as methods that summarize a variant's impact specifically in certain tissues and/or cell-types. Because variants causing one disease may have different characteristics from those causing another disease, disease-specific variant prioritization schemes represent a complementary approach. In this study we assess variant scores' ability to prioritize disease-associated non-coding variants. We propose a simple logistic regression approach for converting tissue/cell-type specific variant scores into disease-specific scores, which also enables calculation of disease similarities based on disease-relevant tissues and cell-types.

Using ~63k non-coding variants from the NHGRI-EBI GWAS catalog with associations spanning 111 disease terms, we find that organism-wide scores (GenoCanyon, LINSIGHT, GWAVA, eigen, CADD) are only moderately successful in predicting disease associations (average precision across scores: 0.129; baseline=0.09), and disease-specific aggregation of tissue/cell-type specific scores (GenoSkyline, FitCons2, DNA accessibility) significantly improves performance (average precision across scores: 0.151; baseline=0.09). Disease similarities based on learned aggregation weights highlight meaningful disease groups (e.g., immune system related diseases and mental/behavioral disorders), while providing information on tissues and cell-types that drive similarities (e.g., lymphoblastoid T-cells for immune-system related diseases). We also show that learned similarities are complementary to purely genetic similarities as quantified by genetic correlation. Overall, this analysis makes the case for disease-specific variant prioritization; our aggregation approach leads to a gain in performance, and it enables interpretable disease models that carry information that is complementary to purely genetic similarities.

Molecular Effects of Genetic Variation Posters - Thursday

PB2560. Dissecting the downstream effects of IL-6 signaling on atheroprotection: a proteome-wide Mendelian randomization study

Authors:

M. Georgakis^{1,2,3}, S. Prapiadou¹, C. Anderson^{4,3}; ¹Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ²Inst. for Stroke and Dementia Res., LMU Munich, Munich, Germany, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Brigham and Women's Hosp., Boston, MA

Abstract Body:

Introduction: Genetic and experimental studies support a causal involvement of interleukin-6 (IL-6) signaling in atheroprotection. While clinical trials targeting IL-6 signaling for lowering the burden of cardiovascular disease are currently underway, any atheroprotective benefits must be balanced against an impaired host immune response, thus highlighting a need for discovery of more specific targets.

Dissecting the downstream mechanisms driving the effects of IL-6 signaling on atherosclerosis could offer insights about novel drug targets with more specific effects.

Methods: Leveraging data from 522,681 individuals of European ancestry in the CHARGE Consortium and UK Biobank, we constructed a genetic instrument of 26 variants in the vicinity of the gene encoding IL-6 receptor that proxied the effects of pharmacological IL-6 receptor inhibition. We explored in Mendelian randomization (MR) the effects of these variants on 3,622 plasma proteins quantified with an aptamer-based multiplex protein assay (SOMAscan) in 3,301 individuals in the INTERVAL cohort. In a mediation MR framework, we explored potential mediators in the effects of genetically proxied IL-6 signaling on cardiovascular disease endpoints (coronary artery disease, large artery atherosclerotic stroke, peripheral artery disease) in disease-specific consortia.

Results: After correction for multiple comparisons, we found significant effects of genetically proxied IL-6 signaling on 83 circulating proteins. A gene ontology enrichment analysis identified immune response, cytokine production and regulation, and cell differentiation pathways. From these 83, the genetically predicted levels of 25 proteins also showed directionally consistent significant associations with risk of coronary artery disease, large artery atherosclerotic stroke, or peripheral artery disease. Genetically predicted circulating levels of CXCL10, a chemokine involved in T cell recruitment and experimentally implicated in atheroprotection, was associated with significantly higher risk of all three endpoints. Mediation analyses revealed that a significant proportion of the effects of IL-6 signaling on cardiovascular disease endpoints were mediated by increases in CXCL10 levels.

Conclusions: Integrating genomic and proteomic data, we found a proteomic signature of IL-6 signaling activation and mediators of its effect on cardiovascular disease. Our analyses suggest CXCL10 to be a potentially causal mediator for atherosclerotic endpoints in three different vascular beds and as such might serve as a promising drug target for atheroprotection.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2561. *Drosophila* assays for variant interpretation in the polycomb repressive complex 2 related syndromes

Authors:

S. Cyrus^{1,2}, W. T. Gibson^{3,1,2}, S. Medina Giró⁴, T. Lian⁴, D. W. Allan^{4,5}; ¹Dept. of Med. Genetics, Univ. of British Columbia, Vancouver, British Columbia, Canada, Vancouver, BC, Canada, ²BC Children's Hosp. Res. Inst., Vancouver, BC, Canada, ³BC Children's Hosp., Vancouver, BC, Canada, ⁴Dept. of Cellular and Physiological Sci., Life Sci. Inst., Univ. of British Columbia, Vancouver, BC, Canada, ⁵Djavad Mowafaghian Ctr. for Brain Hlth., Univ. of British Columbia, Vancouver, BC, Canada

Abstract Body:

Variants of unknown significance pose a challenge for clinical laboratories, physicians and patients. The creation of well-calibrated assays, that accurately predict functional impact of variants, will help resolve these challenges.

The polycomb repressive complex 2 (PRC2) is an epigenetic reader/writer that catalyzes the mono-, di-, and trimethylation of H3K27. Mutations in PRC2 members *EZH2*, *EED* and *SUZ12* cause overgrowth and intellectual disability syndromes, namely Weaver syndrome, Cohen-Gibson syndrome, and Imagawa-Matsumoto syndrome respectively. Pathogenic PRC2 mutations manifest variable expressivity, conferring a wide range of phenotypic effects. Notably, 2.4% of the general population will have a rare missense or nonsense variant in one of these 3 genes. Therefore, attributing or excluding disease causality for rare coding variants may have implications for common complex diseases, in addition to rare syndromes. We have taken advantage of the powerful genetic strategies in *Drosophila* to develop inexpensive and robust assays that can screen hundreds of naturally-occurring variants seen in human populations. Our combination of strategies includes CRISPR/CAS9-based genome editing, Recombination Mediated Cassette Exchange and phiC31 integrase-mediated transgenesis to develop reagents that can test variants in *EZH2*, *EED* and *SUZ12*, including:

- (i) Mimics of the human variant in the fly's orthologous gene (fly mimetics)
- (ii) Direct expression of the human gene variant in the fly

Our assays test the ability of human PRC2 variants to rescue null phenotypes, or to cause known PRC2 phenotypes in *Drosophila*.

Fly mutants in the ortholog *esc* that mimic 6 pathogenic *EED* human variants caused early lethality, or classical *esc* phenotypes such as extra sex combs in males, but 2 benign *EED* mimetic variants displayed wildtype phenotypes. Differences in the severity of phenotypes suggest that our approach may be capable of quantifying the relative functional impact of *EED* variants. Similarly, creating a knock-in allele in the ortholog *E(z)* that mimics a pathogenic human *EZH2* variant generated a lethal phenotype.

Additionally, direct expression of wildtype human *EZH2* and a pathogenic variant in *Drosophila* demonstrated a reduced ability for the variant to cause transcriptional silencing, in a classical position effect variegation assay of the white gene. However, further work is needed to determine if this approach will be as successful as fly mimetics.

Our work to date demonstrates the utility of numerous approaches to functional characterization of *EED* and *EZH2* variants, and in the future *SUZ12* variants, using the *Drosophila* assay platform.

Molecular Effects of Genetic Variation Posters - Thursday

PB2562. Dual inhibition of mTORC1 and mTORC2 activation effectively rescues hyperproliferative lymphatic sprouting in a cell-based model system of complex lymphatic anomalies

Authors:

M. Battig¹, M. E. March¹, L. S. Matsuoka¹, S. E. Sheppard², C. Kao¹, C. Seiler¹, D. Li¹, H. Hakonarson¹; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., Rockville, MD

Abstract Body:

Generalized lymphatic anomaly (GLA) and kaposiform lymphangiomatosis (KLA) are intractable diseases of micro/macrocystic lymphatic malformations in the bones, viscera, and abdominal and thoracic cavities, which can result in respiratory compromise, multiorgan failure, and death. Recently, somatic activating mutations in *PIK3CA* and in *NRAS* have been identified in patients with GLA and/or KLA. *PIK3CA* encodes the PI3K catalytic subunit p110 α while *NRAS* encodes the GTPase N-ras, both of which are strong mediators of cell proliferation. However, despite the identification of these genes, the molecular mechanisms by which these genetic mutations drive the pathogenesis of GLA and KLA are not well understood. Towards this end, we conducted functional characterization using an *in vitro* spheroid-sprouting assay with human dermal lymphatic endothelial cells expressing variants of *PIK3CA* (p.H1047L) or *NRAS* (p.Q61R) that have previously been implicated in these diseases. In comparison to its wild-type control, *PIK3CA*(H1047L) expression resulted in significantly greater lymphangiogenic sprouting, as determined by the cumulative sprout length (1381 μ m vs. 2666 μ m, $p = 0.029$). Similarly, greater lymphangiogenic sprouting was observed when spheroids expressing *NRAS*(Q61R) were compared to its wild-type counterpart (2134 μ m vs. 4750 μ m, $p = 0.0012$). Further analyses *via* immunoblotting revealed over-activation of mTORC1 and mTORC2 targets led to the increased lymphangiogenic activity. Suppression of mTOR activation in mutant cells was tested with OSI-027 (dual mTORC1 and mTORC2 inhibitor) or rapamycin (mTORC1 inhibitor). OSI-027 was as effective as rapamycin in rescuing the hyperproliferative phenotype caused by either mutation, as evidenced by decreases in both the number and the length of capillary-like sprouts to basal levels. Immunoblotting confirmed OSI-027 inhibited lymphangiogenic sprouting by targeting both mTORC1 and mTORC2. These findings demonstrate that somatic activating mutations in *PIK3CA* and in *NRAS* act through over-activation of AKT/mTOR signaling, and inhibitor-based therapy may benefit individuals affected with complex lymphatic anomalies, particularly when multiple signaling pathways are targeted.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2563. Ehlers-Danlos Syndrome: A Likely Pathogenic Variant Found in *COL5A1* gene

Authors:

C. Rodriguez Rodriguez¹, S. Carlo², F. Velez-Bartolomei³, G. Serrano⁴, J. Pascual⁵, A. Cornier⁶; ¹Univ. Autonoma de Guadalajara Sch. of Med., Guadalajara, Mexico, ²Ponce Hlth.Sci. Univ., Cabo Rojo, PR, ³San Jorge Children and Women's Hosp., San Juan, PR, ⁴Ponce Hlth.Sci. Univ. Sch. of Med., Guadalajara, Mexico, ⁵Ponce Hlth.Sci. Univ. Sch. of Med., Ponce, PR, ⁶San Jorge Children s Hosp./ Ponce Hlth.Sci. Univ., San Juan, PR

Abstract Body:

Ehlers-Danlos syndrome is a connective tissue disorder characterized by joint hypermobility, skin hyperextensibility and may also be associated with joint dislocation and berry aneurysm among others. EDS can be subdivided on multiple types depending on inheritance and can express different degrees of severity of symptoms. The *COL5A1* gene, is in charge of the synthesis of pro-alpha 1, a component necessary for the production of type V collagen and a key component in collagen synthesis. Mutations / variations in this gene causes a form of Autosomal Dominant Ehlers-Danlos syndrome, known as the classic type, characterized by the development of soft and hyperextensible skin, and hypermobility of joints. More than 100 *COL5A1* gene mutations have been identified in people affected with EDS. The mutation in *COL5A1* affects the amount of Pro-alpha1 that a cell produces leading to fibrils that contain type V and type I collagen in the skin that are unorganized and larger than usual weakening the connective tissue and causing the clinical manifestations that characterized the disease. We are presenting the case of a 24-year-old female referred to genetic services due to signs of hyperextensible skin, hypermobility of joints and epistaxis. Patient was first referred at age 11, however she did not return for follow-up visits, rendering further testing impossible. Upon returning at the age 22, an Ehlers-Danlos gene panel test, evaluating more than 15 genes related to the disease, was performed. Results showed a pathogenic variant identified as *c.4465G>A(p.Gly1489Arg)*. This sequence replaces glycine with arginine at Codon 189 of the *COL5A1* protein (*p.Gly1489Arg*). Glycine residues are required for the structure and stability of fibrillar collagen. This variant has been observed in people affected with EDS classic type, however it is not present in population databases. Our patient presents most of the classical phenotypical findings consistent with EDS hyperflexible type making this variant a pathogenic one.

Molecular Effects of Genetic Variation Posters - Thursday

PB2564. Elucidating genomic disease associations using transcriptomic, proteomic and metabolomic data in 4,732 individuals

Authors:

E. Persyn¹, A. P. Nath^{1,2}, A. Tokolyi³, J. Marten¹, K. Burnham³, S. C. Ritchie¹, Y. Xu¹, T. Vanderstichele³, S. Lambert¹, INTERVAL Study, D. J. Roberts⁴, E. Di Angelantonio¹, J. Danesh¹, A. S. Butterworth¹, E. E. Davenport³, M. Inouye^{1,2}, D. S. Paul⁵; ¹Dept. of Publ. Hlth. and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, ²Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Inst., Melbourne, Australia, ³Dept. of Human Genetics, Wellcome Sanger Inst., Cambridge, United Kingdom, ⁴Radcliffe Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, ⁵Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom

Abstract Body:

Defining the biological mechanisms underlying the genetic associations with diseases is a major challenge in human genetics. The integration of genomic data with dynamic -omics data (e.g., transcriptomic, proteomic, metabolomic) provides a scalable and systematic approach to identifying candidate causal genes and molecular pathways underlying genetic disease loci. In this study, we generated and analyzed RNA-sequencing data from peripheral blood samples of 4,732 healthy participants in the INTERVAL study. We aimed to (1) identify the genetic determinants regulating gene expression in *cis* and *trans*, (2) integrate dynamic -omics data in INTERVAL to identify candidate causal genes and pathways, and (3) advance the understanding of the mechanisms involving these genes and pathways.

We mapped *cis*-gene expression quantitative loci (*cis*-eQTL), identifying 17,233 *cis*-eGenes (~90% of tested autosomal genes). We further investigated independent signals from *cis*-eGenes with a *trans*-eQTL analysis, discovering 2,058 *trans*-eGenes. The 2,498 *cis*-eGenes with an eSNP in a *trans*-association were enriched in transcription regulation and immune response gene pathways. We also found that 14% of these *cis*-eGenes were transcription factors including many zinc-finger proteins. We further explored the relationship between *cis*-eQTLs and other -omics datasets in INTERVAL by performing colocalization analysis to identify genes whose expression shares genetic regulation with protein and/or metabolite levels. We identified colocalized signals with 185 protein QTLs from the SomaScan platform, 41 protein QTLs from the Olink assay and 223 metabolite QTLs from the Nightingale platform. For example, we found colocalized signals between eQTLs and protein QTLs for genes associated with type I diabetes mellitus, such as *CTSH*, *NCR3* and *SIRPG*. The eQTL data presented here have been integrated into the OMICSPRED database (<https://www.omicspred.org/>), a resource to predict multi-omics data using genotype data and polygenic scores.

Future analyses will further delineate the relationship between molecular QTLs and disease associations through mediation analyses. In summary, the availability of concomitant genomic and multi-omic data in the same individuals provides an opportunity for translating genomic discoveries into therapeutic hypotheses.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2565. Elucidating mechanisms of genetic interactions and oligogenic inheritance in Autism Spectrum Disorder.

Authors:

A. Sertie¹, A. Teles e Silva¹, B. Y. Yokota¹, I. S. Nobrega¹, B. L. Zampieri¹, M. Passos-Bueno²; ¹Hosp. Israelita Albert Einstein, Sao Paulo, Brazil, ²Univ.e de Sao Paulo, Sao Paulo, Brazil

Abstract Body:

Background: Oligogenic inheritance of autism spectrum disorder (ASD) has been supported by several studies showing that a single patient can carry gene-disruptive variants in more than one gene that combined confer a higher risk for disease than their individual risk. However, little is known about how these risk variants interact and converge on causative neurobiological pathways. We identified in an ASD proband deleterious compound heterozygous missense variants in the Reelin (*RELN*) gene, and a *de novo* splicing variant in the Cav3.2 calcium channel (*CACNA1H*) gene. Using induced pluripotent stem cell (iPSC)-derived neural progenitor cells from this proband, we showed that the variants in *RELN* and *CACNA1H* are deleterious and that there is an abnormal interaction between these mutated genes via the mTORC1 pathway. Furthermore, analysis of the sequencing data from two ASD cohorts - a Brazilian cohort of 861 samples, 291 with ASD; the MSSNG cohort of 11,181 samples, 5,102 with ASD - revealed a significantly increased burden of co-occurring risk variants in both alleles of Reelin pathway genes and in one allele of calcium channel genes in ASD (PMID:35668055). However, it is still unknown how these mutated genes alter brain development and lead to ASD. **Objectives:** explore molecular and cellular phenotypes affected by the synergistic action of *RELN* and *CACNA1H* disruptive variants and those that are unique to either one of the alleles by using CRISPR-Cas9 gene editing and cerebral organoids. **Results:** We designed and cloned guide RNAs (gRNAs) to introduce into control iPSC lines the pathogenic variants of *RELN* and *CACNA1H* identified in the ASD proband, and selected the gRNAs with the highest editing rates. These gRNAs will be used for transfection of iPSCs, generating 3 gene-edited iPSC lines (either in *RELN* or *CACNA1H*, as well as in the two genes simultaneously). The edited and isogenic cell lines will then be used to generate cerebral organoids. We generated cerebral organoids from a control iPSC line that model early developmental events in the human forebrain, including the organization of neural progenitor zones similar to the ventricular and subventricular zones, and programmed differentiation of glutamatergic and GABAergic neurons and astrocytes. **Conclusion:** This work integrates risk-variant identification and functional genomics, and will provide further insight into the mechanisms of gene interactions and oligogenicity in ASD.

Molecular Effects of Genetic Variation Posters - Thursday

PB2566. Elucidating the mechanism of a neurological syndrome caused by germline mutations in *H3F3A* and *H3F3B*.

Authors:

A. Sangre^{1,2}, L. Bryant², D. Layo-Carris², R. Angireddy², E. Lubin^{3,2}, D. Brooks¹, K. Musunuru¹, E. Bhoj²; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Children's Hosp. of Philadelphia, Philadelphia, PA, ³Perelman Sch. of Med./Univ. of Pennsylv, Philadelphia, PA

Abstract Body:

Histones are fundamental proteins with key architectural and gene regulatory roles. Histone H3.3, a replication-independent variant of the canonical H3.1 and H3.2, is encoded for by two genes *H3F3A* and *H3F3B*. Among its many functions, H3.3 marks active genes and maintains genome integrity during mammalian development. Highly tumor type specific somatic mutations in H3.3 drive pediatric glioblastoma and chondroblastoma, indicating that H3.3 residues, mutations and genes have distinct functions. Recently, *de novo* missense germline mutations in *H3F3A* and *H3F3B* were identified in a cohort of 56 unrelated patients with progressive neurologic dysfunction and congenital anomalies. Unlike with the cancer variants, there is not a single causative mutation, but rather 31 unique mutations in *H3F3A* and 11 in *H3F3B*, all resulting in phenotypes of varying levels of severity. Interestingly, *H3F3A* and *H3F3B* have been shown to be independently dispensable for development in *Drosophila* and *H3F3A* or *H3F3B* null mice are overtly normal and fertile, suggesting that these missense mutations are not loss of function (LOF).

An overarching question is whether pathology is caused by each specific mutation, or rather by a more global dysregulation of H3.3. It is also not well understood at what level (protein, RNA, or both) this dysregulation is occurring. Given the breadth and distribution of mutations, as well as the apparent lack of genotype-phenotype correlation, large-scale genomic modalities and focused follow-ups are necessary to understand the dysfunction arising in this patient cohort, as well as a more general understanding of the role of missense, synonymous and LOF mutations in histones and their subsequent effects on genome organization and development. We set out to functionally classify the effect of every missense mutation in H3.3 and elucidate the contribution of *H3F3A* and *H3F3B* RNA and H3.3 protein on function in cells. We will use CRISPR technologies to study these mutations in their native contexts in a scalable and unbiased manner to both elucidate how individual mutations, and missense mutations as a broader class, affect H3.3 function. We will also disentangle the importance of RNA and protein on the function of H3.3 in cells to study the specific effects of RNA perturbation. We anticipate that this work will fill key gaps in the field's understanding of the role of histone RNA and protein in the context of development and genome regulation and will also be fundamental in developing a therapy to improve the lives of patients affected by this severe syndrome.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2567. Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease

Authors:

H. Liu¹, T. Doke¹, D. Guo², X. Sheng¹, Z. Ma¹, J. Park¹, H. Vy³, G. Nadkarni³, A. Abedini¹, Z. Miao¹, M. Palmer¹, B. Voight¹, H. Li¹, C. Brown¹, M. Ritchie¹, Y. Shu², K. Susztak¹; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Univ. of Maryland at Baltimore, Baltimore, MD, ³Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

More than 800 million people suffer from kidney disease, yet the mechanism of kidney dysfunction is poorly understood. The epigenome could provide critical insight into disease understanding as it integrates genetic and environmental signals. Here we define the genetic association with kidney function in 1.5 million individuals and identify 878 (126 novel) loci. We map the genotype effect on the methylome (methylation quantitative trait loci, meQTL) in 443 human kidneys, on the transcriptome (eQTL) in 686 samples, and on single cell open chromatin in 57,282 human kidney cells. Integrated analysis of these omics data reveal that methylation variation explains a larger fraction of GWAS heritability than gene expression. We present a multi-stage prioritization strategy based on Bayesian multiple-trait colocalization, mendelian randomization, single cell co-accessibility and enhancer-promoter contacts. Target genes are assigned for 87% of GWAS loci, but the closest gene is only prioritized 34% of the time. We highlight the key role of proximal tubules and metabolism in kidney function regulation. Further, the causal role of *SLC47A1* in kidney disease is defined in mice with genetic loss of *Slc47a1* and in human individuals carrying loss-of-function variants. Our findings emphasize the key role of bulk and single-cell epigenomic information in translating genome-wide association studies into identifying causal genes, cellular origins and mechanisms of complex traits.

Molecular Effects of Genetic Variation Posters - Thursday

PB2568. Epitranscriptomic m6A alterations in *C9orf72* ALS cerebellum

Authors:

J. Ross^{1,2}, D. Rochefort^{1,2}, H. Catoire^{1,2}, C-E. Castonguay^{1,2}, D. Spiegelman^{1,2}, M. Strong³, P. Dion^{1,2}, G. Rouleau^{4,1}, ¹McGill Univ., Montreal, QC, Canada, ²Montreal Neurological Inst. and Hosp., Montreal, QC, Canada, ³Robarts Res. Inst., Western Univ., London, ON, Canada, ⁴Montreal Neurological Inst.-Hosp., Montreal, QC, Canada

Abstract Body:

Background: One of the most common causes of amyotrophic lateral sclerosis (ALS) is an intronic hexanucleotide repeat expansion in *C9orf72*. Widespread RNA dysregulation has been observed in multiple brain regions in *C9orf72*-associated ALS. Mature RNA is chemically modified as a means to modulate gene expression in the cell. The most abundant modification, N6-methyladenosine (m6A), results in altered RNA half-life and potentially overall expression levels. As m6A RNA methylation regulates the amount and duration of transcripts, it may be that alterations to the epitranscriptome are also present in *C9orf72* ALS. We hypothesized that in addition to gene expression and splicing changes in *C9orf72* ALS, m6A RNA methylation might be altered. Further, we hypothesized that the genes in which m6A is altered might point to common pathways that could better explain the pathogenic effect of *C9orf72* expansion. **Methods:** Total RNA was extracted from the cerebellum of five ALS patients and five unaffected controls. m6A RNA was immunoprecipitated and enriched using meRIP-seq, and sequenced in parallel with paired total RNA. Regions with m6A enrichment were detected using clipper following the ENCODE eCLIP pipeline. Differential gene expression and alternative splicing were analyzed using DESeq2 and SUPPA2, respectively. Pathway enrichment analyses were conducted with the R package enrichR. **Results:** We observed substantial amounts of differentially expressed and spliced genes, consistent with previous *C9orf72* cerebellum studies. However, a striking number of regions (891) had significantly downregulated m6A in ALS cases (FDR < 0.05). Very few m6A regions were increased in *C9orf72* cerebellums (44), suggesting a global decrease in m6A methylation due to the repeat expansion. Differentially expressed genes in *C9orf72* cases did not often overlap with differentially methylated m6A regions, suggesting that m6A alterations are independent of gene expression levels. Further, pathway enrichment analysis of genes with downregulated m6A levels showed an enrichment of transcription regulation (padj = 5.53E-06). *C9orf72* itself contained a differentially m6A-methylated region near the ALS-associated expansion, while concurrently being downregulated in ALS cases. **Discussion:** The *C9orf72* expansion causes alterations in transcription that likely lead to or contribute to ALS pathology. Identifying alterations to m6A RNA modifications present in *C9orf72* ALS increases the complexity of this paradigm. Our results could explain the means by which transcriptional alterations are generated by *C9orf72*.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2569*. Escape from Nonsense-Mediated Decay: Annotation of Transcripts with Protein Truncating Variants

Authors:

J. Klonowski¹, Q. Liang¹, Z. Coban Akdemir², C. Lo¹, D. Kostka¹; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²UTHlth.Sch. of Publ. Hlth., Houston, TX

Abstract Body:

Protein truncated variants (PTVs) are genetic variants that shorten the coding sequence of transcripts, thereby changing the protein. They comprise approximately 42% of pathogenic variants in ClinVar, and one-third of transcripts that cause human genetic conditions and cancer mediate their pathogenicity through PTVs. PTVs arise from introduction of premature termination codons (PTCs), generating proteins that are typically degraded by nonsense-mediated decay (NMD), causing loss-of-function (LoF). However, some transcripts with PTCs escape NMD, resulting in translation of altered proteins that may cause a gain-of-function (GoF). While well-established rules have been identified to predict escape from NMD for transcripts with PTVs, there are currently no easy-to-use bioinformatics tools to annotate NMD escape comprehensively.

Here we present a user-friendly software tool for annotating of PTC variants at scale, applying canonical and non-canonical rules for NMD escape. Specifically, annotation is performed not only for canonical but for all variant-overlapping transcripts, resulting in a comprehensive annotation of NMD escape. In addition, we explicitly construct the alternative version of the protein for each PTV and transcript combination. This enables incorporation of indel size and nucleotide content, which leads to more principled application of distance-based rules for NMD escape, compared to methods using pre-computed indices based on reference annotation alone. We provide both a command-line interface that integrates well with variant processing workflows, as well as a web-based interface intended for interactive annotation of variants of interest.

We apply our software to all high-confidence transcripts currently annotated in Ensembl, and we will report results for NMD escape annotation considering variants from two sources. First, to study NMD-escaping transcript-variant pairs on the population level, we consider variants called across approximately 140,000 individuals in gnomAD (version 2.1.1). Second, variants from the NHGRI-EBI GWAS catalog will be investigated for the prevalence of GoF PTVs in the context of complex disease. Overall, our software enables geneticists to better leverage large-scale NMD-escape annotation of PTVs. Given that NMD escape can exert GoF effects, PTV annotation using this new bioinformatics tool may uncover disease causing variants missed by current approaches.

Molecular Effects of Genetic Variation Posters - Thursday

PB2570. Establishing a deep mutational scan for functional classification of MSH6 missense variants

Authors:

A. Scott^{1,2}, S. Jayakody¹, B. Burugula¹, J. O. Kitzman^{1,3}; ¹Dept. of Human Genetics, Univ. of Michigan, Ann Arbor, MI, ²Dept. of Internal Med., Univ. of Michigan, Ann Arbor, MI, ³Dept. of Computational Med. & Bioinformatics, Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Lynch Syndrome is a dominantly inherited colorectal and gynecological cancer predisposition syndrome caused by loss-of-function variants in genes encoding DNA mismatch repair (MMR) factors. *MSH6* is one of the key MMR factors implicated in Lynch Syndrome, and is notable in particular for its high penetrance for endometrial cancer risk. Actionability of clinical MSH6 testing is limited by the burden of missense variant interpretation as demonstrated by MSH6 missense variants in the NCBI ClinVar database, nearly all of which (2745/2802, 98.0%) are either unclassified (i.e., variants of uncertain significance) or have conflicting interpretations. To address this unmet clinical need, we applied deep mutational scanning (DMS) to systematically generate functional data to support variant classification. We have established a DMS-based platform to systematically test missense variants across MSH6, and here we describe proof-of-principle experiments targeting a 50-residue segment of MSH6 (codons 1054-1103). We generated a saturation mutagenesis library representing all 950 distinct missense variants within this region, and introduced these one at a time into human HAP1 *MSH6* knockout cells. These cells are then treated with the nucleotide analog 6-thioguanine, which selects against intact MMR activity, to deplete neutral *MSH6* variants and enrich for pathogenic variants, resulting in a loss of function (LoF) score to quantify MMR activity. Consistent with known functional constraint in this region, 18.8% of missense variants have a deleterious LoF score, especially around the critical p.R1076 residue, and among proline substitutions in regions of secondary structure. The LoF scores of this pilot tile were 100% concordant with known clinical pathogenic (4/4) and benign (5/5) classifications, and are well correlated with effects observed at the equivalent residues in the binding partner and distant paralog MSH2 in a previous DMS study from our group. These experiments demonstrate the feasibility of deep mutational scanning of all possible MSH6 missense variants, toward their prospective interpretation.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2571. Establishment of *in vitro* assay system for evaluating of *RERE* variants

Authors:

B-J. Kim¹, D. Scott²; ¹Baylor Coll. Med., Houston, TX, ²Baylor Coll. Med, Houston, TX

Abstract Body:

Deletions of chromosome 1p36 are the most common terminal deletions in humans. The arginine-glutamic acid dipeptide repeats gene (*RERE*) is located in the proximal critical region on chromosome 1p36. *RERE* is a nuclear receptor coregulator that positively regulates retinoic acid signaling by binding to transcriptional complex of nuclear receptors in the developing embryo. Pathogenic loss-of function variants in *RERE* have been shown to cause many of the defects seen in individuals with proximal 1p36 deletions. Recently, we have demonstrated that genotype/phenotype correlations exist with some *RERE* missense variants causing more severe and/or unusual phenotypes that are not seen in individuals who carry *RERE* loss-of-function variants. For example, the c.4313_4318dupTCCACC [p.Leu1438_His1439dup] variant consistently causes phenotypes that are commonly associated with CHARGE syndrome. Other rare variants have been found to cause congenital diaphragmatic hernia. To determine if these variants function as loss-of-function, dominant negative, or gain-of-function alleles, we are generating a series of variant-specific *RERE* expression plasmids. To date, we have successfully cloned plasmids carrying three truncating variants seen in humans: c.2278C>T, [p.Gln760*], c.3122delC [p.Pro1041Lysfs*], and c.4552C>T, [p.Gln1518*]. Variant *RERE* proteins generated using these plasmids were of expected molecular weight. We then determined the effects of these variants in a functional assay in which the transcription of a luciferase gene is driven by retinoic acid responsible elements (RARE) derived from the promoter of retinoic acid receptor β 2 (RAR β 2DR5-Luc). Overexpression of wild type *RERE* robustly increased luciferase activity when co-transfection of RAR β 2DR5-Luc plasmids in 293T cells. In contrast, luciferase activity was significantly reduced in 293T cells expressing the *RERE* truncating variants. These data provide evidence that the functional assay system we have developed can be used to evaluate the effects of human variants on *RERE*-mediated transcription.

Molecular Effects of Genetic Variation Posters - Thursday

PB2572. Estimating Diagnostic Noise in Panel-Based Genomic Analysis

Authors:

R. Beaumont¹, C. Wright²; ¹Univ. of Exeter Med. Sch., Exeter, United Kingdom, ²Univ. of Exeter, Exeter, United Kingdom

Abstract Body:

Background: Gene panels with a series of strict variant filtering rules are often used for clinical analysis of exomes and genomes. Panels vary in size, which affects the sensitivity and specificity of the test. We sought to investigate the background rate of candidate diagnostic variants in a population setting using gene panels developed to diagnose a range of heterogeneous monogenic diseases. **Methods:** We used the Genotype-2-Phenotype database with the Variant Effect Predictor plugin to identify rare non-synonymous variants in exome sequence data from 200,643 individuals in UK Biobank. We evaluated five clinically curated gene panels: developmental disorders (DD; 1708 genes), heritable eye disease (536 genes), skin disorders (293 genes), cancer syndromes (91 genes) and cardiac conditions (49 genes). We further tested the DD panel in 9,860 proband-parent trios from the Deciphering Developmental Disorders (DDD) study. **Results:** As expected, bigger gene panels resulted in more variants being prioritised, varying from an average of ~0.3 per person in the smallest panels, to ~3.5 variants per person using the largest panel. The number of individuals with prioritised variants varied linearly with coding sequence length for monoallelic disease genes (~300 individuals per 1000 base pairs) and quadratically for biallelic disease genes, with some notable outliers. Based on cancer registry data from UK Biobank, there was no detectable difference between cases and controls in the number of individuals with prioritised variants using the cancer panel, presumably due to the predominance of sporadic disease. However, we observed a marked increase in the number of prioritised variants in the DD panel in the DDD study (~5 variants per proband). Phasing of compound heterozygotes in biallelic genes resulted in a modest reduction in the number of prioritised variants. **Conclusions:** Although large gene panels may be the best strategy to maximize diagnostic yield in genetically heterogeneous diseases, they will frequently prioritise false positive candidate variants potentially requiring additional clinical follow-up. Most individuals will have at least one rare nonsynonymous variant in panels containing more than 500 monogenic disease genes. Extreme caution should therefore be applied when interpreting potentially pathogenic variants found in the absence of relevant phenotypes.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2573. Evaluation of regulatory potential of European ancestry *APOE* ϵ 3 *TOMM40-523*' repeat haplotypes with differential risk effects.

Authors:

M. Lipkin Vasquez, P. Bussies, F. RAJABLI, K. Hamilton, A. Griswold, M. Pericak-Vance, J. young, J. Vance, K. Nuytemans; John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL

Abstract Body:

Introduction: We showed that poly-T repeat length in *TOMM40* (*TOMM40-523*') is associated with AD risk in individuals carrying *APOE* ϵ 3 on European local ancestry (LA). 'Very long' repeats (VL, >29T) have a protective effect compared to 'short' repeats (S, <19T). We hypothesized that variants in linkage disequilibrium (LD) with *TOMM40-523*' on the European LA *APOE* ϵ 3 haplotype can modify risk for AD, potentially through *APOE* regulation. Methods: We used the short tandem repeat detection bioinformatics algorithm HipSTR and whole genome sequencing data of individuals homozygous for the *APOE* ϵ 3 European LA haplotype to type S and VL repeats. Frequency of variants on 16 S and 14 VL haplotypes were compared to determine LD with the repeat. We assessed effect on expression of VL versus S haplotypes in the *APOE* promoter region with or without the known enhancer in *TOMM40* intron 2 using luciferase reporter assays in astrocytes, microglia and neurons. Results: We identified a 16kb LD block surrounding *TOMM40-523*' harboring 21 variants in strong LD ($r^2 > 0.9$) with the repeat (hg19, chr19:45,395-45,411k); including the *APOE* promoter and several potential enhancer regions. Preliminary data indicate the S promoter and VL enhancer infer higher activity in microglia and astrocytes, common combinations of promoter and enhancer haplotypes (S-S) versus (VL-VL) do not differ significantly in activity. Additional analyses including other potential enhancer regions, comprising the repeat itself, are currently ongoing to investigate other cumulative effects. Discussion: Our results indicate that differential activity of promoter and enhancer is dependent on the repeat length haplotype, potentially contributing to the AD risk difference. These regulatory regions may be target of future therapies that could help a large amount of people, given *APOE* ϵ 3's frequency in the general population.

Molecular Effects of Genetic Variation Posters - Thursday

PB2574. Evaluation of the ACMG PM5 variant pathogenicity evidence guideline.

Authors:

G. Maston; Quest Diagnostics, Secaucus, NJ

Abstract Body:

Background: Interpreting the clinical significance of sequence variants remains one of the most challenging tasks in clinical genetic testing. One of the recommended criteria for pathogenicity states “Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before” (ACMG evidence code PM5). This line of evidence is given moderate weight, but to date no formal evaluation of the reliability of this line of evidence has been published. Here, the strength of this guideline is examined using high-throughput functional data (Findlay GM, Daza RM, Martin B, et al. Accurate classification of *BRCA1* variants with saturation genome editing. *Nature*. 2018;562(7726):217-222). **Methods:** In the Findlay study, saturation genome editing was used to produce 1,839 missense changes from 326 different codons located in 2 regions of the *BRCA1* protein and tested them in a cell survival-based assay to determine their functional impact. Functional results were classified as damaging, non-damaging, or intermediate. The study produced 96.5% of all possible single nucleotide variants in these regions, generating data for an average of 5.6 missense changes per codon. The data were re-examined in this study to calculate the proportions of damaging and non-damaging changes within each codon.

Results: At least 1 missense change was functionally damaging at 156 out of 326 (47.85%) codons examined, while 101 sites had 2 or more damaging variants. From these data, we estimate the probability of there being a 2nd damaging variant in a codon given that there was at least 1 damaging variant is 64.74%, which is a 1.35-fold enrichment compared to the number of codons where any damaging variants were found. However, 133 of 156 sites with at least 1 damaging variant also had at least 1 non-damaging variant. In fact, damaging variants are only marginally more common than non-damaging variants at these sites: there was an average of 2.59 damaging variants but also 2.47 non-damaging variants.

Conclusions: These data suggest that this line of evidence needs to be re-examined for its utility, or possibly that the amount of weight it is given be re-calibrated on a gene-by-gene basis. Notably, these data come from 2 regions of the *BRCA1* protein that have known functions (RING domain and BRCT domains), thus there is likely a higher proportion of damaging variants in these data than would be expected in other regions of a protein. Further study is needed, including testing in other regions of the *BRCA1* protein and also in other proteins, to determine if these patterns are universal or unique to the domains of *BRCA1* that were studied.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2575. Evidence from a massively parallel reporter assay and genome engineering for non-coding variants at 1q32-*IRF6* that directly influence risk for orofacial cleft

Authors:

P. Kumari^{1,2}, R. Friedman³, L. Pi⁴, S. Curtis⁵, K. Paraiso⁶, A. Visel^{6,7,8}, E. Leslie⁵, M. White³, B. Cohen³, R. A. Cornell^{1,2}; ¹Current address: Dept. of Oral Hlth.Sci., Univ. of Washington, Seattle, WA, ²Dept. of Anatomy and Cell Biology, Univ. of Iowa, Iowa City, IA, ³Dept. of Genetics, Washington Univ. in St. Louis, St. Louis, MO, ⁴PharmaLex, 1700 District Ave, Burlington, MA, ⁵Dept. of Human Genetics, Emory Univ. Sch. of Med., Atlanta, GA, ⁶Environmental Genomics and Systems Biology Div., Lawrence Berkeley Lab., Berkeley, CA, ⁷U.S. Dept. of Energy Joint Genome Inst., Lawrence Berkeley Lab., Berkeley, CA, ⁸Univ. of California, Merced, CA

Abstract Body:

Variation at the 1q32/*IRF6* locus accounts for about 12% of genetic contribution to isolated orofacial cleft, the most at any single locus. To identify the functional subset of single nucleotide polymorphisms (SNPs) at 1q32 and 7 other loci we conducted a massively parallel reporter assay in a human foetal oral epithelium cell line. SNPs from multiple loci yielded allele-specific effects in this assay. We pursued ten such SNPs at the 1q32/*IRF6* locus with luciferase reporter assays. For two such SNPs, one 10 kilobases upstream, and the other 21 kilobases upstream of the *IRF6* transcription start site, we separately engineered cell clones to be homozygous for the risk-associated allele or the non-risk associated allele. Interestingly, for both SNPs, the expression level of *IRF6* was relatively lower in the former. The *IRF6* - 21kb SNP had allele-specific effects on binding of transcription factor ETS2, nominating *ETS2*, or a paralogue of it, as an OFC risk gene. Conditional analyses suggest that the *IRF6*-10kb and *IRF6*-21kb SNPs together account for most or all of the association signal at 1q32/*IRF6* for cleft lip only and for cleft lip with or without cleft palate. These experiments identify non-coding variants that directly affect risk for orofacial cleft. The methods are generalizable whenever an appropriate cell line is available. Funded by NIH DE027362, USA.

Molecular Effects of Genetic Variation Posters - Thursday

PB2576. Evolutionary study and variant analysis in the *mcph1* gene in patients diagnosed with non-syndromic hearing impairment

Authors:

O. Oluwole¹, K. James¹, A. Wonkam²; ¹Univ. of Cape Town, Cape Town, South Africa, ²Univ Cape Town, Cape Town, South Africa

Abstract Body:

The genetic causes of Hearing impairment (HL) in Africa are unknown despite increasing number of cases. We investigated human mouse orthologous HL genes in patients to identify rare pathogenic variants with possible causal effects in the African individuals. The study identified a homozygous mutation in the exon 13 BRCT2 of microcephalin1 (*MCPHI*) gene in a Cameroonian patient diagnosed with non-syndromic hearing impairment (NSHI). The present study combined evolutionary analyses with DNA sequencing to screened multiplex family; ninety (n=90) NSHI patients and 106 controls recruited from Cameroon and South Africa. The estimated mode of inheritance of NSHI in the cohort was 34.8% autosomal recessive, 34.8% autosomal dominant, 21.74% mitochondrial, and 8.66% X-linked. The approaches include whole exome sequencing, Sanger Sequencing and High-Resolution Melt analyses. We identified rare variants in the investigated genes. However, four rare missense variants and seven novel variants were identified in the candidate gene (*MCPHI*). The variants *MCPHI* c.2222G>A p.(R741Q) (Alt Allele A=0.0000) and novel homozygous *MCPHI* c.2234A>C p.(H745P) identified were absent in the 106 ethnically matched controls. The evolutionary analyses revealed that the *MCPHI* protein evolved in 150 taxa while about 28 condensed in a phylogeny cluster that indicated similar substitution rates, divergent lengths, and positive selections, particularly in the two closest taxa to humans (chimpanzee and gorilla). The protein modelling and surface hydrophobicity analyses suggest a change in atomic charges at the helix-loop that mediates dimerization and DNA binding, such that the wildtype equilibrates at 0.072 nm while the mutant equilibrates at 0.042 nm. The algorithms used also detects an increase in hydrophobicity *in-silico*. Our study suggests a further understanding of the roles of *MCPHI* gene in NSHI using functional assays.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2577. Examination of predicted loss of function (pLoF) variants in the general population reveals high rates of rescue and artifacts

Authors:

S. Gudmundsson^{1,2,3}, M. Singer-Berk¹, N. Watts¹, J. Goodrich¹, M. Wilson¹, Genome Aggregation Database Consortium, S. Baxter¹, H. Rehm^{1,3,4}, D. MacArthur^{1,5,6}, A. O'Donnell-Luria^{1,2,3}; ¹Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, ²Div. of Genetics and Genomics, Boston Children's Hosp., Boston, MA, ³Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, ⁴Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁵Ctr. for Population Genomics, Garvan Inst. of Med. Res. and UNSW Sydney, Sydney, New South Wales, Australia, ⁶Ctr. for Population Genomics, Murdoch Children's Res. Inst., Melbourne, MA

Abstract Body:

One of the most surprising outcomes from genome sequencing of apparently healthy individuals is the prevalence of gene-disrupting variants that are expected to cause severe disease. This could suggest that incomplete penetrance (the proportion of individuals who carry a genotype but do not develop the associated disorder) may be a common feature of human disease. However, it is also possible for such variants to arise from incorrect prediction of functional impact. We investigate this by large-scale analysis of disease-associated predicted loss of function (pLoF) variants in 76,156 genomes from the Genome Aggregation Database (gnomAD).

We assessed all pLoF variants (n=899), in 77 haploinsufficiency genes causing severe, early-onset, highly penetrant disorders expected to be excluded from gnomAD. Variants had passed gnomAD quality control pipelines, and were determined high-confidence by the Loss-Of-Function Transcript Effect Estimator (LOFTEE). We identified evidence for LoF evasion in 595 of 899 variants (66%). Of 595 variants, 49% displayed known mechanisms of rescue, with the top reasons being escape of nonsense-mediated decay because of location in the last exon (22%), location at a site with low per-base expression (pext; 11%), rescue because of being part of a multi-nucleotide variant or in phase with frame-restoring indels (8%), and splice variants with predicted cryptic splice rescues (5%). An additional 17% were filtered as artifacts or of lesser biological importance, with the top reasons being presence in a non-coding transcript (7%), being somatic (allele balance below 25%; 5%), location in a homopolymer (3%), or location in a region with known mapping difficulties (2.7%). The remaining 33% of variants had more than one reason identified, with the most common combinations being last exon with low pext score or low allele balance. Of the remaining 304 pLoF variants, 131 variants were predicted to cause true LoF with no known rescue mechanisms identified. We are now exploring if high-impact splice(s)- and expression(e)- qualitative trait loci (QTL), structural or regulatory variants could explain their incomplete penetrance.

Our results demonstrate the importance of extending LoF assessment beyond standard annotation and reveal the challenges of variant interpretation in general population sequencing. We caution the direct use of pLoF variants, particularly when identified in settings where the prior probability of pathogenicity is low such as genomic screening programs of healthy individuals. Further analysis will be required to understand how LoF variants predicted to cause true loss do not cause disease in these individuals.

Molecular Effects of Genetic Variation Posters - Thursday

PB2578. Exome analysis for bronchial asthma in the Mennonite population reveals an association with *NOTCH4* polymorphisms

Authors:

I. Bucco¹, V. L. Carvalho¹, C. G. Meissner¹, L. C. Oliveira¹, P. I. dos Santos¹, Regeneron Genetics Center, F. McMahon², F. L. Lopes³, A. B. W. Boldt¹, G. C. Kretzschmar¹; ¹Federal Univ. of Paraná, Curitiba, Brazil, ²NIH/NIMH, Bethesda, MD, ³Federal Univ. of Rio de Janeiro, Rio de Janeiro, Brazil

Abstract Body:

Mennonites are a Christian Anabaptist group with cc. 500 years genetic isolation and a closed community structure which, added to three bottleneck events, reduced their genetic diversity and increased their susceptibility to chronic diseases, including bronchial asthma (BA). We screened 213 South-Brazilian Mennonites from Colônia Nova (CON), 157 from Colônia Witmarsum (CWI), and 185 from Curitiba (CWB) for BA and compared the exomes of 49 Mennonites that had BA or reported first-degree affected relatives with 41 that were unaffected and reported unaffected relatives, evaluating the results with multivariate logistic regression. Cc. 12% of the interviewed Mennonites had BA (CON: 8.1%, CWI: 15%, CWB: 14.3%), compared with 4.5% among Brazilians. Thirty-four variants were associated with a dominant effect for BA ($p < 0.005$) in the exome-wide analysis. Among these, 22 are associated with mRNA levels in the lungs/whole blood (GTEx), 8 replace amino acids, and 14 modify CpG sites. Twenty-two (73%) were associated with protection from BA ($p < 0.0005$), but 54% present a lower frequency among Mennonites ($p < 0.006$), compared with those in the non-Finnish Europeans and/or Brazilians. Thus, the higher BA prevalence in Mennonites is associated with a founder effect decreasing the frequency of protective alleles. Among the risk polymorphisms, we genotyped rs423023 of the *NOTCH4* gene with PCR-SSP in an increased sample of 261 Mennonites. This variant breaks a CpG site and is associated with the reduction of gene expression and with the alternative processing of its pre-mRNA in the lung. We genotyped it with another SNP at 5' (rs520803) by PCR-SSP, and found that the *rs520803_rs423023*AG*, with 30.6% frequency was also associated with an additive effect for BA susceptibility (OR = 2.15; CI95% = 1.08 - 4.39; $p = 0.031$), confirming the association previously identified in the exomes. Although the complexity of this genomic region, with high HLA linkage disequilibrium, implies difficulty in identifying a causal haplotype, we suggest that *rs520803_rs423023*AG* contributes to BA susceptibility through alternative splicing and greater expression of common *NOTCH4* transcripts in the lung, increasing the activation of inflammatory pathways.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2579. Exome-wide association analyses of 1,463 proteins in 31k participants of the UK Biobank reveal the spectrum of coding variant effects on protein levels.

Authors:

H. Kim, J. Chiou, UK Biobank Pharma Proteomics Project, E. B. Fauman, M. R. Miller; Internal Med. Res. Unit, Pfizer, Cambridge, MA

Abstract Body:

Circulating proteins are valuable biomarkers of health and disease and can be attractive targets for therapeutic modulation. While genome-wide association studies using imputed variants have identified numerous protein quantitative trait loci (pQTLs), the effects of rare coding variants on protein levels have not been extensively studied. To characterize the genetic architecture of circulating proteins by rare coding variants, we performed exome-wide association analyses of 1,463 proteins quantified by Olink Explore 1563 platform in 30,838 unrelated individuals of European ancestry from the UK Biobank. We identified 4,613 significant cis-pQTLs by coding variants (hereafter referred to as “coding cis-pQTLs”) for 993 proteins (P-value < 2.9e-7). The class of variants predicted to have more severe consequences had higher proportions of variants with significant cis-pQTLs and greater median effect magnitudes in the following order: 1) predicted loss-of-function variants, including frameshift, stop-gain, splice acceptor/donor, and start-loss variants (5.1%, -2.1 sd), 2) in-frame insertion/deletion variants (3.5%, -1.3 sd), 3) missense variants (2.6%, -0.5 sd), and 4) synonymous variants (2.4%, -0.1 sd). The effect magnitudes were also inversely correlated with the minor allele frequency, and for missense variants, positively correlated with the number of algorithms that predicted them to be deleterious. Of the 4,613 coding cis-pQTLs, 3,774 (82%) remained significant when adjusted for GWAS cis-pQTLs within +/-500kb of the gene, and the overall pattern remained comparable.

One challenge in interpreting coding cis-pQTLs arises from the possibility that coding variants may directly alter the binding of antibodies used for protein quantification. Given that coding variants are unlikely to affect the quantification of other proteins in this manner, the presence of trans-pQTLs by the same coding variants can add credibility to the observed coding cis-pQTLs. Of the 4,606 coding variants with cis-pQTLs, 3,622 variants (79%) had at least one significant trans-pQTL (P-value < 8.3e-9), supporting the reliability of 3,629 cis-pQTLs by those variants.

In summary, our study provides a reference for coding variant effects on protein levels and suggests an approach to increase the confidence in cis-pQTLs by coding variants. We are further exploring ways to incorporate the observed effects of coding variants on protein levels into gene burden association test to boost statistical power for disease gene discovery.

Molecular Effects of Genetic Variation Posters - Thursday

PB2580. Expanding the rare *DHCR24*-related sterol biosynthesis disorder: genotype -phenotype correlation.

Authors:

D. Cocciadiferro¹, D. Vecchio², V. Lanari¹, E. Agolini¹, A. Villani¹, F. Petrizzelli³, T. Biagini³, T. Mazza³, D. Martinelli⁴, M. Macchiaiolo², A. Bartuli², A. Novelli¹; ¹Translational Cytogenomics Res. Unit, Bambino Gesù Children's Hosp., IRCCS, Rome, Italy, Rome, Italy, ²Rare Diseases and Med. Genetics Unit, Academic Dept. of Pediatrics, Bambino Gesù Children's Hosp., IRCCS, Rome, Italy, Rome, Italy, ³Bioinformatics Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, S. Giovanni Rotondo (FG), Italy, Rome, Italy, ⁴Dept. of Pediatric Subspecialties, Div. of Metabolic Diseases, Bambino Gesù Children's Hosp. IRCCS, Rome, Italy, Rome, Italy

Abstract Body:

Desmosterolosis is a rare sterol biosynthesis disorder characterized by multiple congenital anomalies, failure to thrive, intellectual disability and elevated levels of desmosterol caused by biallelic mutations of *DHCR24* encoding 3- β -hydroxysterol Δ -24-reductase. In brain cholesterol synthetic metabolism, *DHCR24* protein is known as the heavily key synthetase in cholesterol synthesis, catalyzing the reduction of the delta-24 double bond of sterol intermediates during cholesterol biosynthesis. Until date, less than 10 *DHCR24* pathogenic variants have been identified in association with desmosterolosis. Here, we describe a patient with developmental delay, epileptic encephalopathy, spastic tetraparesis, ventriculomegaly, epicanthus, elevated desmosterol levels, cataract and dolichocephaly harboring the never described before missense variant c.506T>C (p.Met169Thr) in *DHCR24*. Using molecular dynamics simulation techniques, we have investigated the impact of this mutation on the protein's stability and interaction network and preliminary assessed its pathogenic role. In conclusion, this report expands the clinical and molecular spectrum of *DHCR24* related disorder, reporting a novel *DHCR24* variant associated with desmosterolosis syndrome.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2581. Expansion of the histidine-rich domain of *DYRK1A* in a patient with developmental delay, short stature, and microcephaly: A case report.

Authors:

R. Fox, G. Hortopan, E. Yusko, G. Mendiratta-Vij, R. Zimmerman, A. Birch, L. Edelmann; Sema4, Stamford, CT

Abstract Body:

Background: *DYRK1A*-related intellectual disability syndrome is a dominant disorder caused by *de novo* pathogenic variants in the *DYRK1A* gene. It is characterized by intellectual disability, developmental, speech and language delays, microcephaly and facial dysmorphism. We report a novel expansion of the histidine-rich domain (HRD) of the *DYRK1A* gene in a female patient with unexplained developmental delay, short stature, and microcephaly. Our findings provide the first evidence of a critical role for the size of the HRD in the *DYRK1A*-intellectual disability syndrome. **Methods:** Next generation sequencing was performed using a 228-gene autism panel. Sanger sequencing was used for variant confirmation in the proband and parents. Variant analysis was performed using the Agilent Alissa software and variant classification was determined using ACMG and ClinGen recommendations. Consent was obtained to deidentify this sample and use it for research purposes. **Results:** We identified a novel heterozygous in frame duplication variant in the coding region of exon 11 of the *DYRK1A* gene, NM_130436.2:c.1801_1830dup, p.His601_His610dup. The mutation was absent in the mother and father, indicating that this variant is *de novo* in this individual. *DYRK1A* encodes a dual specificity tyrosine-phosphorylation-regulated kinase 1A, that contains a highly conserved HRD that consists of 13 consecutive histidines in the C-terminus. Functional studies of the HRD have shown that it is important for nuclear speckle localization and phosphorylation of serine and threonine residues on exogenous substrates and its own kinase domain. To date, 84.5% of individuals reported with *DYRK1A*-related intellectual disability syndrome have a *de novo* variant. Per ACMG criteria, small in-frame deletions/insertions in repetitive regions without a known function are less likely to be pathogenic. However, this region is not highly polymorphic in the Genome Aggregation Database (<http://gnomad.broadinstitute.org>) with observed duplications not exceeding 2 amino acids. Biochemical characterization of missense mutations occurring in the HRD result in altered DYRK1A catalytic activity. No pathogenic expansions of the HRD have been reported in patients. **Conclusion:** Here, we report a rare duplication variant in the HRD of the *DYRK1A* gene in a female with developmental delay, short stature, and microcephaly. Based on the clinical presentation and family history, the confirmed *de novo* in frame duplication p.601_p.610 in the HRD in *DYRK1A* was classified to be likely pathogenic and disease causing. To our knowledge, this is the first report of a repeat expansion in *DYRK1A* causing this disorder.

Molecular Effects of Genetic Variation Posters - Thursday

PB2582. Explicable prioritization of genetic variants by integration of rule-based and deep learning algorithms for diagnosis of rare Mendelian disorders.

Authors:

K. Lee, H. Lee, J. Woo, D-w. Kim, G. Seo; 3billion, Seoul, Korea, Republic of

Abstract Body:

Using exome, genome and even transcriptome sequencing for making a molecular diagnosis for patients with rare Mendelian disorders has become a common clinical practice. However, determining variant pathogenicity is not a simple task and therefore many variants remain of uncertain significance albeit numerous in silico tools being developed to predict how damaging the variants could be. Here, we report a comprehensive algorithm that prioritizes variants with higher sensitivity than other publicly available tools with added capabilities of annotating variants with the evidence that was used for the prioritization. We integrated three different scores into EVIDENCE, an internally developed variant annotation and classification tool. First, the Bayesian score is based on the 28 criteria defined by the ACMG/AMP variant interpretation guidelines. Second, the symptom similarity score quantifies the semantic similarity between the known symptoms of a specific disorder and those observed in the patient. Third, 3Cnet score, generated by a trained deep-learning model, provides the likelihood of a given amino-acid change impacting the protein function. 3Cnet scores add extra refinement to the variant classification made by the first two scores.

Using an in-house dataset of 1,817 WES patient samples, we compared how our model performs compared to other prioritization algorithms. Our method was able to find the causal gene within the top-5 genes at a recall rate of 93.5 %, while LIRICAL and Exomiser did so at recall rates of 73.9 % and 73.4 % respectively. Reflecting this result, our algorithm performed best in the recent CAGI6 SickKids challenge by successfully identifying ten out of fourteen causal genes. Among them, two genes that showed decreased gene expression in RNA-seq data, *CASK* and *HDAC8*, were identified with the estimated probabilities over 0.96 and 0.93 respectively. Although our tool does not directly calculate the likelihood of a variant affecting gene expression level, we were able to correctly predict that the variants in these two genes might result in decreased expression. These results show that the gene prioritization algorithms can be explicable yet sensitive by integrating rule-based algorithms and machine learning algorithms.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2583. Exploiting a human stem cell-based modeling approach to functionally characterize loss-of-function of lncRNA RMST as a cause of Kallmann Syndrome.

Authors:

L. Stanton, G. Ali, G. Alkhadairi, W. Habbab, Y. Park; Qatar BioMed. Res. Inst., Doha, Qatar

Abstract Body:

Human genetics identified mutations in several genes that cause isolated GnRH deficiency (IGD), a rare Mendelian disorder (1) manifested by abnormal puberty, hypogonadotropism, infertility, and anosmia in Kallmann Syndrome, a phenotypic subset of IGD. A balanced translocation in a KS proband disrupted the gene for the long noncoding RNA (lncRNA) RMST thus implicating it as a genetic regulator of GnRH neuronal development (2). Previously we showed that RMST regulates neuronal development through a physical interaction with SOX transcription factors during neurogenesis (3).

To establish a model system to characterize the role of RMST, we have used Crispr gene editing to delete RMST in human embryonic stem cells (ESC). Small (5kb) and large (40kb) genetic deletions of both alleles of RMST were created to mimic the disruption of RMST resulting from a balanced translocation observed in a KS patient. Several independent, gene edited ESC lines have been generated. Genetic loss of RMST was confirmed by DNA sequencing and absence of RMST transcripts. The RMST mutant ESC lines were shown to retain pluripotency as assessed by conventional means.

The ESC lines with RMST deletion have been differentiated *in vitro* to human neural progenitors and subsequently into human GnRH neurons. We demonstrated that RMST mutant lines generate neural progenitors and GnRH neurons using our directed *in vitro* differentiation protocols. The electrophysiological activity of the mutant neurons relative to wild-type neurons is now being evaluated. To understand the molecular basis of RMST knockout we are performing RNA-seq on wild-type and mutant neural progenitors and neurons.

We will update on the progress of using this *in vitro* model of Kallmann syndrome to dissect the molecular mechanism of RMST in neuronal development and the pathological manifestations of its loss of function on GnRH neurons.

1. Stamou MI, Cox KH, Crowley WF Jr. Discovering genes essential to the hypothalamic regulation of human reproduction using a human disease model: adjusting to life in the “-Omics” era. *Endocr Rev.* 2015;2016(1):4-22. [PubMed: 27454361]

2. Stamou MI, Ng SY, et al, Stanton LW, and Crowley WF. A balanced translocation in Kallmann Syndrome implicates a long noncoding RNA, RMST, as a GnRH neuronal regulator. *J. Clin Endocrinol Metab.* 2020;105(3):231-44. [PubMed:31628846]

3. Ng SY, Bogu GK, Soh BS, Stanton LW. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol Cell.* 2013;51(3):349-359. [PubMed: 23932716]

Molecular Effects of Genetic Variation Posters - Thursday

PB2584. Fast Forward - is Multiomics a resurgence of old?

Authors:

M. Hegde¹, F. Guo², C. Collins³, L. Bean¹, B. Nallamilli¹, N. Guraju¹, H. Liu¹; ¹PerkinElmer, Waltham, MA, ²PerkinElmer Genomics, Pittsburgh, PA, ³PerkinElmer, Boston, MA

Abstract Body:

Multiomics including genome, proteome, transcriptome, and epigenome is an approach to collectively assessing and analyzing data sets. One of the challenges for next-generation sequencing (NGS) based tests is to understand the disease mechanism of the variants of uncertain significance (VOUS) in known Mendelian genes, and newly discovered genes. . Incorporating data generated by RNA sequencing, and biochemical testing, we have reclassified variants from uncertain significance to disease-causing (DMD c.4806T>A; DYSF duplication of exons 10-35; GAA c.1796C>A; ABCD1 c.598G>A). Among 81 cases for GAA enzyme activity and molecular testing, 26 cases with biallelic pathogenic variants in GAA showed over 50% the reduction of the lysosomal alpha-glucosidase enzyme activity (0.42 $\mu\text{mol/L/hr}$; normal range: $\geq 2.10 \mu\text{mol/L/hr}$), despite the GAA variants is associated with mild or severe Pompe disease. Nineteen individuals carrying one pathogenic GAA variant demonstrated average enzyme activity of 1.63 $\mu\text{mol/L/hr}$. The remaining 36 individuals carried two VOUS, or one VOUS and one pseudodeficiency allele, or two pseudodeficiency alleles, or negative. The enzyme activity ranged from 0.72-5.53 $\mu\text{mol/L/hr}$. Similar correlations between the enzyme activity and biallelic pathogenic variants were observed in 36 cases for GBA testing, of which 27 individuals carrying biallelic GBA pathogenic variants with significantly decreased enzyme activity (0.26-0.78; normal range: $\geq 1.60 \mu\text{mol/L/hr}$). Among these 27 individuals, 23 presented increased lyso-Gb1 (19.19->200; normal range: $\leq 17.41 \text{ ng/mL}$) whereas the other 4 in normal range. A remarkable reduction of GLA enzyme with an average of 0.53 $\mu\text{mol/L/hr}$ (normal range: $\geq 1.10 \mu\text{mol/L/hr}$) was observed in 20 out of 33 males tested for Fabry disease. The GLA variants correlated with less severe symptoms and reduced penetrance were associated with close to the normal range of Lyso-Gb3 ($< \text{ or } = 1.11 \text{ ng/ml}$). Our data shed light on the correlation between biochemical results and molecular findings, as well as the genotype-phenotype relationship in some lysosomal storage disorders. These methodologies have been widely used in clinical laboratories for over two decades. The recent technological advances in multiomics approaches will aid in improving the diagnostic performance to tailor personalized treatment, but it's required datasets from multiple measurements, large reference databases for different genes and diseases, integrated analysis methods, and a strong computational infrastructure.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2585*. Fine-mapping causal tissues and genes at disease-associated loci

Authors:

B. Strober, T. Amariuta, M. Zhang, A. Price; Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract Body:

Heritable diseases often manifest in a highly tissue-specific manner; this knowledge is fundamental to our understanding of disease biology and can provide insights into tissue-mediated mechanisms of disease (Hekselman et al. 2020 Nat Rev Genet). Previous studies have generally prioritized relevant tissues for disease based on genome-wide patterns of disease enrichment. However, disease effects at different GWAS loci may be mediated by different tissues, motivating efforts to identify relevant tissues at individual GWAS loci.

Here, we introduce a new statistical method, Tissue-Gene Fine-Mapping (TGFM), to infer the posterior probability that each tissue and gene mediate a given disease association based on GWAS summary statistics and expression quantitative trait loci (eQTL) from diverse tissues. TGFM consists of two steps. First, TGFM estimates gene expression-disease effects for each tissue and gene while accounting for co-regulation (correlations of cis-regulatory SNP effects) across tissues and genes; this step makes use of transcriptome-wide association study (TWAS) statistics, which quantify marginal effects of genetically predicted gene expression on disease (without accounting for co-regulation). Second, TGFM estimates the posterior probability that a specific disease-associated locus is mediated through each tissue and gene given the gene expression-disease effect estimates from the first step. TGFM improves tissue and gene fine-mapping accuracy by incorporating genome-wide estimates of the contribution of gene expression in each tissue to disease as tissue-level priors, and also accommodates gene-level priors.

We apply TGFM to a broad set of 36 diseases/traits and 38 GTEx tissues to fine-map hundreds of disease loci to causal tissues and genes. We evaluate TGFM using 10 independent diseases/traits whose biology is widely believed to involve a single causal tissue. For white blood cell count, TGFM correctly maps 31.3% of GWAS loci to whole blood tissue, vs. 6.4% of loci for a baseline approach that compares tissue-specific probabilities computed by coloc, a widely used gene expression-disease colocalization method. For diastolic blood pressure, TGFM correctly maps 25.4% of GWAS loci to artery tibial tissue, vs. 5.2% of loci for the baseline approach. Despite outperforming the coloc-based baseline, TGFM still fails to correctly fine-map the causal tissue at the majority of disease loci, reflecting the difficulty of this challenge given existing genomic data. In summary, we have developed and evaluated a statistically principled method for fine-mapping the causal tissue and gene at individual GWAS loci.

Molecular Effects of Genetic Variation Posters - Thursday

PB2586. FOXI3 pathogenic variants cause one form of craniofacial microsomia

Authors:

S. Antonarakis¹, K. Mao², C. Borel³, M. Ansar⁴, A. Jolly⁵, P. Makrythanasis⁶, C. Froehlich⁷, J. Iwaszkiewicz⁸, B. Wang⁹, X. Xu², X. Blanc¹⁰, J. Cai², Q. Qi Chen⁹, F. Jin², H. Harinarayana Ankamreddy¹¹, S. Singh¹¹, H. Zhang¹¹, X. Wang², S. Zhao¹², E. Ranza¹³, S. Paracha¹⁴, V. Guida¹⁵, F. Picci-Sparascio¹⁵, D. Melis¹⁶, B. Dallapiccola¹⁷, M. Digilio¹⁸, A. Novelli¹⁸, M. Magliozzi¹⁸, M. Fadda¹⁹, H. Streff²⁰, R. Lewis²¹, V. Zoete⁸, G. Squeo¹⁵, P. Prontera²², G. Giorgia Mancano²³, G. Gori²⁴, M. Mariani²⁵, A. Selicorni²⁵, S. Psoni²⁶, H. Fryssira²⁷, T. Cox²⁸, S. Douzgou²⁹, S. Marlin³⁰, S. Biskup³¹, A. De Luca³², G. Merla³³, A. Groves¹¹, J. Lupski³⁴, Q. Qingguo Zhang⁹, Y-B. Zhang²; ¹Univ. of Geneva, Geneva, Switzerland, ²Beihang Univ., Beijing, China, ³Univ. of Geneva, Geneva 4, Switzerland, ⁴Jules-Gonin Eye Hosp., Univ. of Lausanne, Lausanne, Switzerland, ⁵1 Baylor Plaza, Houston, TX, ⁶Univ. of Geneva, Geneva, Geneva, Switzerland, ⁷CeGaT, Tubingen, Germany, ⁸Swiss Inst. of Bioinformatics, Lausanne, Switzerland, ⁹Chinese Academy of Sci., Beijing, China, ¹⁰Medigenome, Geneva, Switzerland, ¹¹Baylor, Houston, TX, ¹²Capital Med. Univ., Beijing, China, ¹³Medigenome, Swiss Inst. of Genomic Med., Geneva, Switzerland, ¹⁴Kohat Med. Coll. (KIMS), Kohat, Pakistan, ¹⁵Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, ¹⁶Università Univ. degli Studi di Salerno, Salerno, Italy, ¹⁷Inst CSS-Mendel, Rome, Italy, ¹⁸Ospedale 'Bambino Gesù, Rome, Italy, ¹⁹Policlinico Umberto I, Rome, Italy, ²⁰Texas Children s Hosp., Houston, TX, ²¹Baylor Col Med, T-803 Taub Building, Houston, TX, ²²Med. Genetics Unit, Univ. and Hosp. of Perugia, Perugia, Italy, ²³Univ. of Perugia Hosp., Perugia, Italy, ²⁴Meyer Children's Univ. Hosp., Florence, Italy, ²⁵Santa Anna Gen. Hosp., Como, Italy, ²⁶Univ. of Athens, Athens, Greece, ²⁷Clin Med Gen, Athens, Greece, ²⁸Univ. of Missouri-Kansas City, Kansas City, MO, ²⁹Manchester Ctr. for Genomic Med., MANCHESTER, United Kingdom, ³⁰Hôpital Necker, Inst. Imagine, Paris, France, ³¹CeGaT GmbH, Tuebingen, Germany, ³²Ospedale Casa Sollievo della Sofferenza, Rome, Italy, ³³IRCCS Casa Sollievo Della Soff, San Giovanni Ro, Italy, ³⁴Baylor Coll. of Med., Houston, TX

Abstract Body:

Craniofacial microsomia (CFM also known as Goldenhar syndrome), is a craniofacial developmental disorder of variable expressivity and severity with a recognizable set of abnormalities. These birth defects are associated with structures derived from the first and second branchial arches, can occur unilaterally and include ear dysplasia, microtia, preauricular tags and pits, facial asymmetry and other malformations. The inheritance pattern is controversial, and the molecular etiology of this syndrome is largely unknown. 670 patients belonging to unrelated pedigrees with European and Chinese ancestry with CFM, were investigated. We identified 18 likely pathogenic variants in 21 probands in FOXI3 (3.1%). Biochemical and cellular analyses in HEK-293T cells, and 4 different knock-in mouse data support the involvement of FOXI3 in CFM. Our findings indicate autosomal dominant inheritance with reduced penetrance, and/or autosomal recessive inheritance. The phenotypic expression of the FOXI3 variants is variable. The penetrance of the likely pathogenic variants in the seemingly dominant form is reduced, since a considerable number of such variants in affected individuals were inherited from non-affected parents. There is suggestive evidence that common variation in the FOXI3 allele in trans with the pathogenic variant could modify the phenotypic severity and penetrance.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2587. Fragile X related genes and *Dlg4* gene are involved in process formation in neuronal cells

Authors:

H. Shimizu, H. Hohjoh; Natl. Inst NeuroSci., Kodaira, Japan

Abstract Body:

Fragile X syndrome (FXS) is an inherited disorder caused by mutant *Fragile X mental retardation 1 (FMR1)* genes carrying abnormally expanded CGG triplet repeats: aberrant CGG repeats inhibit the expression of *FMR1*. The protein encoded by *FMR1* (FMRP) is an RNA binding protein and involved in the stabilization and translational regulation of mRNAs. FMRP is present in dendrites and dendritic spines of neurons and plays an important role in local translation of mRNAs. Previous studies suggested that FMRP was involved in neuronal development. The decrease in dendritic spines and the reduction of basal dendrite length and branching were observed in *Fmr1* knockout mice. Thus, FXS may have some adverse effects on neurite formation, but the mechanism is not yet fully understood. In this study, to examine the effects of *Fmr1* (FMRP) on process formation, we generated *Fmr1*-deficient cell lines with mouse neuroblastoma Neuro2a (N2a) cells: *Fmr1* was knocked out by gene editing using CRISPR-Cas9. Western blot indicated little or no expression of FMRP in *Fmr1*-deficient N2a cell lines. The *Fmr1*-deficient cells showed impairment of process formation under differentiation conditions. However, supplementation of retinoic acid restored process formation in *Fmr1*-deficient N2a cells. Since FMRP is known to be associated with the *DLG4* mRNA and FXR1 protein, we examined the effects of *Dlg4* and *Fxr1* on process formation. Gene silencing against *Dlg4* and *Fxr1* by RNAi was carried out in naïve N2a cells and *Fmr1*-deficient N2a cells. The knockdown cells exhibited impairment of process formation even under retinoic acid-containing differentiation conditions. In addition, the *Dlg4* mRNA could be detected by qRT-PCR, but the protein was hardly detected in N2a cells by western blotting. Gene silencing against *Dlg4* and *Fxr1* by RNAi in primary mouse hippocampal neurons was also performed. The knockdown neurons exhibited impairment of neurite outgrowth as in N2a cells. We further examined associations between FXR1 protein and *Dlg4* mRNA in *Fmr1*-deficient N2a cells by immunoprecipitation with FXR1 antibody. As a result, the association between *Fxr1* and *Dlg4* mRNA was observed. Taken together, our current findings suggest that the FXR1 protein and *Dlg4* mRNA may be involved in process formation independently of FMRP.

Molecular Effects of Genetic Variation Posters - Thursday

PB2588. Full penetrance of a hypomorphic TYR gene variation in oculocutaneous albinism.

Authors:

C. Velez¹, J. Vazquez², S. Carlo³, F. Velez-Bartolomei¹, A. Cornier⁴; ¹Genetic Diagnostic Group, San Juan, PR, ²Biology, Univ. of Puerto Rico Rio Piedras Campus, San Juan, PR, ³Ponce Hlth.Sci. Univ., Cabo Rojo, PR, ⁴San Jorge Children s Hosp./ Ponce Hlth.Sci. Univ., San Juan, PR

Abstract Body:

We are presenting the case of an 11-year-old male patient referred to genetics due to abnormal metabolic laboratory findings including deficient carnitine levels. Upon physical examination, mild joint hyperlaxity, decreased muscle tone and facial hypoplasia was ascertained. In addition to displaying lack of skin pigmentation, light colored hair and nystagmus. Therefore, an albinism gene panel was requested. Albinism gene panel identified two variants of potential clinical relevance in the *TYR* gene with an autosomal inheritance and heterozygosity for both variants. The *TYR* gene is bounded to the production of tyrosinase and melanin, which affects hair, skin and eyes pigmentation. Mutations in the *TYR* gene have been associated with oculocutaneous albinism (OCA) types IA and IB, and autosomal recessive Waarderburg syndrome type 2 (WS2-OA). The likely pathogenic variant sequence c.140G>A was identified in the patient with a prediction of missense mutation of Gly to Asp at codon 47 in exon 1 of the *TYR* gene. Whereas variant sequence c.1205G>A is classified as a reduced penetrant, hypomorphic pathogenic variant with a missense mutation of Arg to Gln at codon 402 in exon 4 of the *TYR* gene. The amino acid substitutions have been linked with OCA1 and OCA respectively, which is characterized by decreased or absent pigmentation in the hair, skin and eyes. Although reports indicate no clinically significant sequences were identified, and specify only a potential clinical relevance was shown, this patient does meet all the classical clinical criteria for a diagnosis of oculocutaneous albinism due to *TYR* gene mutations.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2589. Functional analysis of genetic variations found in clinical exomes of ASD patients.

Authors:

A. Bhattacharya¹, P. Parlanti¹, L. Cavallo¹, E. Farrow², T. M. Spivey², M. Manzini³; ¹Rutgers Univ. (Child Hlth.Inst.), New Brunswick, NJ, ²George Washington Univ., DC, DC, ³Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ

Abstract Body:

Autism Spectrum Disorder (ASD) and Intellectual Disability (ID) represent genetically heterogeneous neurodevelopmental disorders that are often comorbid and caused by single gene mutations of large effect. The involvement of multiple genes, each affecting a relatively few patients, complicates diagnosis. Genome-wide profiling studies produce an array of ‘variants of uncertain significance’ (mostly missense) which fail to provide clinicians with any specific cues for managing the disease. We sought to develop a framework to characterize these variations using Coiled-coil and C2 domains containing 1A (CC2D1A; MIM *610055, UniProtKB Q24K25), a candidate gene for ASD identified through family studies. We included four ASD/ID cases with compound heterozygous missense mutations [c.980C>T(p.Ser327Leu)/c.1345G>A(p.Val449Met), c.980C>T(p.Ser327Leu)/c.1322G>T(p.Gly441Val), c.956C>T(p.Pro319Leu)/c.2728G>A(p.Glu910Lys), c.1739C>T(p.Thr580Ile)/c.2657G>A,T(p.Arg886His)]. CC2D1A regulates the protein kinase A (PKA)/CREB pathway by repressing phosphodiesterase 4D (PDE4D), via an interaction mediated by its unique DM14 domains. We hypothesized that missense variants in the DM14 region disrupt binding and repression of PDE4D. When wildtype (WT) CC2D1A is overexpressed, PDE4D is repressed leading to an increase in PKA/CREB activity while the missense variants release repression, increasing cAMP degradation and reducing CREB activation. To test this, we cloned and overexpressed WT and mutant CC2D1A (individually and in patient combination) in HEK293 cells. Overexpression of mutant CC2D1A neither affected protein stability nor the survival of the transfected cells at 24 and 48hrs. By employing luciferase reporter assay, we quantified CREB activation 24hrs after transfection with 6hrs forskolin treatment. Compared to the WT, four DM variants G441V, V449M, P319L and T580I lead to blunted response to forskolin induced CREB activation (p=0.0313, N=6-7, Wilcoxon matched pairs signed rank test), suggesting these are most likely loss of function. S327L, V449M, P319L, T580I and R886H also have a concomitant reduction in cAMP levels (p<0.05, N=6, Wilcoxon matched-pairs signed rank test). Our work suggests mutations in the DM14 domain affect CC2D1A binding to PDE4D and disrupts signaling.

Molecular Effects of Genetic Variation Posters - Thursday

PB2590. Functional and *in silico* analyses of variants found in mitochondrial trifunctional protein deficiency patients

Authors:

E. Vieira Neto^{1,2}, M. Wang², A-W. Mohsen¹, Y. Wang¹, C. Van't Land¹, E. Koppes¹, T. S. Anthonymuthu¹, H. Bayir^{1,2}, J. Vockley^{1,2}; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²UPMC Children's Hosp. of Pittsburgh, Pittsburgh, PA

Abstract Body:

Background: Mitochondrial trifunctional protein (TFP) catalyzes the last three steps of the β -oxidation spiral for long-chain fatty acids as well as an unrelated step in cardiolipin metabolism. It is a heteromeric enzyme composed of two α and two β -subunits, encoded by *HADHA* and *HADHB* genes, respectively. More than 150 mutations in *HADHA* and *HADHB* are reported, typically resulting in complete TFP deficiency, while a common *HADHA* missense mutation (p.E510Q) usually leads to an isolated LCHAD deficiency. TFP/LCHAD deficiency shows a wide variety of clinical phenotypes that may range from severe, fatal neonatal to late onset with recurrent rhabdomyolysis. We previously showed defective cardiolipin remodeling and mitochondrial bioenergetics in this disorder. Our objective here is to address its genetic heterogeneity correlating it with cardiolipin profile and bioenergetics function. Methods: Genotype was confirmed by sequencing of DNA from skin fibroblasts. α and β -TFP proteins were analyzed by western blot (WB). Reverse-transcription droplet digital PCR (RT-ddPCR) was used to evaluate mRNA expression. Structure analysis of the variant proteins was performed by the Insight II (2000) software package using TFP atomic coordinates. Mitochondrial bioenergetics were measured with a Seahorse Analyzer. Cardiolipins (CLs) and monolyso-CLs were quantified by LC-MS/MS. The databases HGMD, ClinVar, and gnomAD were consulted. The *in silico* predictive tools PROVEAN, SIFT, PolyPhen-2, and Mutation Taster were applied. Results: WB of cells with truncating variants in *HADHA* or *HADHB* revealed significant decrease in levels of both subunits. Genotypes comprising *HADHA* or *HADHB* missense mutations had protein levels comparable to controls. Truncating variants had variable effects on mRNA levels. Structure analysis of the variant proteins predicted disruption of catalytic sites, substrate binding and destabilization of subunit interaction. Mitochondria from cells with genotypes homozygous for p.E510Q or having truncating variants in *HADHA* and *HADHB* presented increased levels of oxidized CLs and/or monolyso-CLs. These same cells presented mitochondrial dysfunction reflected in clear reduction of maximal respiration and spare respiratory capacity. Conclusion: Genotypes with truncating mutations were the most affected in terms of protein expression, but their impact on mRNA levels varied. These genotypes plus homozygous p.E510Q were also very deleterious for mitochondrial function and cardiolipin metabolism. There was a high level of concordance between the results of these functional analyses, and *in silico* structure analysis, and predictive tools.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2591*. Functional characterization of a heat shock regulated enhancer at coronary artery disease associated locus that controls *CALCRL* expression in endothelial cells

Authors:

I. Selvarajan¹, J. Li², P. Pölönen³, T. Örd¹, K. Öunap¹, K. Mäklin¹, A. Toropainen¹, L. Stolze⁴, C. ROMANOSKI⁴, M. Heinäniemi³, Y. Fang², M. Kaikkonen¹; ¹A. I. Virtanen Inst. for Molecular Sci., Univ. of Eastern Finland, Kuopio, Finland, ²Dept. of Med., The Univ. of Chicago, Chicago, IL, ³Inst. of Biomedicine, Univ. of Eastern Finland, Kuopio, Finland, ⁴The Univ. of Arizona, Tucson, AZ

Abstract Body:

Coronary artery disease (CAD) is one of the major causes of mortality worldwide. Recent genome-wide association studies have started to unravel the genetic architecture of the disease. Such efforts have identified Calcitonin receptor-like (*CALCRL*), an important mediator of the endothelial fluid shear stress response, associated with CAD risk variants. Better understanding of *CALCRL* gene regulation and the role of SNPs in modulation of *CALCRL* activity could provide important steps towards understanding genetic predisposition in shear stress. In this study we functionally characterized the non-coding regulatory elements carrying CAD risks SNPs and studied their role in the regulation of *CALCRL* expression in endothelial cells. We demonstrated that rs880890-harboring regulatory element exhibits high enhancer activity and significant allelic bias with A allele showing 40% more activity than G allele. We also observed that under shear stress the A allele of rs880890 is favored over the G allele. CRISPR deletion of rs880890-enhancer resulted in downregulation of *CALCRL* expression. EMSA showed that heat shock factors are binding to the enhancer with a preference to A allele compared to G allele. Interestingly, HSF1 knockdown under shear stress showed a significant decrease in *CALCRL* expression. RNA-Seq of *CALCRL* knockdown and enhancer deletion results confirmed the role of *CALCRL* in eNOS, apelin, adrenomedullin and angiotensin signaling pathways while Real-Time Cell Analyzer (RTCA) based assay further showed decrease in cell proliferation and tube formation. Overall, our results demonstrate the existence of an endothelial-specific heat shock factor regulated transcriptional enhancer carrying a CAD risk SNP rs880890 that regulates *CALCRL* expression.

Molecular Effects of Genetic Variation Posters - Thursday

PB2592. Functional characterization of new non-coding pathogenic variants creating upstream Open Reading Frames in the 5'UTR of the Endoglin gene and causing Hereditary Hemorrhagic Telangiectasia.

Authors:

O. Soukarieh¹, E. Tillet², C. Proust¹, F. Soubrier³, A. Goyenville⁴, M. Eyries³, D-A. Trégouët¹; ¹INSERM UMR 1219, Bordeaux Population Hlth.Res. Ctr., Université de Bordeaux, Bordeaux, France, ²INSERM UMR U1292, laboratoire BIOSANTE, Université Grenoble Alpes, Inserm, CEA, Grenoble, France, ³Département de génétique, Hôpital Pitié-Salpêtrière, Sorbonne Université, Assistance Publique-Hôpitaux de Paris, Paris, France, ⁴Université Paris-Saclay, UVSQ, Inserm, END-ICAP, Versailles, France

Abstract Body:

Hereditary Hemorrhagic Telangiectasia (HHT) is a rare vascular disorder causing abnormal vessel formation and characterized by autosomal dominant transmission and considerable variability in symptoms and clinical severity. In clinical diagnosis, 3 main genes, *ALK1*, *ENG* and *SMAD4* are routinely screened for pathogenic variants at the origin of HHT. About 80% of HHT cases are caused by pathogenic coding mutations in *ALK1* and *ENG*. However, at least 15% remain with no molecular explanations in the 3 main genes.

We here report the identification of 2 new variants, c.-79C>T and c.-68G>A, in the 5'UTR of *ENG* in 2 unrelated HHT patients. These 2 variants are predicted to create Open Reading Frames located upstream (upORF) of the CoDing Sequence (CDS). upORF are naturally present in human transcripts and result from the presence of an upstream translation initiation site (uTIS) located within the 5'UTR. The association of an uTIS with a stop codon located within the 5'UTR leads to a fully upstream ORF (uORF). Elongated CDS (eCDS) and overlapping ORF (oORF) result from uTIS being in-frame and out-of-frame with the CDS, respectively.

Interestingly, 4 additional upORF-creating variants in the 5'UTR of *ENG* have been previously reported in HHT patients (c.-142A>T, c.-127C>T, c.-10C>T and c.-9G>A), without extensive functional investigations of their pathogenicity. To assess the potential effect of all of these 6 variants (5 oORF- and 1 eCDS-creating variants) on *ENG*, we performed *in vitro* functional assays based on the expression of wild-type and mutant constructs in human cells. We found that all were associated with a decrease of protein levels in HeLa and HUVECs cell lines. A comparison of our experimental results with patients' clinical characteristics suggests that oORF-creating variants leading to *ENG* protein levels less than 40% *in vitro* would be associated with a severe form of HHT whereas eCDS-creating ones, at the origin of *ENG* levels around 80% would lead to a mild form of HHT. Additional experiments relying on artificial deletions in our mutated constructs revealed that some, but not all, created upORF could be translated into small encoded peptides. We are now using a dedicated functional assay to explore the effect of these 6 *ENG* variants on the signaling activity of ALK1 in response to BMP9, a mandatory element before providing a definitive molecular diagnosis.

Overall, we here identified two never reported 5'UTR *ENG* variations in HHT patients and shed new lights on the role of upORF on *ENG* regulation. Our findings contribute to ameliorate molecular diagnosis in HHT.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2593. Functional Characterization of Novel UBE3A Variants Associated with Angelman Syndrome

Authors:

M. Cousin^{1,2}, G. E. Boyum^{1,3,4}, T. L. Schwab^{1,3}, C. T. Schmitz^{1,3}, A. Jain⁵, E. C. Wirrell⁶, J-M. Tillema⁶, D. Babovic-Vuksanovic^{1,7}, B. C. Lanpher^{1,7}, R. H. Gavrilo^{1,6,7}, K. J. Clark^{1,3}, E. W. Klee^{1,2,7}; ¹Ctr. for Individualized Med., Mayo Clinic, Rochester, MN, ²Dept. of Quantitative Hlth.Sci., Mayo Clinic, Rochester, MN, ³Dept. of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN, ⁴Univ. of Wisconsin-Madison Genetics Training Program, Madison, WI, ⁵Mayo Clinic Graduate Sch. of BioMed. Sci., Jacksonville, FL, ⁶Dept. of Neurology, Mayo Clinic, Rochester, MN, ⁷Dept. of Clinical Genomics, Mayo Clinic, Rochester, MN

Abstract Body:

Angelman syndrome (AS, MIM: 105830) is a rare neurodevelopmental disorder characterized by developmental delay or intellectual disability, speech impairment, gait ataxia, limb tremor, microcephaly, seizures, and a happy demeanor with frequent laughing, smiling, and excitability. AS is caused by alterations of the 15q11-13 chromosomal region with a subset caused by pathogenic variants within the paternally imprinted *UBE3A* (MIM: 601623, NM_130838.2) gene. Mechanisms of *UBE3A* variant pathogenesis include loss of the nuclear isoform or abnormal catalytic activity by the encoded E6-associated protein (E6AP) ubiquitin-protein ligase. We identified and functionally characterized novel *UBE3A* variants in 3 unrelated individuals with a neurodevelopmental disorder consistent with AS. Proband 1 has a maternally inherited c.91G>A; p.(Glu11Lys) variant which was not found in the maternal grandmother. Proband 1 is female with a history of an abnormal brain MRI, intractable epilepsy, absent speech, a hemiparetic gait disorder, and a happy demeanor. Proband 2 has a maternally inherited c.1772T>C; p.(Leu591Pro) variant occurring de novo in the mother. Proband 2 is male and has global developmental delay, language delay, abnormal gait with uplifted arms flexed at the elbow, frequent smiling, history of GI reflux, sleep disturbance, and a light complexion. Proband 3 has a de novo c.2360C>T; p.(Thr787Met) variant. Proband 3 is female with a clinical diagnosis of Landau-Kleffner syndrome with intractable seizures, development and speech delays, language regression, auditory agnosia, abnormal movement, sleep disturbance, and hyperactivity. Two additional entries for this variant have since been deposited in ClinVar as a VUS associated with AS. All variants are extremely rare in the population (gnomAD) and affect highly conserved residues. The p.(Glu11Lys) lies in the AZUL domain and the p.(Leu591Pro) and p.(Thr787Met) fall in the N-lobe and C-lobe, respectively, of the HECT domain. By expressing GFP-tagged wild-type or variant *UBE3A* in HEK293 cells, we demonstrate that p.(Glu11Lys) and p.(Leu591Pro) variants are associated with almost complete loss of nuclear localization and the p.(Thr787Met) variant with reduced nuclear localization. Using ligase-dead in comparison to ligase-active constructs we show p.(Glu11Lys) has no effect on protein stability, where p.(Leu591Pro) was unstable and p.(Thr787Met) had reduced protein with ligase-active construct only suggesting ligase hyperactivity. This study demonstrates the functional characterization of these novel *UBE3A* variants confirming they are disease causal.

Molecular Effects of Genetic Variation Posters - Thursday

PB2594. Functional characterization of polycystic ovary syndrome-associated risk loci identifies genetic regulatory regions

Authors:

L. Sankaranarayanan¹, G. Johnson¹, A. Barrera¹, A. Dunai², T. Reddy¹; ¹Duke Univ., Durham, NC, ²Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Polycystic ovary syndrome (PCOS) is a leading cause of infertility and type 2 diabetes and affects 10-15% of menstruating people worldwide. PCOS is characterized by increased circulating testosterone levels, ovulatory dysfunction and increased insulin resistance. Recent genome wide association studies (GWAS) have identified ~20 PCOS susceptibility loci in European and Han Chinese cohorts. The identified loci are implicated in neuroendocrine, reproductive, and metabolic function. The associated variants are predominantly non-coding. Rare non-coding variants in the *DENNDIA* locus have also been found in ~50% of families with PCOS by whole genome sequencing. Together, these associations suggest non-coding genetic contributions to PCOS.

We tested the hypothesis that PCOS-associated non-coding genetic variants contribute to PCOS pathogenesis by altering gene regulation. We first identified ~1400 functioning regulatory elements across the PCOS-associated loci using a high throughput reporter assay, STARR-seq, in a cell line model of testosterone production. Adding to evidence that the identified regulatory regions have function, ~60% of the identified regions are in previously documented open chromatin regions from DNase-seq and ATAC-seq studies.

To then investigate the effects of noncoding genetic variants in these regulatory regions, we tested for allele specific regulatory activity in a multi-ethnic pool of five genomes from healthy individuals. *DENNDIA* codes for a protein of the connectin family functioning as guanine nucleotide exchange factors for the GTPase Rab35. *DENNDIA* has also been identified as a regulator of androgen biosynthesis in ovarian cells. We identified 35 genetic variants that alter regulatory activity in the *DENNDIA* locus, including one that is ancestry specific. About half of the identified variants reside in previously documented open chromatin regions. Four of those genetic variants (rs12237685, rs28441318, rs10120705 and rs10117455) are also eQTLs for *DENNDIA* from the GTEx project and reside within open chromatin regions or annotated as ENCODE candidate cis-regulatory elements.

Taken together, these studies suggest a new mechanism for PCOS where rare and common genetic variants cause altered gene expression, including altered *DENNDIA* expression; and in which some of those variants may contribute in an ancestry-specific genetic risk for PCOS.

Molecular Effects of Genetic Variation Posters - Thursday

PB2595. Functional impact of rare variants and sex across the x chromosome and autosomes

Authors:

R. Ungar¹, T. Li², N. Ersaro¹, N. Gates Vetr¹, A. Battle³, S. Montgomery⁴; ¹Stanford, Stanford, CA, ²Johns Hopkins Sch. of Med., Baltimore, MD, ³Johns Hopkins Univ., Baltimore, MD, ⁴Stanford Univ., Stanford, CA

Abstract Body:

Sex differences are widespread in humans, 37% of genes are sex biased in at least one tissue across the autosomes and X-chromosome (Oliva et al. Science. 2020). Yet many studies fail to analyze sex as a biological variable, and the X-chromosome is largely ignored; 67% of genome-wide association studies exclude the X-chromosome (Wise et al. AJHG 2013). Furthermore, these studies typically consider only common variants excluding rare variants despite their abundance and potential for larger effect sizes. We have hypothesized that understanding how rare variants functionally operate on the X and in a sex-stratified manner across the X and autosomes can provide novel insight into the biological underpinnings of sex differences in complex traits and diseases.

We use the GTEx resource of population-scale, multi-tissue gene expression to extend analyses of rare variant associated gene expression outliers to identify effects present on the X or masked by sex. We identified 109 multi-tissue gene expression outliers on the X-chromosome. We further identified 27 multi-tissue outliers that were only detectable at $abs(Z) > 2.5$ in one of the sexes and masked in a joint sex analysis. Per individual, we observed approximately 0.20 outliers per individual on the X while autosomal chromosomes range from 0.09 to 0.66 outliers. Males averaged 0.05 more outliers on the X than females and were notably more enriched for outlier-associated rare variants compared to females on the X. Sex masking of outlier effects also highlighted specific genes. We identified an individual with a rare variant at a frequency of $1.5e-5$ in the gene ADGRD2 that has a z-score of 1.6 in a combined-sex distribution but a z-score of 3.1 when stratifying to a female-only distribution. ADGRD2 has been implicated to predict uterine corpus endometrial carcinoma survival (Lei et al. International Immunopharmacology 2022), and truncating mutations cause congenital bilateral absence of vas deferens (Patat et al. AJHG 2016). We leveraged both observed outlier status and genomic annotations to prioritize 124,325 rare variants on the X-chromosome using a Bayesian hierarchical model. When assessing sex-specific distribution of rare variant posteriors we found CD99L2, which reduces inflammatory response in mice knockouts (Rutledge et al. Experimental & Molecular Pathology 2015), showed higher splicing posteriors in females. Our study demonstrates that while rare variants are by definition difficult to individually study, providing detailed generalizations of effects in a sex-specific manner is an important step in understanding biological underpinnings of sex differences and their role in personalized genomics.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2596. Gene discovery and pathogenic mechanisms of disease in worms, flies, and fish: The Model Organisms Screening Center for the Undiagnosed Diseases Network

Authors:

M. Wangler¹, D. Baldrige², H. Bellen¹, A. Bowman², S. Pak², J. Postlethwait³, T. Schedl⁴, L. Solnica-Krezel², M. Westerfield³, S. Yamamoto⁵; ¹Baylor Coll. of Med., Houston, TX, ²Washington Univ. in St Louis, St Louis, MO, ³Univ. of Oregon, Eugene, OR, ⁴Washington Univ. Sch. of Med., St Louis, MO, ⁵Baylor Coll. of Med., HOUSTON, TX

Abstract Body:

Model organisms allow for functional studies of human genetic variants that are candidates for disease while also providing avenues to explore in depth pathogenic mechanisms. The Undiagnosed Diseases Network (UDN) has incorporated a Model Organisms Screening Center (MOSC), and for the past four years the MOSC has conducted studies on UDN cases in Phase II of the project. Undiagnosed human subjects undergo sequencing and variant interpretation in the UDN, and candidate variants are nominated for the MOSC for further evaluation. The MOSC human genetics team initially evaluates each variant using available gene- and variant-level human genetic data, after which, clinicians nominating the UDN cases interface with model organism investigators at monthly video conferences. Subsequently, variants are experimentally studied by experts in *C. elegans*, *Drosophila melanogaster* and Zebrafish. Over the course of Phase II, 243 variants in 173 genes from 158 human subjects have been nominated for MOSC studies. Laboratory based model organism studies in Worm, Fly and Fish have been or are being conducted for 26, 55 and 35 variants, respectively (48% of nominated variants in total). MOSC studies to date in Phase II have yielded data for 33 genes in favor of variant causality across all three organisms and for 24 genes no evidence for variant causality after thorough study. These studies also allow for in depth mechanistic studies; for example, *Drosophila* studies of a novel *MRTFB de novo* variant determine this gene's role in neurodevelopmental disease and actin binding. These MOSC studies have led to additional new disease gene discoveries from Phase II, including *CDK19*, *BICRA*, *TOMM70*, *RAB5B*, and *TNPO2*. This model organism work ultimately aids in diagnosing otherwise undiagnosed patients by providing support for pathogenicity of candidate disease-causing variants. In summary, the MOSC is a uniquely collaborative effort which brings clinicians seeing undiagnosed patients into collaboration with basic scientists in model system labs. In addition, model organism studies of newly uncovered rare disease provide opportunities to exploring molecular mechanisms and future therapeutic approaches.

Molecular Effects of Genetic Variation Posters - Thursday

PB2597. Gene expression profiling with direct long-read RNA sequencing uncovers functional variation affecting transcripts production.

Authors:

A. Real¹, C. Borel², N. M. R. Lykoskoufis¹, G. PugaYung^{1,3}, J. D. Seebach^{1,3}, A. Brown⁴, E. T. Dermitzakis¹, A. Viñuela⁵, A. Ramisch¹; ¹Univ. of Geneva, Geneva, Switzerland, ²Univ. of Geneva, Geneva 4, Switzerland, ³Univ. Hosp. Geneva (HUG), Geneva, Switzerland, ⁴Univ. of Dundee, Dundee, United Kingdom, ⁵Newcastle Univ., Newcastle, United Kingdom

Abstract Body:

Expression quantitative trait loci (eQTLs) studies have uncovered thousands of genetic effects acting on gene expression. However, the current short-read RNA-seq technologies do not characterize transcripts in their full-length form, leaving us incomplete information to understand the molecular regulatory mechanisms of eQTLs. In this study, we produced long-read native poly(A) RNA-seq data using the Oxford Nanopore Technologies (ONT) platform for 60 genetically different lymphoblastoid cell lines (LCLs) from the 1000 Genomes/Geuvadis project. We identified 11,154 protein-coding genes and lincRNAs expressed in at least 50% of the samples, for which we see a good agreement with published gene expression quantifications of the same samples based on Illumina short-read sequencing (pairwise sample correlation 0.61-0.79). We identified 48,539 transcripts using the FLAIR pipeline (Tang, 2020), of which 39% are already annotated based on GENCODE version 19. While 35% of these annotated transcripts were expressed in each of the 60 LCL samples, this was true only for 7% of unannotated ones. A genome-wide QTL analysis on the 14,447 annotated transcripts expressed in at least 50% of samples identified 72 transcript QTLs (trQTLs) (FDR 5%) of which 71 are novel compared to the published list of eQTLs from the larger Illumina dataset (317 samples). Based on the gene-level quantifications from our long-read data, we detected 45 eQTLs of which 12 were also trQTLs. We observed that eQTL genes had a significantly lower number of annotated transcripts than genes with trQTLs (Wilcoxon test p -value = $5.23e-05$), suggesting that genetic effects on genes with higher transcript diversity were missed using gene-level quantifications. Overall, we were able to identify new trQTLs with a small number of samples, whose effect on expression were missed when using short read technology. Hence, we could show that by using long-read data to analyze transcriptomes, we are one step closer to understand the key role of genetic variants on transcription abundance and structure, which might also improve the characterization of disease-associated variants.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2598. Gene prioritization for rare diseases integrating genotype, RNA-seq and phenotype - lessons from a CAGI 6 challenger team

Authors:

V. Yépez¹, C. Mertes², N. H. Smith³, I. F. Scheller¹, J. Gagneur¹; ¹Technical Univ. of Munich, Garching bei München, Germany, ²Technical Univ. of Munich, Garching, Germany, ³Technical Univ. of Munich, München, Germany

Abstract Body:

RNA sequencing has emerged as a complementary tool to DNA sequencing for rare disease diagnostics. However, gene prioritization methods integrating genotype, RNA-seq and phenotype have been lacking. To address this need, the SickKids Genome Clinic released a CAGI6 diagnostics challenge and provided nearly 80 genomes and RNA-seq samples [1] from children suffering from a rare genetic disorder. We developed a gene prioritization model integrating mono-allelic expression, gene expression [2] and splicing outliers [3] extracted through the workflow DROP [4] together with variant annotations and HPO-encoded phenotypes. For the samples without genetic data, we called the variants from the RNA-seq data using GATK. The model is a gradient boosting machine implemented using XGboost trained on a cohort of 209 mitochondrial disease patients [5] from which half are diagnosed. Our model prioritizes the causal gene first for almost half of the diagnosed cases, and among the top 5 in more than 70% of them. Application to the CAGI6 SickKids cohort revealed a known splice-disrupting pathogenic variant and reported several promising candidates. Our approach and publicly available software [6] can help find and prioritize candidate genes found by DNA and RNA sequencing and can be especially useful to reduce the burden of manual inspection in cohorts of hundreds of samples.

References

1. <http://genomeinterpretation.org/cagi6-sickkids.html>
2. Brechtmann et al, AJHG (2018)
3. Mertes et al, Nat Commun (2021)
4. Yépez et al, Nat Protoc (2021)
5. Yépez, Gusic et al, Genome Med (2022)
6. https://github.com/gagneurlab/cagi6_sickkids

Molecular Effects of Genetic Variation Posters - Thursday

PB2599. Genetic control of mRNA splicing as a potential driver for incomplete penetrance of coding variants

Authors:

J. Einson¹, D. Glino², S. Castel², P. Mohammadi³, T. Lappalainen⁴; ¹Columbia Univ., New York, NY, ²New York Genome Ctr., New York, NY, ³Scripps Res., La Jolla, CA, ⁴KTH Royal Inst. of Technology & NY Genome Cntr, New York, NY

Abstract Body:

Alternative splicing is a fundamental molecular process that can be influenced by common genetic variants across individuals and populations. In this study, we propose a model where common splicing quantitative trait loci (ψ QTLs) affect the dosage of rare coding variants (cSNPs) in their target exons, which consequently reduces the cSNP's penetrance. We hypothesize that haplotype configurations undergo selective pressure against those that increase penetrance in the general population, and that this can be inferred from the perspective of purifying selection. This model may explain some instances of incomplete penetrance, and could ultimately help predict the clinical severity of carrying a putatively damaging variant. First, we calculated exon PSI scores from bulk RNA-seq data across 18 GTEx tissues. We found that PSI Z-scores were significantly different between exons carrying deleterious (N = 19,178) and non-deleterious (N = 49,575) rare variants (Mann-Whitney U-Test: $p = 2.577e-4$), with exons carrying deleterious rare variants spliced in less often, on average. To test if this result was potentially driven by regulatory variants, we cataloged ψ QTLs (N = 5,196 across 18 tissues), which use exon PSI as a molecular phenotype. We developed an enrichment test for haplotype combinations that takes into account ψ QTL allele frequencies. The test outputs " ϵ ", where the sign represents if high penetrance haplotypes are depleted or enriched in a dataset, and an associated p-value. We first applied our test to phased WGS data in GTEx V8, and observed a depletion of putative high-penetrance haplotypes ($\epsilon = -0.0156$, $p = 1.006e-6$). When splitting the data into groups based on synonymous vs. non-synonymous status of the cSNP, we did not observe a significant difference in depletion between for high-penetrance configurations. We next tested our model in donors from the Simons Simplex Collection (SSC), to investigate if high penetrance haplotypes are associated with autism spectrum disorder (ASD). Across ASD associated genes in probands, deleterious high-penetrance haplotypes were depleted ($\epsilon = -0.055$, B.P. $p = 0.046$). The result was similar in parents, but did not reach significance ($\epsilon = -0.028$, B.P. $p = 0.11$). This indicates that splicing modifiers may provide some protection in families with evidence of heritable autism risk variants. Overall, we have demonstrated a relationship between common splice-regulatory variants and rare pathogenic cSNPs, two typically separate areas of study.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2600*. Genetic determinants of blood transcript splicing and impact on molecular phenotypes in 4732 healthy individuals

Authors:

A. Tokolyi¹, E. Persyn², K. Burnham¹, A. Nath^{2,3}, J. Marten², INTERVAL Study, D. Roberts⁴, E. Di Angelantonio², J. Danesh², A. Butterworth², M. Inouye^{2,3}, D. Paul^{2,5}, E. Davenport¹; ¹Wellcome Sanger Inst., Hinxton, United Kingdom, ²Univ. of Cambridge, Cambridge, United Kingdom, ³Baker Heart and Diabetes Inst., Melbourne, Australia, ⁴Univ. of Oxford, Oxford, United Kingdom, ⁵AstraZeneca, Cambridge, United Kingdom

Abstract Body:

Untangling the pathways by which genetic variants modulate molecular phenotypes and disease risk requires comprehensive integrated analyses in large, deeply phenotyped cohorts of individuals. Splicing quantitative trait loci (sQTLs) are major contributors to complex traits, with a similar contribution as those affecting gene expression levels. However, the mechanisms that link these variants to downstream molecular and disease phenotypes through transcript splicing remain to be explored. INTERVAL is a deeply phenotyped cohort of healthy blood donors across England, including 4,732 individuals with matched genotypes and whole-blood RNA-seq, as well as proteomic, lipidomic, and metabolic measurements.

Here we utilize the split reads present in RNA-seq to describe the genetic architecture of splicing in whole blood, yielding 31,319 sQTLs (at FDR<0.05) across 8,799 genes. Comparing and colocalizing these sQTLs to expression QTLs (eQTLs) mapped in the same individuals reveals both commonalities and differences, with 9422 splice events (in 47.9% of tested genes) possessing a colocalizing eQTL. Splicing also appears to share genetic regulation with plasma protein QTLs (pQTLs), assayed by SOMAscan, with 544 splice events having genetic signals colocalizing with pQTLs (in 50.3% of 346 proteins) after conditional association analysis. These colocalizations demonstrate a genetic basis for known splicing differences impacting the plasma levels of proteins, such as the solubility of FAS, as well as revealing many other splicing events through which genetic variants may impact downstream plasma protein levels. Subsequent colocalization of sQTLs and pQTLs with COVID-19 HGI summary statistics recapitulates established associations of splicing and protein levels with disease risk in *OAS1*, as well as providing novel splicing associations in risk genes such as *IFNAR2* and *OAS3*.

This largest to-date splicing QTL catalog and associated preliminary analyses will be a useful resource for future study of the genetic architecture of transcript splicing and the subsequent impact on molecular phenotypes and disease risk.

Molecular Effects of Genetic Variation Posters - Thursday

PB2601. Genetic determination of fetal hemoglobin levels in patients with sickle cell disease

Authors:

O. Ojewunmi¹, T. Adeyemo^{2,3}, A. Oyetunji⁴, A. Akinrodoye⁴, H. Patel¹, H. Rooks¹, C. Grace⁵, A. Goel⁶, H. Ahmad⁷, O. Nnodu⁸, J. Strouboulis¹, S. Menzel¹; ¹King's Coll. London, London, United Kingdom, ²Univ. of Lagos, Lagos, Nigeria, ³Lagos Univ. Teaching Hosp., Lagos, Nigeria, ⁴Sickle Cell Fndn. Nigeria, Lagos, Nigeria, ⁵Oxford Univ., Oxford, Oxfordshire, United Kingdom, ⁶Univ Oxford, Oxford, Oxfordshire, United Kingdom, ⁷Ahmadu Bello Univ. Teaching Hosp., Zaria, Nigeria, ⁸Univ. of Abuja, Abuja, Nigeria

Abstract Body:

Sickle cell disease is an inherited blood disorder caused by a point mutation in the β -globin. It is a major cause of childhood disability and mortality in sub-Saharan Africa. A major modifier of sickle cell disease (SCD) is fetal hemoglobin (HbF). HbF ($\alpha_2\gamma_2$) is heritable and persistent HbF ameliorates the disease severity and improves survival. Previously, genetic polymorphisms in three loci, *BCL11A*, *HMIP*, and β -globin like cluster have been associated with high HbF in healthy persons and in individuals with SCD. Nigeria has the largest population of SCD patients world-wide, enabling the assembly of powerful study cohorts to discover new disease modifier loci.

In this study, we recruited 1,117 SCD patients (HbSS/HbS β^0) comprising three major ethnic groups in Nigeria to conduct the first genome-wide association study of HbF in Nigeria.

Our aim was to identify new loci and perform fine-mapping to discover causal variants in persons with SCD. Genotyping was performed using H3Africa Illumina chip array and imputation using 1000 Genomes Phase 5 version 3. Association with HbF was analyzed using a linear regression model assuming an additive allele effect (adjusting for age, sex, and population structure).

Our preliminary results of 379 patients confirmed *BCL11A* rs7606173 (EAF: effect allele frequency = 0.43, $\beta = -0.461$, $p = 7.19 \times 10^{-15}$) showing the strongest association with HbF. We also identified two novel *BCL11A* SNPs: rs62142644 (EAF = 0.35, $\beta = -0.429$, $p = 9.9 \times 10^{-14}$) and rs72962585 (EAF = 0.21, $\beta = -0.378$, $p = 2.43 \times 10^{-8}$). In addition, we found a novel locus on chromosome 9: rs13284968 (EAF = 0.25, $\beta = -0.345$, $p = 2.54 \times 10^{-8}$). Regional association plot revealed two credible SNPs (rs6706648 and rs62142644) in linkage disequilibrium with rs7606173 (R^2 : 0.911 and 0.738 respectively). Stepwise conditional regression analysis on rs7606173 indicated rs45595933 ($p = 1.07 \times 10^{-4}$) as an independent signal. A third conditional analysis using rs7606173 and rs45595933 identified rs1427407 ($p = 0.0024$). The HbF-boosting genotypes of rs7606173 (GG) and rs1427407 (TT) are always inherited together, but they have independent genetic effect on HbF levels. Providing further insight into the contribution of rs1427407 (T/G), rs7606173 (G/C), and rs45595933 (G/A) to the *BCL11A* HbF association signal, haplotype analysis showed that haplotype TGA was associated with the highest HbF ($\beta = 0.391$, 95%CI: 0.16 - 0.62, $p = 0.001$). Functional Genome-wide Complex Trait Analysis suggests functional role for rs7606173 as it is contained within enhancers characterized by histone modification patterns in HUDEP-2 cells. This region is enriched with sites for GATA-1, POLR2A, and TAL-1 binding.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2602. Genetic effects on chromatin accessibility and nucleosome positioning using snATAC data across 284 skeletal muscle biopsies

Authors:

X. Wang¹, C. C. Robertson^{1,2}, A. Varshney¹, N. Manickam¹, P. Orchard¹, M. Laakso^{3,4}, J. Tuomilehto⁵, T. A. Lakka³, K. L. Mohlke⁶, M. Boehnke¹, L. Scott¹, H. A. Koistinen⁵, F. Collins², S. C. J. Parker¹; ¹Univ. of Michigan, Ann Arbor, MI, ²NIH, Bethesda, MD, ³Univ of Eastern Finland, Kuopio, Finland, ⁴Kuopio Univ Hosp., Kuopio, Finland, ⁵Finnish Inst for Hlth.and Welfare, Helsinki, Finland, ⁶Univ North Carolina, Chapel Hill, NC

Abstract Body:

Insulin resistance in skeletal muscle is a central process contributing to type 2 diabetes (T2D) and related metabolic traits. Among 13 cell types we have identified in snRNA and snATAC data from skeletal muscle, type 1 muscle fiber is the most abundant. Studying genetic variants that affect chromatin accessibility and nucleosome positioning in type 1 muscle fiber can provide mechanistic insights at GWAS signals for T2D and related trait, and allow for causal inference of local chromatin dynamics (i.e. which transcription factors (TFs) causally position nucleosomes?). We profiled the chromatin landscape using snATAC of muscle from 284 individuals and focused our analyses on type 1 fibers. Using a hidden markov model (HMMRATAC), we inferred nucleosome free regions (NFR, N=44,763 peaks), which represent TF-occupied regions, and their adjacent nucleosomal chromatin regions (N=89,526 peaks, two per NFR), which represent nucleosomal phasing that may be induced by a primary TF-binding event. For each of the 284 samples, we used Multi-Otsu statistical thresholding on the observed ATAC-seq fragment length distribution to classify fragments as originating from NFR or nucleosomal regions. Using QTLtools, we performed a cis-QTL analysis (cis window of 100kb) on both chromatin accessibility (caQTL) and nucleosome positioning (nucQTL). We identified 10,203 caQTLs and 5,292 nucQTLs at 5% FDR. The same lead variant was shared across 968 caQTL and nucQTL signals. For 2,543 out of 5,292 (48%) nucQTLs, the lead variant was in strong linkage disequilibrium ($r^2 > 0.8$) with a caQTL lead variant. The remaining 2,749 (52%) nucQTLs may represent independent nucleosomal signals that would not be identified in a standard QTL scan with ATAC-seq data. We are now performing formal colocalization analyses of the overlapping caQTL and nucQTL signals and T2D and related trait GWAS signals, followed by Mendelian Randomization, to infer causal directional effects. Motif enrichment analyses may reveal specific TFs that causally affect nucleosome positioning. Overall, these analyses expand our understanding of regulatory mechanisms in skeletal muscle underlying T2D and related trait genetic risk.

Molecular Effects of Genetic Variation Posters - Thursday

PB2603. Genetic inactivation of zinc transporter SLC39A5 improves liver function and hyperglycemia in obesogenic settings

Authors:

S. Chim¹, K. Howell¹, J. Dronzek¹, W. Weizhen², C. Van Hout¹, M. Ferreira¹, B. Ye¹, A. Li¹, S. Brydges², V. Arunachalam¹, A. Marcketta¹, A. Locke¹, J. Bovijn¹, N. Verweij¹, T. De¹, L. Lotta¹, L. Mitnau¹, M. LeBlanc¹, Regeneron Genetics Center, GHS-RGC DiscovEHR Collaboration, D. Carey³, O. Melander⁴, A. Shuldiner¹, H. Nistala¹, K. Karalis¹, A. Economides¹; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Regeneron Pharmaceuticals, Tarrytown, NY, ³Geisinger Hlth.System, Danville, PA, ⁴Lund Univ., Malmo, Sweden

Abstract Body:

Recent studies have revealed a role for zinc in insulin secretion and glucose homeostasis. Randomized placebo-controlled zinc supplementation trials have demonstrated improved glycemic traits in patients with type II diabetes (T2D). Moreover, rare loss-of-function variants in the zinc efflux transporter *SLC30A8* reduce T2D risk. Despite this accumulated evidence, mechanistic understanding of how zinc influences systemic glucose homeostasis and consequently T2D risk remains unclear. We explored the relationship between zinc and metabolic traits by searching the exome database of the Regeneron Genetics Center-Geisinger Health System DiscovEHR cohort for genes that regulate zinc levels and associate with changes in metabolic traits. We previously reported that rare loss-of-function (LOF) variants (MAF<1%) in Solute Carrier Family 39, Member 5 (*SLC39A5*) associated with increased circulating zinc ($p=4.9 \times 10^{-4}$). Trans-ancestry meta-analysis across four studies exhibited nominal association of *SLC39A5* LOF variants with decreased T2D risk. To explore the mechanisms underlying these associations, we generated mice lacking *Slc39a5*. *Slc39a5*^{-/-} mice display improved liver function and reduced hyperglycemia when challenged with congenital or diet-induced obesity. Furthermore, our recent findings showed that under conditions of diet-induced non-alcoholic steatohepatitis (NASH), *Slc39a5*^{-/-} mice display significantly attenuated fibrosis and inflammation. We also show that *Slc39a5*^{-/-} mice, in addition to their elevated hepatic zinc levels, have activation of hepatic AMPK and AKT signaling, at least in part due to zinc mediated inhibition of hepatic protein phosphatase activity. Taken together, these results suggest SLC39A5 maybe considered as a potential therapeutic target for non-alcoholic fatty liver disease (NAFLD) and its associated metabolic derangements including T2D.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2604. Genetic regulation of OAS1 nonsense-mediated decay underlies association with COVID-19 hospitalization in patients of European and African ancestries

Authors:

O. Florez-Vargas¹, A. Banday¹, M. Stanifer², O. Onabajo¹, B. Papenberg¹, M. Zahoor³, L. Mirabello¹, T. Ring¹, C-H. Lee¹, G. Barsh⁴, L. Biesecker⁵, M. Carrington⁶, C. Dalgard⁷, J. Edberg⁸, H. Feigelson⁹, S. Holland¹⁰, A. Hutchinson¹¹, B. Korf¹², S. Lipkin¹³, J. Pacheco¹⁴, D. Rader¹⁵, M. Ritchie¹⁵, B. Rosenblum¹⁶, M. Ross¹³, M. Yeager¹, COVNET and Genetic and functional analysis of COVID-19 severity working group, S. Boulant², S. Chanock¹, J. Feld³, L. Prokunina-Olsson¹; ¹Natl. Cancer Inst., Rockville, MD, ²Univ. of Florida, Gainesville, FL, ³Univ. of Toronto, Toronto, ON, Canada, ⁴Hudson Alpha Inst. for Biotechnology, Huntsville, AL, ⁵Natl. Human Genome Res. Inst., Bethesda, MD, ⁶Natl. Cancer Inst., Frederick, MD, ⁷Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, ⁸Univ. of Alabama at Birmingham, Birmingham, IN, ⁹Inst. for Hlth.Res., Aurora, CO, ¹⁰Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, ¹¹Natl. Cancer Inst., Bethesda, MD, ¹²Univ. of Alabama at Birmingham, Birmingham, AL, ¹³Weill Cornell Med., New York, NY, ¹⁴Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, ¹⁵Univ. of Pennsylvania, Philadelphia, PA, ¹⁶Natl. Human Genome Res. Inst., Washington DC, DC

Abstract Body:

COVID-19 is an infectious disease caused by the virus SARS-CoV-2 and, as in any other infectious disease, its clinical outcomes are determined by a complex interplay between host and pathogen factors. Heterogeneity in the clinical outcomes of COVID-19 strongly suggests the involvement of genetic factors in its pathogenesis. Genome-wide association studies (GWAS) have reported a series of genetic variants in distinct loci associated with COVID-19 susceptibility. One of these associated loci is the chr12q24.13, which harbors three genes encoding antiviral 2',5'-oligoadenylate synthetase (OAS) enzymes: OAS1, OAS2, and OAS3. While these associations were performed by comparing patients with COVID-19 and the general population, the interindividual variability in response to SARS-CoV-2 infection ranging from asymptomatic to fatal disease remains poorly understood. To address this critical gap, we have provided genetic, functional, and clinical insights into this locus in relation to COVID-19 severity. We analyzed patients of European (n=2249) and African (n=835) ancestries with hospitalized vs. non-hospitalized COVID-19. Patients were recruited by the COVNET Consortium. COVID-19 status was defined as non-hospitalized (mild), hospitalized moderate (not requiring mechanical ventilation), or hospitalized severe (mechanical ventilation or death due to COVID-19). In both ancestries, protection from hospitalized disease was consistently associated with one ancestral/introgressed Neandertal haplotype comprising several OAS1 markers. Since OAS1 is an interferon-induced antiviral protease, we further tested the haplotypes in a clinical trial with pegIFN- λ 1 and found that the OAS1 haplotypes were also associated with SARS-CoV-2 clearance. We then carried out in vitro functional studies of OAS1 with specific haplotypes and demonstrated that the combined effect of the associated splicing rs10774671 and missense rs1131454 markers increased OAS1 protein abundance through the regulation of splicing and nonsense-mediated decay. Our results suggest that the OAS1 expression is elevated in the presence of genetic variants or due to treatment with interferons, improving SARS-CoV-2 clearance and decreasing the risk of hospitalization for COVID-19 in patients of European and African ancestries.

Molecular Effects of Genetic Variation Posters - Thursday

PB2605. Genetic regulation of splicing in macular and peripheral retina reveals novel molecular mechanisms for age-related macular degeneration

Authors:

P. A. Mehta^{1,2,3}, J. M. Rouhana^{1,2,3}, A. R. Hamel^{1,2,3}, S. Mehrotra^{1,2,3}, J. Advani⁴, M. English⁴, D. A. Ferrington⁵, A. Swaroop⁴, A. V. Segrè^{1,2,3}; ¹Massachusetts Eye and Ear, Boston, MA, ²Harvard Med. Sch., Boston, MA, ³Broad Inst. of Harvard and MIT, Cambridge, MA, ⁴NIH (NIH), Natl. Eye Inst. (NEI), Bethesda, MD, ⁵Doheny Eye Inst., Univ. of California Los Angeles, Pasadena, CA

Abstract Body:

Alternative splicing of mRNA that increases diversity of the transcriptome and proteome is highly complex in the retina. The molecular mechanisms underlying 52 known genetic associations with age-related macular degeneration (AMD), a leading cause of central vision loss in the elderly, are not well understood. We hypothesized that genetic variants that alter splicing in the retina may also contribute to AMD risk. We therefore performed RNA-sequencing of 185 macular and 403 peripheral retina samples from postmortem healthy and AMD cases, and imputed the donor's genotype data to GTEx v8 whole genome sequencing variant calls. Following quality control, we used LeafCutter to detect and quantify splicing events based on reads spanning splice junctions. We identified 98,616 splice events for 12,720 genes with 76% overlap between macula and peripheral retina. About 25% of the splice events were specific to the retina compared to 49 GTEx tissues. We identified 34 splice events in 24 genes that were significantly altered between advanced AMD and healthy macula samples (Wilcoxon rank sum test $P=4E-05$). These genes were enriched in cytoskeletal protein binding and cell adhesion ($FDR<0.05$). Using FastQTL and a regression model that corrects for sex, age, AMD grade, top genotype principal components, and intron excision covariates (PEERs), we uncovered variants associated in *cis* (1Mb) with alternative splicing (splicing quantitative trait loci, sQTLs) for 1,390 and 2,850 genes in macula and peripheral retina, respectively ($FDR<5%$); the sGenes showed enrichment of respiratory electron transport, photoreceptor outer segment, and phototransduction ($FDR<0.01$). The sQTLs were most significantly enriched in splice donor/acceptor sites, followed by missense, synonymous, and intron variants (TORUS). Using colocalization analysis (eCAVIAR) between retina sQTLs that overlapped any of the 34 AMD genome-wide association study (GWAS) loci, we detected significant colocalization for 6 loci, which included a *TBC1D23* sQTL and an sQTL acting on *BLOC1S1-RDH5* transcripts that we are experimentally inspecting. Analyzing AMD associations beyond genome-wide significance, we found that macular and peripheral sQTLs were enriched for AMD associations (QTLEnrich $P<1E-05$), to a larger extent than the retina eQTLs. Notably, the target genes of retina sQTLs with AMD GWAS $P<0.05$ were enriched for rare inherited retinal degeneration (IRD) genes ($P=8E-05$), suggesting a link between alternative splicing of Mendelian disease genes and AMD. Our study suggests that regulation of splicing may play an important role in AMD susceptibility and propose new therapeutic targets.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2606*. Genetically Decreased CPS1 Activity Attenuates Atherosclerosis in Humans and Mice Through Sexually Dimorphic Patterns

Authors:

J. Hilser¹, N. Woodward², P. Huang¹, J. Gukasyan³, Z. Fouladian¹, S. Charugundla⁴, A. Lusic⁵, L. Ma⁶, J. Bjorkegren⁶, Z. Wang⁷, W. Tang⁷, S. Hazen⁷, H. N. Hodis¹, W. Mack¹, J. Hartiala⁸, H. Allayee⁹; ¹Univ. of Southern California, Los Angeles, CA, ²Univ. of Southern California, Los Angeles, CA, ³USC, Van Nuys, CA, ⁴Univ. of California Los Angeles, Los Angeles, CA, ⁵UCLA, Los Angeles, CA, ⁶Mount Sinai Sch. of Med., New York, NY, ⁷Cleveland Clinic, Cleveland, OH, ⁸Univ of Southern California, Los Angeles, CA, ⁹USC Keck Sch. of Med., Los Angeles, CA

Abstract Body:

Background: We previously reported that a variant (rs715) of *CPS1*, the rate-limiting enzyme of the urea cycle, exhibited a female-specific pattern of association with lower urea cycle metabolite levels and decreased risk of coronary artery disease (CAD). In the present study, we sought to independently replicate these findings in multi-ancestry cohorts and genetically modified mice. **Results:** Consistent with prior observations, rs715 was significantly associated with decreased risk of CAD in women in the UK Biobank (OR=0.96, 95% CI 0.94-0.99; $p=4.2 \times 10^{-3}$) and Biobank Japan (OR=0.91, 95% CI 0.87-0.95; $p=3.0 \times 10^{-4}$). This association increased in significance when meta-analyzed together (OR=0.94, 95% CI 0.92-0.96; $p=8.3 \times 10^{-8}$) along with data from the CARDIoGRAM+C4D Consortium. A similar meta-analysis in men did not reveal association of rs715 with CAD (OR=0.99, 95% CI 0.98-1.01; $p=0.21$). A meta-analysis with several hormone or nutraceutical clinical trials in women also revealed cross-sectional association between rs715 and decreased carotid intima-media thickness ($\beta=-0.046$, SE=0.011 $p=1.6 \times 10^{-5}$). To complement these studies, we characterized heterozygous *Cps1* deficient (*Cps1*^{+/-}) mice, which exhibited the same pattern of metabolomics associations observed in humans. On a hyperlipidemic background, genetically decreased *Cps1* activity also significantly decreased aortic lesions in male mice compared to wildtype (WT) ($452,855 \pm 23,550$ vs $343,481 \pm 29,752 \mu\text{m}^2/\text{section}$; $p=0.009$). A similar male-specific pattern was observed when atherosclerosis was assessed along the entire aorta by *en face* analysis in WT and *Cps1*^{+/-} mice (6.4 ± 1.2 vs $2.3 \pm 0.6\%$; $p=0.01$). By comparison, *Cps1* deficiency had no effects on atherosclerotic plaque development in female mice. Gene expression analysis in humans revealed that women have ~20% lower hepatic *CPS1* expression compared to men (828 ± 385 vs 1024 ± 582 TPM; $p=4.6 \times 10^{-6}$) regardless of rs715 genotype whereas the opposite pattern is observed in WT mice with males exhibiting ~50% lower hepatic *Cps1* expression than WT females (0.73 ± 0.22 vs 1.3 ± 0.37 ru; $p=1.6 \times 10^{-4}$). Thus, human females and male mice have naturally lower *CPS1* expression, which taken together with our genetic and atherosclerosis analyses, provide an explanation for the opposite but sexually dimorphic pattern of associations observed in the two species. **Conclusions:** Our results provide further evidence for the causal and sex-specific relevance of *CPS1* to atherosclerosis in both humans and mice. Further studies will be needed to elucidate the biological mechanism for the atheroprotective effect of genetically decreased *CPS1* activity.

Molecular Effects of Genetic Variation Posters - Thursday

PB2607. Genome-First Approach to Explore Prevalence and Cancer Risk of Adults with Pathogenic and Likely Pathogenic Variants in RASopathy Genes

Authors:

E. Astiazaran Symonds¹, G. M. Ney¹, J. Kim¹, J. S. Haley², M. N. Frone¹, A. Pemov¹, M. E. Yohe³, A. M. Gross³, D. Collaboration², N. T. Strande², H. Rao², M. Shadrina⁴, D. J. Carey², B. D. Gelb^{4,5}, D. R. Stewart¹; ¹Clinical Genetics Branch, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, ²Geisinger Clinic, Danville, PA, ³Pediatric Oncology Branch, Ctr. for Cancer Res., Natl. Cancer Inst., Bethesda, MD, ⁴Dept.s of Pediatrics and Genetics and Genomic Sci., Icahn Sch. of Med., New York, NY, ⁵Mindich Child Hlth.and Dev. Inst., Icahn Sch. of Med., New York, NY

Abstract Body:

The RASopathies are a group of disorders with multisystemic manifestations, including tumor predisposition, arising from germline pathogenic/likely pathogenic variants (GPVs) in RAS-MAPK pathway genes. Estimations of prevalence and penetrance have been focused on pediatric populations and are limited by syndrome rarity and incomplete ascertainment. To address these limitations, a genome-first approach was used to analyze exome sequencing and phenotypic data from electronic health records (EHRs) from 1) Mount Sinai's BioMe (n=32,344), 2) the UK Biobank (UKB, n=200,600) and 3) Geisinger MyCode Community Health Initiative (n=175,499). We focused on 24 genes from the ClinGen RASopathy Gene Curation Expert Panel (EP) and applied the ClinGen RASopathy Variant Curation EP's methods to identify individuals with heterozygous (VAF>25%) GPVs. We retrieved ICD codes from EHRs for 1) RASopathies and 2) neoplasms. We identified 32 individuals with GPVs in BioMe (1:1,011), 112 in UKB (1:1,791) and 139 in MyCode (1:1,262). Overall, we found a prevalence of 1:1,745 for heterozygotes for GPVs associated with Noonan syndrome (NS), 1:12,762 for Legius syndrome (LS), 1:68,065 for cardiofaciocutaneous (CFC) syndrome and 1:40,839 for *LZTR1*-related schwannomatosis. The prevalence of GPVs for any RASopathy across all cohorts was 1:1,433. Subsequently, we performed analyses to estimate cancer risk in MyCode individuals with GPVs using phecodes, a strategy that groups different ICD codes into meaningful phenotypes. We saw significantly elevated relative risks (RRs) for acute and chronic myeloid leukemia (9.2, 95% CI 2.3-36.5 and 14.4, 95% CI 3.6-57.7), myeloproliferative disease (4.2, 95% CI 2.3-7.7) and brain cancer (3.4, 95% CI 1.6-7.5) in NS, for melanoma (12, 95% CI 3.5 to 41.4) in CFC, and for bone cancer (5.2, 95% CI=1.2-20) and reticulosarcoma (4.7, 95% CI 1.5-14.6) in *LZTR1* heterozygotes. When combining all individuals with GPVs in RASopathy genes, associations with testicular cancer (7.4, 95% CI 1.8-29.7), and cancer of nasal cavities (5.0, 95% CI=1.3-20.2) were uncovered. The ICD code that includes NS (Q87.19) was linked to only 4.6% of heterozygotes. This genome-first approach identified heterozygotes of RASopathy gene GPVs while mitigating ascertainment bias and providing more precise prevalence estimates. Additionally, we describe for the first time the spectrum of cancers in adult RASopathy gene heterozygotes, including new associations to confirm in larger studies. A substantial number of individuals identified had not been diagnosed with a RASopathy, suggesting the potential for improved and earlier detection and treatment using this approach.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2608. Genome-wide association study of serum metabolites and lipids in Multiple Sclerosis

Authors:

C. Yang¹, E. Marshall¹, N. Sadhu¹, F. B. Briggs², K. C. Fitzgerald³, H-H. Tsai¹, K. Trinh¹, H. Mclaughlin¹, C. Singh¹, R. Wei¹, D. Conway⁴, M. Comabella⁵, H. Wiendl⁶, E. M. Mowry³, P. Calabresi³, S. Belachew¹, S. John¹, C. Shen¹, H. Runz¹, A. Gafson¹, P. G. Bronson¹; ¹Biogen, Cambridge, MA, ²Case Western Reserve Univ., Cleveland, OH, ³Johns Hopkins Univ., Baltimore, MD, ⁴Cleveland Clinic, Cleveland, OH, ⁵Vall d'Hebron Univ. Hosp., Barcelona, Spain, ⁶Univ. of Münster, Münster, Germany

Abstract Body:

Introduction: Multiple sclerosis (MS), a neurodegenerative disease, affects millions worldwide. Dysregulation of cellular metabolism and abnormal metabolite levels have been reported in MS. An individual's metabolic phenotype is influenced by multiple levels of biologic interaction including genetics. Metabolite levels can serve as intermediate traits between genotype and disease phenotype. We aim to identify single nucleotide polymorphisms (SNPs) associated with metabolite levels (mQTLs), which may highlight causal pathways in MS.

Methods: MS PATHS (Partners Advancing Technology and Health Solutions) is a longitudinal study collecting clinical and MRI data in >20,000 persons with MS (pwMS). We profiled 663 metabolites (Biocrates platforms: MxP Quant 500, Eicosanoid, Oxysterol) in non-fasting serum from 863 MS PATHS participants of European ancestry (n = 863). Inclusion criteria: availability of serum and genetic data. Exclusion criteria: hypercholesterolemia, statin use. After stringent quality control, we integrated metabolomics with genetics to identify mQTLs. We tested 517 metabolites and 10.6 million SNPs with linear regression, including age, sex, and the top 10 principal components for ancestry. After Bonferroni correction for multiple testing ($5 \times 10^{-8}/517$), the significance threshold was $P < 9.7 \times 10^{-11}$. Colocalization analyses with MS risk, UK Biobank traits, methylation QTLs, and known mQTLs were conducted for our significant mQTLs.

Results: We identified 17 significant ($P < 9.7 \times 10^{-11}$) associations between SNPs and metabolites, of which five were novel: three triacyl glyceride metabolites [*FADS2*], ceramide [*BANF2*], and phosphatidylcholine [*MYRF*]. *MYRF* regulates differentiation of myelinating cells (oligodendrocytes). Three known mQTLs had evidence for shared causal variants (PP4>0.95). *CPS1* rs1047891*A (missense) and *SLC16A9* rs1171617*T (intronic) were associated with increased glycine and carnitine, respectively, as well as increased fat-free mass traits. *ASPG* rs2282377*G (intronic) was associated with increased asparagine and reduced methylation.

Conclusions: This is the first large-scale mQTL study in pwMS. We identified 17 significant mQTLs, of which five were novel, including an mQTL for phosphatidylcholine that mapped to *MYRF*. We replicated 12 mQTLs, including mQTLs for the amino acid neurotransmitters glycine and asparagine. Next steps include replicating our novel mQTLs, and testing whether loci for longitudinal MRI traits in MS share a causal variant with our mQTLs. Understanding genetic predictors of metabolite levels may provide biological insight into MS pathogenic mechanisms.

Molecular Effects of Genetic Variation Posters - Thursday

PB2609. Genome-wide characterization of selective constraint on variation within CTCF Binding Sites.

Authors:

C. Tubbs¹, M. Benton², E. McArthur¹, J. Capra³, D. Ruderfer⁴; ¹Vanderbilt Univ., Nashville, TN, ²Baylor Univ., Waco, TX, ³Univ. of California San Francisco, San Francisco, CA, ⁴Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Background: CCCTC-binding factor (CTCF) binds DNA at tens of thousands of conserved sequence motifs. These CTCF Binding Sites (CBSs) facilitate gene expression by regulating 3D genome organization. Mutations in CBSs have clinical impact; their disruption has been linked to the etiologies of both rare and common disorders. However, we currently lack a comprehensive understanding of when and how CBS variation contributes to disease. We hypothesize that mutations that disrupt CBSs will be enriched for deleterious effects and carry signals of purifying selection. **Methods:** To test our hypothesis, we assessed the allele frequency (AF) of genetic variation at CBSs from 76k whole-genome sequenced samples representing diverse global ancestries. We defined CBSs by considering all genomic regions that have significant match to the 19 bp CTCF core binding motif (JASPAR, MA0139.1). To characterize the AF spectrum of variants at these sites, we queried the gnomAD database (v3.1.2) for all overlapping, mono allelic SNVs. Finally, we classified our CBSs by their genomic context and putative regulatory role in the genome by annotating whether they cluster, associate with insulatory boundary regions, overlap an ENCODE candidate cis-regulatory element, or act as a chromatin loop anchor. **Results:** We defined 767,259 CBSs that possess a significant ($p < 10^{-5}$) sequence match to the CTCF binding motif and 153,535 with functional support using the JASPAR and ENCODE databases. We identified 3.41M overlapping SNVs (mean AF=0.00363) in total and detected significantly lower AF (mean AF=0.00358, $p=2.62 * 10^{-11}$) for those that overlapped a CBS with a putative functional annotation. We further observed variants that overlapped chromatin loop anchors to be the most depleted of variation. We next hypothesized that a SNV's predicted impact on the binding affinity of a CBS would further modulate its frequency in the population. We stratified variants by their position in the CBS motif and calculated their impact on binding affinity by taking the difference in motif scores for the reference and alternate alleles. We observed lower AF (mean AF=0.00327, $p=5 * 10^{-3}$) for variants that significantly modify the binding motif compared to those that do not (mean AF=0.00381). **Conclusion:** Currently, we lack a framework for interpreting the effects of variants that disrupt CBSs. Here, we present an approach that integrates population genetic data with regulatory genome annotations to characterize the constraint on CBSs. This work represents an important first step in characterizing the contribution of CBS variation to human disease and will allow for future interpretation of their pathogenicity.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2610. Genome-wide microRNA expression quantitative trait loci from 596 adult brains

Authors:

S. Vattathil¹, S. Canon¹, A. Lori¹, E. S. Gerasimov¹, Y. Liu¹, D. A. Bennett², T. S. Wingo¹, A. P. Wingo^{1,3}; ¹Emory Univ., Atlanta, GA, ²Rush Univ. Med. Ctr., Chicago, IL, ³VA Atlanta Hlth.Care System, Decatur, GA

Abstract Body:

MicroRNAs (miRNAs) are essential post-transcriptional regulators that suppress gene expression by binding complementary mRNA sequences and causing either transcript degradation or translation repression. These non-coding RNAs are too small (~22 nucleotides) to be captured in RNA sequencing experiments targeting protein-coding transcripts, and while expression quantitative trait loci (eQTL) for protein-coding genes have been profiled across human tissues, to date there have been few large studies focusing on miRNA eQTLs (miR-QTLs). In this study, we aimed to map and characterize miR-QTLs in adult human brain tissue. Brain dorsolateral prefrontal cortex specimens were collected after autopsy of participants from the Religious Order Study (ROS) or the Rush Memory and Aging Project (MAP). Both studies are U.S.-based longitudinal community-based cohort studies of cognitive decline and aging. The 596 ROS/MAP participants included here were all of European ancestry and mean age at death was 90.3 years (standard deviation 6.2 years). MiRNA abundance was quantified using small-RNA sequencing, and genotypes were called from whole-genome sequencing or imputed from SNP array genotypes. We focused on common (minor allele frequency > 5%) single nucleotide polymorphisms (SNPs) within 500 Kb of miRNA genes. For each miRNA, we used linear regression to test for association between SNP genotype and miRNA abundance while controlling for the effect of age at death, sex, clinical diagnosis, post-mortem interval, RNA integrity number (RIN), sequencing batch, genetic principal components, and surrogate variables estimated from the miRNA data. We found miR-QTLs for 310 miRNAs (50% of 623 tested miRNAs) at false discovery rate < 1%. After LD-based clumping of SNPs for each miRNA (e.g. pruning SNPs with $R^2 > 0.5$ in 250 Kb windows), the results include 2,419 index miR-QTLs. Among these, 8 miR-QTLs are located within the corresponding miRNA gene, and another 938 miR-QTLs (~40% of index miR-QTLs) are located within 50 Kb of the gene. MiR-QTLs are enriched for SNPs that overlap brain promoters compared to non-miR-QTL SNPs (Mantel-Haenszel common odds ratio 1.44, p-value 5.6×10^{-9}), and a subset of miR-QTLs overlap promoters predicted to regulate the specific associated miRNA. To our knowledge, this is the first genome-wide adult brain miR-QTL study. This miR-QTL catalog complements existing and future miRNA analyses in other tissues, developmental stages, and cohorts, and the results may be integrated with transcriptomic, proteomic and phenotypic data to better understand the contribution of miRNAs to pathogenesis of brain illnesses.

Molecular Effects of Genetic Variation Posters - Thursday

PB2611*. Genome-wide pleiotropy study identifies association of *PDGFB* with Age-Related Macular Degeneration and COVID-19 infection outcomes.

Authors:

X. Sun¹, J. Chung², V. Vig¹, X. Han¹, G. O'Connor¹, X. Chen¹, M. DeAngelis³, M. Subramanian¹, L. Farrer⁴; ¹Boston Univ. Sch. of Med., Boston, MA, ²Boston Univ., Boston, MA, ³Jacobs Sch. of Med. and BioMed. Sci., Buffalo, NY, ⁴Boston Univ Sch Med, Boston, MA

Abstract Body:

Background Patients with age-related macular degeneration (AMD) have been shown to be at higher risk of infection and complications of COVID-19. We leverage large-scale data of genome-wide association studies (GWAS) for AMD and COVID-19 to assess the relationship between the two diseases and identify shared genetic susceptibilities between AMD and COVID-19. **Methods** We evaluated the genetic architecture shared between AMD (n = 33,976) and three COVID-19 outcomes (n ≥ 1,388,342) including critical illness, hospitalization, and infections by assessing genetic correlations and conducting a pleiotropy analysis of these two diseases using the previous GWAS summaries and Multi-Trait Analysis of GWAS (MTAG) software. The functional relevance of our most significant finding was examined by analyses of gene expression and DNA methylation quantitative trait locus (eQTL), transcriptome datasets for AMD and COVID-19, and Mendelian randomization (MR). **Results** A modest, but significant, genetic correlation of AMD was observed with COVID-19 infection ($r_G = 0.10$, $P = 0.02$). Our pleiotropy analysis for the two diseases identified one novel genome-wide significant ($P < 5 \times 10^{-8}$) association with several single nucleotide polymorphisms (SNPs) located 18.1kbp upstream of *PDGFB* (best SNP: rs130651; MTAG $P = 2.4 \times 10^{-8}$; AMD OR = 0.96, $P = 1.4 \times 10^{-7}$; COVID-19 OR = 0.90, $P = 0.02$). The rs130651 disease risk allele was significantly associated with increased *PDGFB* expression in multiple tissues (best eQTL, $P = 1.8 \times 10^{-11}$ in blood) and decreased DNA methylations in the *PDGFB* promoter region (best mQTL, cg11247378, $P = 9.74 \times 10^{-64}$). Moreover, *PDGFB* expression was higher in AMD cases than AMD controls ($p = 0.067$), as well as in the peak COVID-19 symptom stage (11-20 days after onset) compared to early/progressive stage (0-10 days) among COVID-19 patients over age 40 ($P = 0.03$). Our MR analysis found that the liability to AMD risk derived from the complement system dysfunction (OR [95% CI]; hospitalization = 1.02 [1.01- 1.03], infection = 1.02 [1.01-1.03]) and the increased levels of serum cytokine PDGF-BB (β [95% CI]; critical illness = 0.07 [0.02-0.11]) are significantly associated with the COVID-19 outcomes. **Conclusion** Our study demonstrated that AMD is associated with an increased risk of COVID-19 infection and *PDGFB* may be responsible for the severe clinical outcomes of COVID-19 among the AMD patients. Our findings may provide some insights about pathophysiology and potential interventions for these disorders.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2612. Genome-wide transcriptome correlation analysis identifies shared pathways between coronary heart disease and hematologic traits

Authors:

H. Abe¹, E. Gamazon²; ¹Vanderbilt Univ., NASHVILLE, TN, ²VUMC Clare Hall, Univ. of Cambridge, Nashville, TN

Abstract Body:

Recent studies have reported an increased risk of developing coronary heart disease (CHD) in patients who have undergone treatment for hematologic malignancies. Furthermore, clonal hematopoiesis of indeterminate potential (CHIP), a pre-malignant condition that results from clonal expansion of blood or immune cells, has been shown to increase risk of both coronary heart disease (CHD) and hematologic malignancies. Nevertheless, our understanding of the underlying biology and of the specific genes involved in the shared etiology of CHD and hematologic traits remains limited. Evaluation of gene expression regulation may identify novel genes and pathways that can serve as targets for therapeutic application and functional analysis.

Here, we curated genome-wide association summary statistics from large cohorts such as the UK Biobank and examined the shared genetics between CHD (N= 32926), 8 hematologic malignancies (N range= from 209 to 2279) and 23 blood cell traits(N= 427154). We apply PrediXcan/TWAS, which estimates the genetically regulated expression (GReX) of a gene and tests its association with traits. Specifically, we use the recently developed joint-tissue imputation (JTI) models that leverage tissue similarity of gene expression and the regulatory information from reference functional genomic resources such as GTEx and Roadmap to estimate GReX. Using the zscore estimates from this method, we calculated the correlation of the genetically regulated effect spanning 1800 biological pathways. This allows us to examine the extent to which pathways associated with hematologic traits have similar effects on CHD. Our results identified pathways, including the MAP kinase activation and mTORC1 mediated signaling, as strongly correlated between CHD and hematologic malignancies at the level of GReX. Moreover, we observed inflammation, cell division, DNA replication, as well as mitochondrial fatty acid oxidation, as the top shared pathways across the blood cellular traits. Our results highlight potentially causal molecular mechanisms and genes whose phenotypic effect on disease classes is mediated mainly through gene expression. The approach we present has broad implications for the search for genetic relationships between other cardiovascular and hematologic traits.

Molecular Effects of Genetic Variation Posters - Thursday

PB2613. Genomic characterization of the immunoglobulin light chain lambda locus from individuals of European, Asian and African origin.

Authors:

W. Gibson¹, O. Rodriguez¹, K. Shields², A. Silver¹, A. Dorgham³, M. Sorensen⁴, G. Deikus⁵, R. Sebra⁵, E. Eichler⁴, A. Bashir⁶, M. Smith³, C. Watson³; ¹Univ. of Louisville, LOUISVILLE, KY, ²Univ. of Louisville, Kentucky, KY, ³Univ. of Louisville, Louisville, KY, ⁴Univ. of Washington, Seattle, WA, ⁵Icahn Sch. of Med. at Mount Sinai, New York, NY, ⁶Google, Palo Alto, CA

Abstract Body:

The adaptive immune system relies on a diverse set of over one hundred immunoglobulin (IG) genes across three genomic loci that are variably combined to form antibodies (Ab). Studies show that the IG loci are highly variable between individuals and populations. However, the complexity of the IG loci severely limits the effective use of standard short read sequencing, limiting our knowledge of population diversity in these loci. We leveraged existing long read whole-genome sequencing (WGS) data, fosmid technology, and targeted long-fragment DNA capture (IG-cap) combined with single molecule, real time (SMRT) long read sequencing (Pacific Biosciences) to create haplotype-resolved assemblies of the IG Lambda (IGL) locus from 6 ethnically diverse individuals. In addition, we generated 10 diploid assemblies of IGL utilizing IG-cap combined with SMRT sequencing from a diverse cohort of individuals. These data represent highly accurate base-level assemblies of the IGL region. From these 16 individuals, we identified significant allelic diversity, including 37 novel IGLV alleles. In addition, we observed highly elevated single nucleotide variation (SNV) in IGLV genes relative to average IGL intergenic and genomic background SNV density. By comparing SNV calls between our high quality assemblies and existing short read datasets from the same individuals, we show a high propensity for false-positives in the short read datasets. Finally, for the first time we haplotype-resolved, at nucleotide resolution, common 5-10 Kb duplications in the IGL constant region that contain functional IGLJ and IGLC genes. Together these data represent a significant advancement in our understanding of genetic variation and population diversity in the IGL locus.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2614*. Glucocorticoids unmask silent non-coding genetic risk variants for common diseases.

Authors:

T. Nguyen, H. Gao, D. Liu, Z. Ye, J-H. Lee, G. Shi, K. Copenhaver, L. Zhang, L. Wei, J. Yu, Z. Cheng, H. Li, L. Wang, T. Ordog, R. Weinshilboum; Mayo Clinic, Rochester, MN

Abstract Body:

Most disease-associated sequence variants are non-coding, and understanding their function remains a major challenge for biology. Many diseases are the result of complex gene-by-environment interactions, increasing the difficulty in understanding mechanistic function of variants. Using the glucocorticoid signaling system, we discovered and exhaustively characterized environment-responsive genetic variants. We first identified 134 glucocorticoid receptor (GR)-mediated expression quantitative trait loci (eQTL) across 30 human B-lymphoblastoid cell lines (LCLs) with exposure to a GR agonist and antagonist. Using machine learning-based chromatin modeling to integrate 16 LCL epigenomic datasets, we predicted that three-quarters of the GR-modulated SNPs mapped to enhancers, half of which with promoter-looping properties. This observation fits with current knowledge of GR being primarily an enhancer mediator.

Based on these predictions, we utilized a massively parallel reporter assay to directly test variant x stimulus interactions by determining transcriptional activities of all GR-mediated loci across different drug treatments and cell lines. Up to 81% of the loci displayed allele-specific transcriptional activity in response to drugs, with high consistency across two tested cell lines. To understand the regulatory architecture of these variants, we performed H3K27ac HiChIP before and after GR activation. We observed that in 75% of SNP-gene pairs in active enhancers, GR either induced the connections between SNP-gene pairs, or mediated transcription of already established loops.

Lastly, we investigated disease risk SNP mechanisms overlapping with GR-mediated eQTLs, identifying 30 associations with GR-related diseases across autoimmunity, metabolism, mood disorders, osteoporosis and cancer. Without “unmasking” of variant function after glucocorticoid exposure, the causal variants and their target genes would have been overlooked. For example, SNPs previously associated with obesity mediated cortisol-responsive expression of *FBXL19*, an adipogenesis-controlling gene, and *HSD3B7*, a cholesterol metabolizing enzyme. We validated a GR-mediated eQTL for breast cancer with GR-targeted ChIP-seq, RNA-seq and H3K27ac HiChIP in breast cancer cell line and observed similar interaction between the risk SNP and the novel risk gene *MAST4* identified originally in LCLs. Furthermore, *MAST4* expression was dramatically repressed in breast cancer tissue and predictive of treatment outcome. Together, these results provide a mechanistic framework for understanding the function of non-coding genetic variants across environments.

Molecular Effects of Genetic Variation Posters - Thursday

PB2615*. Hematopoietic Loss of Y Chromosome Leads to Cardiac Fibrosis and Heart Failure Mortality

Authors:

L. Forsberg¹, S. Sano², K. Horitani², H. Ogawa², J. Halvardson¹, N. Chavkin², J. Mattisson¹, M. Danielsson¹, A. Zaghlool¹, K. Walsh²; ¹Uppsala Univ., Uppsala, Sweden, ²Univ. of Virginia Sch. of Med., Charlottesville, VA

Abstract Body:

Mosaic loss of chromosome Y (mLOY) in blood is the most common acquired human mutation and detectable in more than 10% of blood leukocytes in at least 5%, 20% and 40% of 50, 60 and 70 year old men, respectively. Remarkably, mLOY is associated with a diverse and growing list of diseases, notably many forms of hematological and non-hematological cancers, Alzheimer's disease, cardiovascular events, autoimmune conditions, age-related macular degeneration and type 2 diabetes. While associated with the most common causes of death in general populations; men with mLOY display increased mortality in prospective analyses. As a common and male specific somatic mutation, mLOY helps explain why men in the entire world live shorter lives compared with women. Beyond the risk factor of age itself, genomewide association studies (GWAS) identifies germline risk loci located near genes involved in cell-cycle regulation and genome instability. A polygenic score suggest that individuals with high genetic risk could be up to five times more likely to be affected during aging. Furthermore, mLOY often co-occur with point mutations associated with clonal hematopoiesis of indeterminate potential (CHIP). A number of preventable risk factors for mLOY such as smoking, air pollution and other environmental exposures has also been described. Despite recent attention, the causal relationship between mLOY in leukocytes and disease manifesting in other organs have been largely unknown. Using mouse models reconstituted with bone marrow cells lacking the Y chromosome by CRIPR-Cas9 editing, we here describe a causal link between hematopoietic mLOY and age-dependent cardiac dysfunction. We found that Y chromosome-deficient cardiac macrophages over-activate a profibrotic signaling network, leading to cardiac fibroblast proliferation and activation, excessive matrix production, and diminished heart function. We also found that a neutralizing TGF β antibody reversed the pathological cardiac phenotypes caused by mLOY in the mouse model. Consistent with this, survival analyses of men in the UK Biobank showed that mLOY in leukocytes at study entry is associated with risk for death caused by cardiovascular disease overall during follow up; an association mainly driven by heart failure, hypertensive heart disease and aortic aneurysm. Tissue fibrosis is a hallmark of aging and estimated to contribute to 45% of deaths in industrialized countries. In view of recent efforts to treat heart failure, idiopathic pulmonary fibrosis, and some cancers with antifibrotic approaches, men with mLOY could represent a patient subpopulation that exhibits a superior response to this class of therapeutic agents.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2616. Human genetic variation reveals regulators of *Yersinia pestis* cellular Infection.

Authors:

R. Keener, L. Wang, D. Ko; Duke Univ., Durham, NC

Abstract Body:

Yersinia (Y.) pestis is the gram-negative pathogenic bacterium responsible for the deadliest pandemic in recorded human history, the Black Death. This pathogen causes such catastrophic pandemics because of its variable 30-100% mortality rate in infected humans when antibiotic intervention is unavailable. Factors like time to treatment and nutrition can cause some of this variation but, like other infectious diseases, we hypothesize host genetic variation also plays a role. These genetic differences can normally be identified by human challenge studies or genome-wide association studies (GWAS) of natural infection. Both quantify disease symptoms, physiology, and outcomes then compare these phenotypes to the genetics of the individual. Unfortunately, using these techniques for *Y. pestis* is difficult: high lethality makes human challenges unethical, and the infrequency of present-day *Y. pestis* outbreaks coupled with non-random population exposure make GWAS too underpowered and confounded. To overcome these hurdles, we developed a genetically diverse cellular GWAS approach. This method, called Hi-HOST (high-throughput human *in vitro* susceptibility testing), is unhindered by pre-existing conditions, medical care access, diet, or other environmental and life history confounders found in traditional human disease related GWAS. Additionally, Hi-HOST previously identified variation in *VAC14* and *ARHGEF26* loci that modulate *Salmonella* infection. For *Y. pestis* Hi-HOST, we infected LCLs (lymphoblastoid cell lines; immortalized B cells) from 978 genetically distinct individuals in eight global populations with GFP-tagged *Y. pestis*. Using flow cytometry, we measured percent infection (proportion of cells that are GFP+) as well as bacterial burden (median fluorescence of GFP+ cells). Each phenotype was tested for association with approximately 15 million single nucleotide polymorphisms (SNPs), using family-based association to control for population stratification while also allowing for heritability estimates. This uncovered a genome-wide significant hit ($p = 3 \times 10^{-9}$) in the intracellular survival phenotype. We hypothesize the gene containing this SNP helps modulate the intracellular niche of *Y. pestis*. Additionally, stratifying for likely functional variants revealed a nonsynonymous SNP associated with 24-hour percent infection with greater significance ($p=3.5 \times 10^{-07}$) than expected based on the neutral distribution. Experimental follow-up studies suggest the protein product of this gene acts as a previously unrecognized receptor for *Y. pestis* in lymphoid lineage cells, regulating *Y. pestis* attachment and uptake.

Molecular Effects of Genetic Variation Posters - Thursday

PB2617. Human pancreatic islet microRNAs implicated in diabetes and related traits by large-scale genetic analysis

Authors:

H. Taylor^{1,2}, Y-H. Hung³, N. Narisu¹, M. Erdos¹, M. Kanke³, T. Yan¹, C. Grenko¹, A. Swift¹, L. Bonnycastle¹, P. Sethupathy³, F. Collins¹, D. Taylor¹; ¹Ctr. for Precision Hlth.Res., Natl. Human Genome Res. Inst., NIH, Bethesda, MD, ²Cardiovascular Epidemiology Unit, Dept. of Publ. Hlth.and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, ³Dept. of BioMed. Sci., Coll. of Vet. Med., Cornell Univ., Ithaca, NY

Abstract Body:

Genetic studies have identified ≥ 240 loci associated with risk of type 2 diabetes (T2D), yet most of these loci lie in non-coding regions, masking the underlying molecular mechanisms. Recent studies investigating mRNA expression in human pancreatic islets have yielded important insights into the molecular drivers of normal islet function and T2D pathophysiology. However, similar studies investigating microRNA (miRNA) expression remain limited. Here, we present data from 63 individuals, representing the largest sequencing-based analysis of miRNA expression in human islets to date. We characterize the genetic regulation of miRNA expression by decomposing the expression of highly heritable miRNAs into *cis*- and *trans*-acting genetic components and mapping *cis*-acting loci associated with miRNA expression (miRNA-eQTLs). We find (i) 81 heritable miRNAs, primarily regulated by *trans*-acting genetic effects, and (ii) 5 miRNA-eQTLs. We also use several different strategies to identify T2D-associated miRNAs. First, we colocalize miRNA-eQTLs with genetic loci associated with T2D and multiple glycemic traits, identifying one miRNA, miR-1908, that shares genetic signals for blood glucose and glycated hemoglobin (HbA1c). Next, we intersect miRNA seed regions and predicted target sites with credible set SNPs associated with T2D and glycemic traits and find 32 miRNAs that may have altered binding and function due to disrupted seed regions. Finally, we perform differential expression analysis and identify 13 miRNAs associated with T2D status—including miR-187-3p, miR-21-5p, miR-668, and miR-199b-5p—and 4 miRNAs associated with a polygenic score for HbA1c levels—miR-216a, miR-25, miR-30a-3p, and miR-30a-5p.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2618. Hypotonia due to homozygous mutation of *NALCN*

Authors:

O. Ope^{1,2}; ¹St. George Univ. Sch. of Med., Detroit, MI, ²Ascension St. John Hosp., Detroit, MI

Abstract Body:

NALCN encodes a sodium ion leak channel expressed in the nervous system. As a leak channel, it conducts a persistent influx of sodium ions to facilitate formation of action potentials. Homozygous or compound heterozygous loss of function variants in *NALCN* cause infantile hypotonia with psychomotor retardation and characteristic facies-1 (IHPRF1; OMIM 615419). Through exome sequencing, we identified homozygous *NALCN* p.Arg735Ter pathogenic variants in two siblings of Afro-Caribbean ancestry. Both individuals had failure to thrive, severe hypotonia, and dolichocephaly. Diagnosis was obtained as a result of a successful international collaboration originating from a resource limited setting on a small Caribbean island. Due to its geographical isolation and low socioeconomic status, the island lacks many specialty medical services, including clinical genetics. As with many rare syndromes, individuals of Afro-Caribbean ancestry with *NALCN* variants are rarely reported in the medical literature. This is an especially important genetic diagnosis, as it is associated with sudden cardiac death, and the older sibling in this family died suddenly without a known etiology after evaluation but before molecular diagnosis.

Molecular Effects of Genetic Variation Posters - Thursday

PB2619. Identification of bidirectional regulatory regions for *FADS1* and *FADS2*

Authors:

S. Yang, H. Zhang, K. Ye; Univ. of Georgia, Athens, GA

Abstract Body:

The *FADS1* and *FADS2* genes encode two rate-limiting enzymes in the biosynthesis process of long-chain polyunsaturated fatty acids (LC-PUFAs). These two genes are located adjacent to each other in a head-to-head orientation, resulting in a potential shared regulatory region between them. However, whether and how the shared region affects their expression is still largely unknown. Using GTEx data, many eQTLs in this region were found to be associated in opposite directions with *FADS1* and *FADS2* expression. By integrating ChIP-seq data in HepG2 cell lines from ENCODE, we revealed that two causal variants for *FADS* expression, rs174557 and rs968567, both lie in binding sites for transcription factor SP1 and SREBP1c. Our reporter assay showed that overexpression of SP1 and SREBP1c enhances promoter or enhancer activities of the shared regulatory regions. We will apply CRISPRi approach to interfere their binding sites in these regions to investigate the potential competition on transcription factors between *FADS1* and *FADS2*. In addition, by analyzing genotypes of rs174557 and rs968567 in 5005 individuals across multiple human populations, we found that A allele of rs174557 and T allele of rs968567 do not appear together; the two alleles increase the transcription of *FADS1* and *FADS2*, respectively. To examine if the combinations of alleles at these two SNPs are sufficient to explain the bidirectional regulation on *FADS1* and *FADS2*, we will apply prime editing in HepG2 cell lines. Taken together, findings from this study will enhance our understanding of *FADS* expressions and facilitate the development of genome-informed LC-PUFAs supplementation for preventing associated complex diseases.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2620. Identification of polyadenylation signals relevant to Mendelian disease variant interpretation.

Authors:

H. Shiferaw¹, C. Hong², J. Johnston³, L. Biesecker⁴; ¹NIH, Bethesda, MD, ²Natl. Inst.S OF HEALTH, BETHESDA, MD, ³NIH, Rockville, MD, ⁴NIH NHGRI, Bethesda, MD

Abstract Body:

Polyadenylation is essential in maintaining nascent mRNA stability. Variants in polyadenylation signal hexamers (PASH) can result in reduced polyadenylation at associated polyA sites and lead to altered gene expression. Only 31 PASH variants in 22 genes have been associated with Mendelian disorders. Of the 22 genes, 14 genes have at least one PASH variant listed as disease causing ('DM') in HGMD. We hypothesize that this is an under-representation of this class of variant. Here, we aimed to identify clinically important PASH by examining the usage of polyA sites and the traits of associated PASH. Polyadenylated 3' UTRs from EST data were collected from the PolyA Site 2.0 database, re-annotated, and examined for predominant polyA site usage activity (defined by >50% of overall EST representation). Associated PASH were selected based on the relative location to the predominant polyA site and the strength of the hexamer motif. We identified 15,212 predominant PASH for further examination. To understand constraint in the identified predominant PASH, we compared the frequency of variants in the predominant PASH vs. control sequences by examining variants in gnomAD. The predominant PASH were significantly more constrained than control sequences ($p < 0.001$, Mann-Whitney U test). Additionally, distribution of deleteriousness scores from *in-silico* predictors (FATHMM-MKL and CADD) for potential variants in predominant PASH were predicted to be more deleterious than potential variants in control sequences ($p < 0.001$, Mann-Whitney U test). To understand the effects of variants occurring in PASH, RNA sequencing and 3' end sequencing was performed on 76 ClinSeq[®] individuals with variants in these hexamers. Twenty-five of 64 genes with predominant PASH variants showed alternative polyadenylation, which suggests that variants in predominant PASH may lead to altered gene function. Toward pathogenicity classification of PASH variants, 30 PASH variants identified in HGMD were annotated for evidence that supports pathogenicity according to ACMG/AMP criteria adapted for polyadenylation variants. Sixteen variants had sufficient evidence to support a classification of pathogenic or likely pathogenic. We suggest that classification of PASH variants as pathogenic or likely pathogenic is limited by patient data and could be aided by *in vitro* functional studies not typically pursued for PASH variants. These data reinforce that PASH sequences are critical in polyadenylation and that variants in these sequences should be considered and assessed as candidates for pathogenic, disease-associated variation.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2622. Identification of susceptibility loci for frailty through a genome-wide scans and genetic association study of aging-related variants in Korean population

Authors:

Y. Kwak¹, S. Yang², M. Yuk¹, J. Lee¹, J. Youn², N. Song²; ¹Coll. of Pharmacy, Chungbuk Natl. Univ., Cheongju-si, Korea, Republic of, ²Coll. of Pharmacy, Chungbuk Natl. Univ., Cheongju, Korea, Republic of

Abstract Body:

BACKGROUND: Although frailty occurs with chronological aging with a state of increased physiologic vulnerability for adverse health outcomes, the genetic mechanisms underlying frailty were not understood. Genome-wide association studies (GWAS) have identified approximately 1,300 single-nucleotide polymorphisms (SNPs) for aging but their involvement in frailty has not been evaluated and GWAS of frailty has not identified the susceptibility SNPs with a nominal genome-wide significance. Thus, we performed GWAS to identify SNP of frailty and evaluated whether GWAS-identified aging susceptibility SNPs are associated with frailty in Korean population. **METHODS:** Among 8,840 study participants from the Korean Genome and Epidemiology Study (KoGES)_Ansan and Ansung study with genome-wide SNP data, 1,690 participants with frailty data which was defined as the short-physical performance battery (SPPB) score were selected. The 344,928 and 459,909 SNPs, genotyped by Affymetrix 5.0 and Korean Biobank Array (K-CHIP), were analyzed in GWAS, respectively. GWAS was performed to identify novel SNPs associated with frailty using multivariable logistic regression model after adjustment for age, sex, and history of diabetes and hyperlipidemia. The 283 previous GWAS-identified aging susceptibility SNPs ($P \leq 5 \times 10^{-8}$) were selected to evaluate the association with frailty. **RESULTS:** Among 1,690 study participants (age, years: mean=70.08, sd=6.14), frailty was observed in 89 (5.27%) participants. In the current GWAS, the most significantly associated SNPs with the frailty ($P < 1 \times 10^{-5}$) were rs10119203 (TTLL11, OR=3.88, 95% CI=2.17-6.94, $P=4.83 \times 10^{-6}$), rs17075127 (LOC107986479, OR=4.34, 95% CI=2.31-8.16, $P=5.13 \times 10^{-6}$), rs12532855 (HECW1, OR=10.36, 95% CI=3.77-28.47, $P=5.73 \times 10^{-6}$), rs6558938 (CSMD1, OR=5.68, 95% CI=2.64-12.22, $P=8.65 \times 10^{-6}$), rs117280191 (RBFOX1, OR=9.87, 95% CI=3.59-27.17, $P=9.39 \times 10^{-6}$), and rs16977369 (SLC39A11, OR=5.86, 95% CI=2.75-12.66, $P=6.60 \times 10^{-6}$). Among the 283 GWAS-identified aging susceptibility SNPs, 14 SNPs (rs1008084, rs1159806, rs2069837, rs2075650, rs2736100, rs3959554, rs4420638, rs4838595, rs6224, rs72738736, rs7412, rs7705526, rs8089807, and rs9386796) were associated with frailty ($P < 0.05$). **CONCLUSION:** Although the current study results did not satisfy the nominal genome-wide significance threshold ($P < 5 \times 10^{-8}$), we identified novel and strong genetic susceptibility loci associated with the frailty in Korea population. Furthermore, we found that 14 GWAS-identified aging susceptibility SNPs were also associated with frailty suggesting that aging and frailty may have shared biological mechanisms.

Molecular Effects of Genetic Variation Posters - Thursday

PB2623. Identifying context-specific genetically regulated expression changes in monocytes associated with resilience to Alzheimer's Disease

Authors:

Y. Mustafa¹, W. Bush², A. Naj³, J. Below⁴, T. Hohman⁵, L. Dumitrescu⁶; ¹Case Western Reserve Univ., Cleveland, OH, ²Case Western Reserve Univ, Cleveland, OH, ³Univ. of Pennsylvania, Philadelphia, PA, ⁴Vanderbilt Univ Med Ctr., Nashville, TN, ⁵Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Alzheimer's Disease (AD) is a complex, multifactorial disease and remains the most common form of dementia worldwide. Previous AD genetic research, particularly genome-wide association studies (GWAS), have focused on the identification of genetic risk variants; however, the effect of each variant is small and biologically difficult to interpret. Transcriptome-wide association studies (TWAS) aggregate genetic variants in a gene-based association approach, capturing greater variation and generating concrete mechanistic hypotheses bridging the association. We conducted a TWAS of circulating monocytes to test the hypothesis that variation in genetically regulated gene expression (GR_{EX}) alters their function and response during inflammation, leading to differences in AD pathogenesis. Monocyte genetic data (Fairfax et al, 2014) imputed against the Haplotype Reference Consortium (HRC r1.1) on the Michigan Imputation Server, and corresponding expression data of monocytes in a naïve state, 12 and 24 hours after LPS stimulation, and 24 hours after stimulation with IFN- γ from ArrayExpress served as our reference. We trained elastic-net regression models of GR_{EX} on the basis of the surrounding *cis*-regulatory SNP profile for all genes. We then imputed GR_{EX} levels of 5108 individuals from an AD resilience GWAS (Dumitrescu et al, 2020) based on summary statistics, and tested for association between GR_{EX} and AD using the MetaXcan software suite. The results of the *cis*-eQTL analysis showed concordance between our results and those of the Fairfax group, including variants with a context-specific direction of effect—illustrating the importance of studying these cells in an appropriate environment. The TWAS analysis revealed three genes with GR_{EX} significantly associated with cognitive and global resilience to AD symptoms: *CRISPLD2*, *NETO2*, and *DHRS9*, in IFN- γ , LPS-24h, and LPS-2h, respectively. We contribute evidence of potentially novel genetic associations between components of the innate immune system and AD in a time-, cell-, and context-specific manner. Further, we expand the available tissues and cell types for TWAS of AD by generating reference panels that account for the dynamic biology of monocytes in different states. These panels enable analyses that may have importance for Alzheimer's Disease, cancer, autoimmune conditions and more; moreover, they may provide concrete biological hypotheses as to how genetic variation can influence these complex phenotypes through monocyte activity.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2624*. Identifying functional variants related to heart development in African populations.

Authors:

N. Xie¹, Y. Feng¹, S. Fan^{2,1}, T. Nyambo³, S. W. Mpolok⁴, G. G. Mokone⁵, A. Njamnshi⁶, C. Folkunang⁷, D. Woldemeskel⁸, G. Belay⁸, S. Tishkoff^{1,9}; ¹Dept. of Genetics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ²State Key Lab. of Genetic Engineering, Human Phenome Inst., Zhangjiang Fudan Intl. Innovation Ctr., Sch. of Life Sci., Fudan Univ., Shanghai, China, ³Dept. of Biochemistry, Kampala Intl. Univ. in Tanzania, Dares Salaam, Tanzania, United Republic of, ⁴Dept. of Biological Sci., Faculty of Sci., Univ. of Botswana, Gaborone, Botswana, ⁵Dept. of BioMed. Sci., Univ. of Botswana Sch. of Med., Gaborone, Botswana, ⁶Dept. of Neurology, Central Hosp. Yaoundé; Brain Res. Africa Initiative (BRAIN), NeuroSci. Lab, Faculty of Med. and BioMed. Sci., The Univ. of Yaoundé I, Yaoundé, Cameroon, ⁷Dept. of Pharmacotoxicology and Pharmacokinetics, Faculty of Med. and BioMed. Sci., The Univ. of Yaoundé I, Yaoundé, Cameroon, ⁸Dept. of Biology, Addis Ababa Univ., Addis Ababa, Ethiopia, ⁹Ctr. for Global Genomics and Hlth.Equity, Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

African populations have high levels of genomic and phenotypic diversity. To increase our knowledge of African genomic diversity, we analyzed a novel high coverage whole genome sequencing dataset consisting of 180 individuals from 12 indigenous African populations and identified more than 30 million SNPs. We identified candidate loci that may be targets of local adaptation to diverse environments by using a population differentiation-based analysis (Di) to identify alleles showing highly differentiated frequencies between populations. We selected the top 0.1% of SNPs based on Di values (Di-SNPs) and studied genomic functional enrichment of nearby genes using GREAT (genomic regions enrichment of annotations tool). Interestingly, we observed an enrichment of SNPs near genes involved in heart development in the Hadza hunter-gatherers from Tanzania, possibly due to local adaptation. We observed that 9,325 (99.5%) of the 9,370 total Hadza Di-SNPs locate in non-coding regions, only 45 (0.5%) locate in coding regions, and no nonsynonymous variants are found in heart-related genes. To find putative regulatory variants associated with heart function, we overlapped Hadza Di-SNPs with genomic regions predicted to be functional regulatory regions, based on human heart-derived DNase-seq and ATAC-seq, as well as ChIP-seq for the cardiac transcription factors GATA4, TBX5, and NKX2.5. Intersection of these data sets identified Hadza Di-SNPs located within predicated enhancer regions close to genes important for heart development. Specifically, we identify a variant common only in the Hadza population (allele frequency of 46.7% in the Hadza but <6.7% in global populations) located in a predicted heart-specific enhancer of the HAND2 (heart and neural crest derivatives expressed 2) gene, a key transcription factor for cardiac morphogenesis. We demonstrate that this variant impacts enhancer activity in cardiomyocytes. We are employing CRISPR genome editing in human pluripotent stem cell-derived cardiomyocytes to determine the impact of this variant on cardiac development. This study demonstrates an approach for identifying functionally important variants that may play a role in local adaptation. Funded by NIH grant 1U01HG012047-01, 1R35GM134957, R01AR076241 and ADA 1-19-VSN-02.

Molecular Effects of Genetic Variation Posters - Thursday

PB2625. Impact of rare heterozygous mutations of *PCSK1* on obesity: implication for treatment with MC4R agonists.

Authors:

M. Baron¹, L. Folon¹, M. Derhourhi¹, B. Balkau², G. Charpentier³, S. Franc³, R. Roussel⁴, M. Canouil¹, P. Froguel^{1,5}, A. Bonnefond^{1,5}; ¹Inserm UMR1283, CNRS UMR8199, Lille, France, ²Inserm U1018, Villejuif, France, ³CERITD, Evry, France, ⁴Inserm U1138, Paris, France, ⁵Imperial Coll. London, London, United Kingdom

Abstract Body:

Background and aims: Pathogenic mutations in key genes involved in the leptin-melanocortin pathway are well-established causes of monogenic forms of obesity. Recently, the melanocortin-4 receptor (MC4R) agonist setmelanotide has been developed to efficiently treat obese patients with homozygous mutations in *POMC* and *LEPR*. Its relevance to other forms of monogenic obesity is currently under investigation. Among those forms is *PCSK1* deficiency. Bi-allelic pathogenic mutations in *PCSK1* lead to early-onset obesity associated with severe endocrinopathy, but the clinical impact of heterozygous pathogenic *PCSK1* mutations on obesity is still elusive. We performed large-scale functional genetic studies to clarify the link between heterozygous *PCSK1* mutations and obesity. **Material and methods:** All 14 coding exons of *PCSK1* were sequenced in 10,000 individuals (including obese and/or diabetic patients) by next-generation sequencing. The detected heterozygous variants were created by mutagenesis and inserted into plasmids. Functional assays of each variant were performed in HEK293 cells: enzymatic activity was analysed using a fluorescent PC1/3 substrate. **Results:** We identified 66 rare heterozygous variants of *PCSK1* (minor allele frequency < 1%). The variants were clustered into 5 groups (A-E) based on their enzymatic activity. The 17 variants in group A had complete loss of enzymatic activity whereas group B included 11 variants with partial loss of enzymatic activity. They strongly affect the structure of the protein preventing the expression of the mature active form of PC1/3. We found that class A variants increased the risk of developing obesity by 6-fold in both children ($P=0.097$; OR=6.0) and adults ($P=0.038$; OR=5.6). In contrast, class B variants were not associated with obesity risk ($P=0.86$; OR=1.1). Importantly, we observed a lack of sensibility and specificity of the largely used *in silico* tools predicting putative heterozygous pathogenic *PCSK1* mutations. Indeed, the pathogenic variants predicted by REVEL showed 4 false negatives and 11 false positives compared to the results of *in vitro* functional assays. **Conclusion:** Our results strongly suggest that only obese individuals carrying *PCSK1* heterozygous mutations proven pathogenic through *in vitro* functional testing (*i.e.* category A, 25%) should be eligible for putative setmelanotide treatment, but not the carriers of mutations with only intermediate *in vitro* deleterious effect. Therefore, *in vitro* tests are required following the molecular diagnosis of rare heterozygous *PCSK1* mutations before initiating expensive medication.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2626. Improving generalizability of gene expression prediction models from African-ancestry populations.

Authors:

E. Esquinca¹, M. Boorgula¹, C. H. Arehart², M. Campbell¹, S. Chavan², N. Rafaels², C. Cox¹, A. Greenidge³, P. Maul¹, T. Maul¹, D. Walcott³, T. Brunetti¹, I. Ruczinski⁴, K. Kammers⁵, H. Watson³, R. Landis³, M. A. Taub⁴, M. Daya¹, E. Kenny⁶, R. A. Mathias⁴, C. R. Gignoux¹, K. Kechris¹, K. C. Barnes¹, R. K. Johnson¹; ¹Univ. of Colorado, Anschutz Med. Campus, Aurora, CO, ²Univ. of Colorado, Aurora, CO, ³Univ. of the West Indies, Bridgetown, Barbados, ⁴Johns Hopkins Univ., Baltimore, MD, ⁵Johns Hopkins Univ., Sch. of Med., Baltimore, MD, ⁶Icahn Sch. of Med. at Mt Sinai, New York, NY

Abstract Body:

Prior gene expression prediction studies demonstrated that differences in out-of-sample prediction across populations of different ancestries were related to genetic differentiation between source populations. Therefore, we aimed to quantify the relationship between population genetic differentiation and gene expression prediction performance to develop more generalizable gene expression prediction models. Previously, we used the PrediXcan suite of tools to create a database of SNPs that contribute to significant prediction of unstimulated CD4+T cell gene expression among 260 African-ancestry individuals participating in the Barbados Asthma Genetics Study. This database was used to predict CD4+T gene expression in subjects from the 1000 Genomes Project (TGP), of African (AA: YRI, N=86) and European (EA: CEU, FIN, & GRB N=276) ancestry who also have cell line RNA sequencing measures. Prediction accuracy (R²) for each gene (N=3,279) was calculated as the square of the Pearson correlation coefficient between predicted versus observed values. We tested whether prediction accuracy differed across AA and EA populations using a Wilcoxon rank sum test. Next, we used the R package KRIS to summarize the pairwise genetic fixation index (F_{st}) for each SNP between AA and EA populations in TGP. We averaged the F_{st} for all SNPs used to predict a gene's CD4T expression as a measure of population differentiation and tested whether average F_{st} was associated with prediction performance using Pearson's correlation. Average prediction performance for the 3,227 genes was like prior studies applying prediction models developed in one ancestry to an out- f-ancestry or multi-ancestry population, with median R²=0.00631 (IQR=0.001, 0.027). Prediction performance was significantly better (p<0.001) in the AA population with a median R²=0.0155 (IQR=0.003, 0.052, max=0.91) compared to the EA population median R²=0.00597 (IQR=0.001, 0.026, max=0.82). The mean F_{st} of SNPs used to predict CD4+T gene expression was not significantly associated with increased difference in prediction performance between AA and EA (rho=0.0166, p=0.35), and no association with prediction performance among the AA or EA population (p>0.05). We showed that prediction accuracy of gene expression from SNP data was increased among subjects from the same ancestry as the model training dataset compared to the out-of-ancestry subjects, but this difference in prediction accuracy was not significantly association with mean F_{st}. Ongoing work evaluates the improvement in generalizability of gene expression prediction models through restriction of eligible SNP predictors by ancestry specific MAF.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2627. Inhibiting aurora kinase A in the HMC3 immortalized human microglia cell line increases phagocytosis of amyloid beta *in-vitro*.

Authors:

K. Cook¹, S. H. Littleton², J. A. Pippin¹, E. A. Burton³, S. A. Anderson¹, A. Chesi², S. F. A. Grant⁴, ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Univ. of Pennsylvania, Philadelphia, PA, ³The Univ. of Pennsylvania, Philadelphia, PA, ⁴Children's Hosp. of Philadelphia/Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Alzheimer's Disease (AD) represents the leading cause of dementia worldwide. Despite its increasing prevalence and burden on public health, there are still not widely available and efficacious therapeutics. This relative dearth of available pharmaceutical interventions is a result of the multi-factorial nature of AD associated neurodegeneration. Among the key contributors to AD pathology is neuroinflammatory processes, and central to those mechanisms are the microglia, representing the central nervous system's resident macrophage. Integrating reported AD GWAS statistics with our combined ATAC-seq/high-resolution promoter-focused Capture-C datasets has implicated microglial genes in the pathophysiology of the disease. Among the targets resulting from our variant-to-gene efforts for this disease is the gene encoding Aurora Kinase A (*AURKA*). *AURKA* is a serine/threonine kinase involved in NFkB signaling, as well as involved in the promotion of M1 pro-inflammatory polarization in macrophages. To investigate the effects of *AURKA* inhibition on key aspects of AD pathology we have obtained a research grade inhibitor (MK8475) and applied it to the human microglial HMC3 model system. Cells exposed to MK8475 following stimulation with AB displayed a 30% ($P=0.0033$) increased uptake of fluorescently conjugated Amyloid Beta (AB_{42}). Following up on this functional phenotype we sought to characterize the transcriptome of cells treated with MK8475 using RNA-seq. We found key GO analysis terms associated with the gene module upregulated in cells treated with the inhibitor after exposure to AB_{42} to be 'organonitrogen compound metabolic processes' ($P=1.5 \times 10^{-6}$), 'interferon type I response' ($P=1.63 \times 10^{-7}$), and 'cellular response to cytokine signaling' ($P=1.19 \times 10^{-11}$). Indeed, these categories have been previously implicated as dysregulated in AD pathology. Additionally, the transcriptomic profile of drug treated cells displayed upregulation of genes previously published as enriched in the transcriptomic profile of microglia actively phagocytosing fibrillar AB (Grubman *et al.* 2021). Taken together, results from a FLOW cytometry-based phagocytosis assay, as well as transcriptomic profiling, reveal that inhibition of *AURKA* is a promising therapeutic avenue for promoting microglial phenotypes and holds promise for ameliorating AD pathology.

Molecular Effects of Genetic Variation Posters - Thursday

PB2628. Integrated analysis of retinal eQTLs, meQTLs and eQTM identifies target genes and epigenetic mechanisms contributing to age-related macular degeneration.

Authors:

J. Advani¹, A. R. Hamel^{2,3}, S. Mehrotra^{2,3}, C. Kiel⁴, T. Strunz⁴, J. Rouhana^{2,3}, M. Kwicklis¹, F. V. Asten¹, R. Ratnapriya¹, X. Corso-Díaz¹, D. G. Hernandez⁵, R. J. Kappahn⁶, D. Ferrington⁶, B. Weber^{4,7}, A. V. Segre^{2,3}, A. Swaroop¹; ¹Neurobiology, Neurodegeneration and Repair Lab., Natl. Eye Inst., NIH, Bethesda, MD, ²Ocular Genomics Inst., Dept. of Ophthalmology, Massachusetts Eye and Ear, Harvard Med. Sch., Boston, MA, ³The Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Inst. of Human Genetics, Univ. of Regensburg, Regensburg, Germany, ⁵Lab. of Neurogenetics, Natl. Inst. of Aging, NIH, Bethesda, MD, ⁶Dept. of Ophthalmology and Visual NeuroSci.s, Univ. of Minnesota, Minneapolis, MN, ⁷Inst. of Clinical Human Genetics, Univ. Hosp. Regensburg, Regensburg, Germany

Abstract Body:

Age related macular degeneration (AMD) is a multifactorial progressive neurodegenerative disease and a leading cause of blindness in people with advanced age globally. A large genome-wide association study (GWAS) has previously identified 52 genetic variants at 34 AMD susceptibility loci (Fritsche et al., *Nat. Genet.* 2016). However, a vast majority of associated variants localize to non-coding genome and causal genes remain largely undetermined at most loci. Our previous retinal eQTL and TWAS analyses have uncovered several candidate genes in AMD GWAS loci (Ratnapriya et al., *Nat. Genet.* 2019; Strunz et al., *PLoS Genet.* 2020). However, complex traits, such as AMD, are affected by environmental and genetic factors, epigenetic mechanisms and their genetic regulation can help dissect the impact of aging and environment on disease onset and severity. In this study, we defined the inter-relationship among gene expression, genetic variability, and DNA methylation in the human retina and investigated their potential causal effect on AMD. We performed DNA methylation profiling of postmortem retina from 152 controls and AMD cases using Illumina MethylationEPIC bead chip. Applying QTLtools to methylation and imputed genotype data, we identified 37,453 variants associated in *cis* (<1 Mb) with CpG sites (methylation quantitative trait loci, meQTLs) (FDR≤0.05). Gene ontology analysis of mGenes identified genes involved in regulation of GTPase activity, actin filament and synaptic signaling. Applying colocalization analyses with eCAVIAR and Summary data-based Mendelian Randomization (SMR) between meQTL and eQTL or AMD GWAS signals have revealed meQTLs that may impact 9 disease-associated loci; e.g., meQTLs mapped to *CFH*, *ARMS2/HTRA1* or *LINC01004*, and 10,246 meQTLs that may mediate expression of 2,637 genes, e.g., *NLRP2* or *NDUFA10*. Furthermore, integration of expression with methylation identified 13,747 expression quantitative trait methylation (eQTM) (FDR≤0.05). Notably, eQTM target genes were enriched in oxidative phosphorylation, electron transport chain and translational initiation. Thus, our study, for the first time, describes relationship between genetic regulation and DNA methylation in the retina, and association of meQTLs with AMD. The high correlation of genetic effects between methylation and gene expression highlights the target genes that may mediate the effect of aging and/or environmental factors on AMD. Future incorporation of age, environment or lifestyle factors, such as smoking, to meQTL and eQTM detection, can potentially augment understanding of complex interplays among divergent components contributing to AMD pathology.

Molecular Effects of Genetic Variation Posters - Thursday

PB2629. Integrative 3D structural analysis of *de novo* missense variants and their associated protein structures inferred by AlphaFold2 identifies the importance of β -strand/sheet features for congenital heart disease

Authors:

Y. Xie¹, Y. Yang², C. Moth³, B. Li³, J. Gu¹, W. Dong¹, J. Capra⁴, S. Jin⁵, W. Zheng², H. Zhao¹; ¹Yale Univ., New Haven, CT, ²Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ³Vanderbilt Univ., Nashville, TN, ⁴Univ. of California San Francisco, San Francisco, CA, ⁵Washington Univ. Sch. of Med., St. Louis, MO

Abstract Body:

Multiple studies have been conducted to characterize the association between missense variants and protein structures using data from Protein Data Bank (PDB). However, only 35% of human proteins map to a PDB entry, and in many cases the structure covers only a portion of the protein sequences, not the whole proteins. The release of protein structures from AlphaFold2 (AF2) covers 98.5% of human proteins, with 58% of them with high confidence.

In this study, we surveyed the associations between congenital heart disease (CHD) *de novo* missense variants and structural features using protein structures from both PDB and AF2. Compared with *de novo* missense variants from healthy controls, we identified a significant association of β -strand/sheet (a three-class secondary structure) with CHD *de novo* missense variants after multiple testing adjustments (marginal $p=0.0019$). Structural features including α , π , 3_{10} -helix and core (relative solvent accessible area $<5\%$) are marginally significant (marginal $p<0.05$). If we only include PDB structures in the analysis, all features are not significant.

Under the assumption that the AF2 predicted structures included in our analysis are accurate, our preliminary results 1) show the increased power of extending protein structure information using AF2; 2) call upon novel statistical methods that can integrate protein structural features with *de novo* variants; and 3) may shed new light on the relationship between molecular effects of *de novo* missense variants and CHD.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2630*. Investigating the impact of non-coding structural variants flanking epilepsy-associated gene, *FOXG1* on chromatin architecture and gene regulation.

Authors:

A. Ramamurthy¹, J. Calhoun¹, M. Bandouil¹, M. Nobrega², G. Carvill¹; ¹Northwestern Univ., Chicago, IL, ²Univ. of Chicago, Chicago, IL

Abstract Body:

Structural variants (SVs) are genetic variants ranging from 50 bases to a few megabases in size and include deletions, duplications, insertions, inversions and translocations. Epilepsy is a heterogeneous neurological disorder characterized by recurrent seizures. Up to 12% of genetic epilepsies can be explained by SVs that disrupt the coding regions of the genome, causing epilepsy through gene dosage effects. SVs in non-coding genomic regions can perturb *cis*-regulatory elements (CREs) and/or topologically associated domains (TADs). These loci are important for 3D genome architecture and gene regulation. Whether SVs in these loci cause epilepsy and the associated putative pathogenic mechanisms are unknown. The goal of this study is to define the molecular mechanisms by which SVs in non-coding regions influence epilepsy susceptibility. We identified microdeletions in the *14q12* genomic region flanking the gene, *FOXG1*, in individuals with epilepsy and associated neurodevelopmental disorders. This gene encodes the Forkhead Box G1 protein, a transcription factor essential for mammalian cerebral corticogenesis, and *FOXG1* haploinsufficiency is associated with epilepsy. We hypothesized that the *14q12* microdeletions flanking *FOXG1* alter gene expression. We utilized induced pluripotent stem cell (iPSC) derived neuronal models for our studies, as *FOXG1* expression is highest in neuronal cell types. We introduced the microdeletions via CRISPR based genome editing to create stable iPSC lines and reprogrammed the iPSCs to neural progenitor cells (NPCs). Our preliminary findings demonstrate reduced *FOXG1* expression and protein abundance in NPCs harboring the microdeletion downstream of *FOXG1*, compared to our isogenic controls. Additionally, during the selection of CRISPR edited iPSCs, we identified iPSC clones which harbored an inversion of the region of the intended microdeletion downstream of *FOXG1*. Interestingly, NPCs derived from this iPSC line demonstrate increased *FOXG1* expression. Further validation of these findings is currently ongoing. Future steps will involve evaluating TAD architecture and CRE interactions as well as assessing changes in neuronal phenotypes in the presence of the proposed microdeletions. This study will be the first to explore molecular mechanisms of non-coding SVs associated with epilepsy. A comprehensive understanding of genome wide SVs in the non-coding regions associated with epilepsy will enable the identification of misregulated genes that can function as potential targets of pharmacological agents and aid the development of therapy for those individuals with active genetic epilepsies.

Molecular Effects of Genetic Variation Posters - Thursday

PB2631. Investigating the molecular and cellular effects of pathogenic variants of *GNAI1* in developmental and epileptic encephalopathy

Authors:

E. Fathi¹, H. Chen², E. Almanza Fuerte¹, H. Mefford³; ¹St. Jude Children's Res. Hosp., Memphis, TN, ²St. Jude Children's Res. Hosp., Memphis, TN, ³St. Jude Children's Res. Hosp., Memphis, TN

Abstract Body:

Developmental and epileptic encephalopathies (DEE) are characterized by intellectual disability, significant developmental delay, intractable seizures, and autistic features. Recently our group identified 16 unique *de novo* mutations in *GNAI1* (G protein subunit alpha I1) in 24 individuals with DEE. *GNAI1* encodes $G\alpha_{i1}$, a G-protein subunit that binds to and hydrolyzes GTP and interacts with a variety of downstream effectors in signaling cascades inside the cell. However, the effect of the identified variants on $G\alpha_{i1}$ functions such as GDP/GTP binding, GTP hydrolysis, and affecting downstream signaling pathways is not known. Although *GNAI1* is predicted to be intolerant to loss-of-function variants, only a single disease-causing variant has been studied to date. Therefore, there is an urgent need to determine the molecular and cellular consequences of disease-causing variants and develop a cellular model in which to study disease and test potential therapies. To investigate the molecular and cellular function of *GNAI1* disease-causing variants, we selected four pathogenic mutations, each found in two or more patients, including three missense variants in different positions of the GDP binding domain (Gly40Cys, Thr48Lys, Lys270Arg), and one single amino acid deletion just outside of the GDP binding domain (Gln172del). We designed a series of *GNAI1* mutations in pcDNA3.1 vector with a flag tag using site-directed mutagenesis. We also designed one early termination (Thr48Ter) as a control for haploinsufficiency and loss of function effects. In addition, we are generating CRISPR-engineered cells with the same selected variants and have patient cells for three variants. We are investigating the effect of disease-related $G\alpha_{i1}$ variants on protein localization, protein-protein interactions, GTP binding and downstream pathways. We transiently transfected HEK-293T cells with each mutant construct (Gly40Cys, Thr48Lys, Lys270Arg, Gln172del, Thr48Ter). We showed that the protein expression of all five *GNAI1* mutant constructs is significantly decreased compared with wild type (WT) control in HEK-293T cells. Preliminary results suggest that the Lys270Arg and Gln172del variants disrupt protein localization compared with control. Our engineered and patient-derived cells represent a valuable tool for studying the role of *GNAI1* mutations in the pathogenesis of $G\alpha_{i1}$ mutant proteins. We will validate our preliminary results from *in vitro* studies in 3D brain organoids from patient cells, which will serve as a beneficial disease model to investigate molecular and cellular function of $G\alpha_{i1}$ variants and to test promising therapeutic strategies for *GNAI1*-related DEE.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2632. Investigating the potential roles of Importin-7 and Importin- β in LINE-1 retrotransposition.

Authors:

S. Mortimer, J. V. Moran; Univ. of Michigan Med. Sch., Ann Arbor, MI

Abstract Body:

Sequences derived from Long Interspersed Element-1 (LINE-1 or L1) retrotransposons comprise ~17% of human genomic DNA. The L1 retrotransposition cycle begins with transcription of a full-length bicistronic L1 RNA, which exits the nucleus and undergoes translation to generate the L1-encoded proteins (ORF1p and ORF2p). Both ORF1p, an RNA binding protein, and ORF2p, which has endonuclease and reverse transcriptase activities, preferentially associate with their encoding RNA transcript to form a ribonucleoprotein (RNP) particle by a process known as cis-preference. Components of the L1 RNP then must gain access to genomic DNA, where the ORF2p endonuclease activity generates a single strand endonucleolytic nick in genomic DNA to expose a 3'-hydroxyl group (3'-OH) that is used as a primer by the ORF2p reverse transcriptase activity to initiate the reverse transcription of L1 RNA. However, how the L1 RNP enters the nucleus requires elucidation. A 2006 study (Kubo et al. PNAS) demonstrated that L1 retrotransposition can occur in G1/S arrested cells, indicating that components of the L1 RNP can cross an intact nuclear membrane. Additionally, a 2019 study (Miyoshi et al. Molecular Cell), which used an immunoprecipitation-mass spectrometry screen in HEK293T cells, identified the nuclear import proteins Importin-7 (Imp-7) and Importin- β (Imp- β) as putative L1 ORF2p interaction partners. Here, I describe progress toward investigating whether Imp-7 and/or Imp- β might play a role in L1 retrotransposition. My preliminary results suggest that Imp-7 siRNA knockdown leads to a modest reduction in retrotransposition as compared to mock siRNA treatment in HEK293T cells, but not in HeLa-JVM cells. Ongoing experiments will determine whether Imp- β knockdown has a similar effect on L1 retrotransposition and whether co-immunoprecipitation experiments provide evidence for ORF2p/Imp-7 and/or ORF2p/Imp- β interactions.

Molecular Effects of Genetic Variation Posters - Thursday

PB2633. KDM5A, a chromatin remodeler and newly identified autism gene, has a cell type-specific function in the hippocampus.

Authors:

L. El Hayek, D. DeVries, K. Kaur, M. Chahrour; Univ. of Texas Southwestern Med. Ctr., Dallas, TX

Abstract Body:

Autism spectrum disorder (ASD) is a constellation of neurodevelopmental disorders with high clinical and genetic heterogeneity. It is characterized by impaired communication skills, social behavior deficits, restricted interests, and repetitive behaviors. Although 1 in 44 children are diagnosed with ASD, the causative genes that have been identified to date account for only 30% of cases, with no single gene individually contributing to more than 1% of cases. Therefore, the genetic architecture of ASD suggests that it is a collection of individually rare disorders. Designing targeted therapeutics for these rare genetic subtypes of ASD will require mechanistic understanding of the disrupted pathways. Importantly, insights into the normal function of ASD genes continue to inform the mechanisms underlying neurodevelopment. Epigenetic chromatin regulation is essential for proper gene expression and for brain development, and is a key pathway disrupted in ASD with causative mutations identified in several genes encoding chromatin remodelers (e.g. *ARID1B*, *CHD8*, *SETD5*). We identified a novel ASD gene, *KDM5A*, and showed that its loss results in neurobehavioral abnormalities and transcriptional dysregulation in the cortex and hippocampus. *KDM5A* is a chromatin regulator that belongs to the *KDM5* family of lysine-specific histone H3 demethylases. Consistent with its function, we found that loss of *KDM5A* results in global increase in trimethylated H3 lysine 4 (H3K4Me3) in the cortex. Thus, evidence from our studies suggests that *KDM5A* is critical for brain development and function. It has been shown that ASD genes are enriched in specific cell types in the brain during development and that changes in gene regulatory networks in ASD occur in a cell type-specific manner in the brain. *KDM5A* is ubiquitously expressed across cell types in the brain, its expression pattern throughout development is dynamic, and it is differentially expressed at different stages of cell lineage birth and maturation, indicating that *KDM5A* function in the brain is likely to be cell type-specific. We performed single-nuclei isolation and RNA sequencing from hippocampal tissue from wildtype and *Kdm5a* knockout mice and identified specific subtypes of excitatory and inhibitory neurons that are particularly vulnerable to *KDM5A* loss of function. These cell types have common and distinct transcriptional networks that are disrupted upon loss of *KDM5A*. Our findings define a cell type-specific role for *KDM5A* in the hippocampus and more broadly, advance our knowledge of the role chromatin regulation plays in brain development and function.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2634. KMT5B is required for early motor development.

Authors:

H. Feser Stessman, J. Hulen, D. Kenny, R. Black, J. Hallgren, K. Hammond, E. Bredahl, R. Wickramasekara, P. Abel; Creighton Univ., Omaha, NE

Abstract Body:

Heterozygous disruptive variants in lysine methyl transferase 5B (*KMT5B/SUV420H1*) have been identified as likely-pathogenic among humans with developmental delays, intellectual disability, and autism spectrum disorder phenotypes. Interestingly, this patient population presents with high rates of motor deficits (i.e., hypotonia and motor delays); however, the role that KMT5B plays in early motor development is largely unknown. Using a mouse gene trapping strategy, we have previously reported early motor reflex deficits and reduced body size in the germline *Kmt5b* heterozygous (HET) state compared to wild-type (WT) littermates. The goal of this study was to further characterize the developmental motor phenotype in the HET condition to better understand the effects of KMT5B haploinsufficiency on skeletal muscle growth and maturation. Using our gene trap model, we assessed neuromuscular strength, skeletal muscle weight (i.e., muscle mass), neuromuscular junction (NMJ) structure, and myofiber type, size, and distribution in both sexes and at two developmental times (postnatal days P17 and P44) to compare juvenile to adult skeletal muscle structures. Data were collected in hindlimb soleus (SOL) and extensor digitorum longus (EDL) muscles to compare slow- to fast-twitch muscle types, respectively. Prior to the onset of puberty at P17, slow-twitch SOL muscle weight was significantly reduced in HET compared to WT males ($p = 0.0072$; Tukey's test) but not females suggesting sexual dimorphism at this developmental stage. At the young adult stage (P44), we identified decreased neuromuscular strength ($p = 0.0367$; 2-way ANOVA) and increased NMJ fragmentation in the SOL ($p = 0.0157$; 2-way ANOVA) in HET animals. Further, skeletal muscle weights were significantly decreased at P44 in both the HET SOL ($p = 0.0015$; 2-way ANOVA) and EDL ($p < 0.0001$; 2-way ANOVA) muscles. HET muscle weight remained lower than WT even after correcting for known body weight differences; this was not found among other body organs suggesting skeletal muscle specificity. Smaller HET muscles were composed of a significantly increased numbers of fast-twitch myofibers, yet the overall size of myofibers was significantly smaller in P44 muscles ($p < 0.05$; 2-way ANOVA). We conclude that KMT5B haploinsufficiency results in a skeletal muscle developmental deficit causing reduced muscle mass and body weight (i.e., hypotonia) that phenocopies the KMT5B patient population.

Molecular Effects of Genetic Variation Posters - Thursday

PB2635. Large-scale integrative analysis of 1.5 million single cells across diverse human tissues and immune cell types reveals dynamic regulatory effects on *HLA* gene expression.

Authors:

J. Kang^{1,2,3}, A. Z. Shen^{1,2,3}, Y. Luo^{1,2,3}, S. Sakaue^{1,2,3}, S. Gurajala^{1,2,3}, A. Nathan^{1,2,3}, V. R. C. Aguiar^{2,3,4}, C. Valencia^{1,2,3}, L. Rumker^{1,2,3}, F. Zhang^{1,2,3}, A. Jonsson^{1,2}, S. Yazar⁵, J. Alquicira-Hernandez⁵, K. Dey^{6,3}, K. Jagadeesh^{6,3}, Accelerating Med. Partnership (AMP): Rheumatoid Arthritis and Systemic Lupus Erythematosus Network, J. E. Powell⁵, M. Gutierrez-Arcelus^{2,3,4}, D. A. Rao^{1,2}, L. Donlin^{7,8}, J. Anolik⁹, M. B. Brenner^{1,2}, S. Raychaudhuri^{1,2,3}; ¹Brigham and Women's Hosp., Boston, MA, ²Harvard Med. Sch., Boston, MA, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Boston Children's Hosp., Boston, MA, ⁵Garvan Inst. of Med. Res., Sydney, NSW, Australia, ⁶Harvard T. H. Chan Sch. of Publ. Hlth., Boston, MA, ⁷Hosp. for Special Surgery, New York, NY, ⁸Weill Cornell Med., New York, NY, ⁹Univ. of Rochester Med. Ctr., Rochester, NY

Abstract Body:

The major histocompatibility (MHC) region on chromosome 6, containing the human leukocyte antigen (*HLA*) and ~260 other genes, has the strongest genetic associations with immune-mediated diseases. However, the exact molecular mechanisms behind MHC disease risk are yet unsolved. Previous research has extensively explored how *HLA* coding variation contributes to disease risk, but altered *HLA* expression might also confer disease risk, for example through higher levels of antigen presentation to autoreactive T cells.

--

Here, we hypothesize that MHC genetic variants modulate *HLA* expression in a cell-state-dependent manner. Single-cell RNA-seq enables modeling expression quantitative trait loci (eQTLs) at the single-cell level in disease-relevant tissues and dynamic cell states. To this end, we collected and uniformly processed genotype and transcriptional data from four existing datasets in total comprising >1,495,000 single-cell profiles from 1,073 individuals: inflamed synovial tissue from rheumatoid arthritis patients and controls (n=69), inflamed intestinal tissue from ulcerative colitis patients and controls (n=22), peripheral blood mononuclear cells (PBMCs) treated *in vitro* with influenza A and controls (n=73), and PBMCs from a population cohort (n=909).

--

Because standard expression quantification pipelines are confounded by the polymorphic nature of *HLA* genes, we developed a personalized read alignment strategy that imputes 4-digit classical *HLA* alleles (*HLA-A, B, C, DPB1/B1, DQA1/B1, DRB1*) per individual then quantifies single-cell expression using a personalized genome. We find that *HLA* class II expression is more variable across cell states within a given cell type compared to class I and is present in both professional antigen-presenting cells as well as other cell states, such as proliferating T cells. We then performed a single-cell eQTL analysis across the MHC to identify variants regulating classical *HLA* gene expression in myeloid, B, and T cells. First, we aggregated cohorts together to increase power and used a pseudobulk linear model, identifying eQTLs for all eight *HLA* genes in every cell type (all lead P-vals < 4e-9). Then, we demonstrated multiple cases where *HLA* regulatory variants are cell-state-dependent by modeling the eQTLs at single-cell resolution using a negative binomial mixed effects model that includes interaction terms with continuously defined cell states. For example, a T cell *HLA-DQA1* eQTL (rs3104371) becomes strongest in cytotoxic T cells (interaction P = 6e-322). This work offers a large-scale, cross-tissue atlas of *HLA* expression and its genetic regulation across diverse disease-relevant cell states.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2636. Leveraging Mosaic Epileptogenic Human Brain Tissue to Study Cell-type Specific Transcriptional Changes Associated with a Pathogenic *PIK3CA* Variant

Authors:

M. Gade, D. Lai, E. Heinzen; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract Body:

Post-zygotically acquired mutations that arise during embryonic cortical development can lead to abnormal neuron morphology and/or migration defects, both of which can often result in seizures and developmental disabilities. Somatic mutations that cause abnormal activation of proteins in the PI3K-AKT-mTOR signaling pathway have been implicated in focal cortical dysplasia type 2 and hemimegalencephaly (HMEG). Surgically resected mosaic brain tissue from affected individuals offers a unique opportunity to determine the burden of mutations and transcriptional effects of the mutations across cell types. To tease apart the transcriptional effects of the mutation within cell types, we modified the G&Tseq¹ protocol to simultaneously perform single-cell genotyping and transcriptomic profiling to identify the comprehensive transcriptomic signature of mutation-positive and mutation-negative nuclei. We benchmarked the protocol by comparing nuclei isolated from surgically resected brain tissue of a patient with HMEG harboring a pathogenic somatic variant in *PIK3CA* (E545K) to an age-matched control. We used a fluorescently labeled NeuN antibody to sort neuronal and non-neuronal nuclei for bulk (n=20,000 nuclei/total nuclei) and single-cell (n=44 NeuN+ nuclei/sample) preparations. RNA extracted from bulk (NeuN+ and NeuN-) and single NeuN+ nuclei for both cases and controls was sequenced at a depth of 15 million (MM) reads/sample. In parallel, DNA was genotyped in mutation-positive bulk and single nuclei using droplet digital PCR and TaqMan genotyping assay, respectively. A variant allele fraction of 30% (NeuN+) and 35% (NeuN-) was observed in bulk samples from the mosaic *PIK3CA* tissue, which correlated with that observed in single nuclei. RNA from bulk nuclei populations consistently yielded ~10 MM uniquely mapped reads (UMR) while RNA from single nuclei yielded ~5-10 MM UMR. Transcriptomic analysis of bulk nuclei (NeuN- & NeuN+) and single nuclei (NeuN+) confirmed the activation of the mTOR pathway. We also observed transcriptomic signature associated with *PIK3CA* variants specific to NeuN+ bulk and single nuclei; namely, the CREB phosphorylation pathway and an upregulation of *SPDYE5*, a gene implicated in low-grade glioma. Analyses of genotype-informed single nuclei RNAseq are currently underway. This approach will allow for novel insights into the impact of pathogenic somatic mutations on cell-autonomous and non-cell-autonomous functions, high-resolution assessments of cell-type-specific mutation burden, and may help identify antiseizure targets for treating mTOR related intractable seizures.

1. Macaulay IC, et al. Nat Protoc. 2016

Molecular Effects of Genetic Variation Posters - Thursday

PB2637. Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery

Authors:

H. Mostafavi, J. Spence, S. Naqvi, J. Pritchard; Stanford Univ., Stanford, CA

Abstract Body:

Most signals in genome-wide association studies (GWAS) of complex traits point to noncoding genetic variants with putative gene regulatory effects. However, currently identified expression quantitative trait loci (eQTLs) explain only a small fraction of GWAS signals. By analyzing GWAS hits for complex traits in the UK Biobank, and cis-eQTLs from the GTEx consortium, we show that these assays systematically discover different types of genes and variants: eQTLs cluster strongly near transcription start sites, while GWAS hits do not. Genes near GWAS hits are enriched in numerous functional annotations, are under strong selective constraint and have a complex regulatory landscape across different tissue/cell types, while genes near eQTLs are depleted of most functional annotations, show relaxed constraint, and have simpler regulatory landscapes. We describe a model to understand these observations, including how natural selection on complex traits hinders discovery of functionally-relevant eQTLs. Our results imply that GWAS and eQTL studies are systematically biased toward different types of variants, and support the use of complementary functional approaches alongside the next generation of eQTL studies.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2638. Long-chain fatty acid oxidation disorder gene variants identified through a gene panel sponsored program: a diverse gene variant landscape.

Authors:

V. Rangel Miller¹, H. McLaughlin², D. Marsden¹, O. K. Japalaghi¹, A. Willcock³, N. Miller⁴; ¹Ultragenyx Pharmaceutical, Novato, CA, ²Invitae, South Lyon, MI, ³Invitae Corp., San Francisco, CA, ⁴Ultragenyx, Brisbane, CA

Abstract Body:

Long-chain fatty acid oxidation disorders (LC-FAOD) are rare, life-threatening, autosomal recessive conditions that impair the utilization of fats for energy production. Undiagnosed LC-FAOD may present with hypoglycemia, cardiomyopathy, cardiac arrhythmias, and neuromuscular signs. **Methods:** Patients in the US, Canada, and Mexico of all ages who have either a clinical diagnosis or suspicion of LC-FAOD with a confirmatory acylcarnitine test ordered or performed are eligible for this no-charge genetic testing program. The next generation sequencing gene panel with copy number variant (CNV) detection includes the 6 genes associated with LC-FAOD (*ACADVL*, *CPT1A*, *CPT2*, *HADHA*, *HADHB*, *SLC25A20*) plus 19 additional genes associated with disorders that cause a similar acylcarnitine profile. **Results:** 561 patients were tested as of 30 MAR 2022 and 244 LC-FAOD gene variants were identified among 196 patients: 150 (61%) pathogenic/likely pathogenic (P/LP), 94 (39%) variant of uncertain significance (VUS). 4 patients had variants in 2 or more LC-FAOD genes. Additional evidence, including segregation, clinical/biochemical information, or functional modeling, led to the reclassification of 8 *ACADVL* VUS to P (3) or LP (5), impacting 11 individuals. 50 patients had a positive (n=30 P/P, LP/P, LP/LP) or potential positive (n=20 LP/VUS, P/VUS) LC-FAOD genetic diagnosis: 60% *ACADVL*, 16% *CPT2*, 8% *HADHA*, 4% *CPT1A*, 4% *HADHB*, 2% *SLC25A20*. Clinical signs reported for 21 of these patients include elevated creatine kinase (14), rhabdomyolysis (10), myopathy (8), myoglobinuria (4), cardiomyopathy (3), retinitis pigmentosa (3), hypoketotic hypoglycemia (1), and peripheral neuropathy (1). While most (56% of 561) patients tested were under 1 year (y) of age, the rate of LC-FAOD genetic diagnoses (50) varied with age: 9.9% under 1y, 4.2% 1y-12y, 15% 13y-20y, 9.7% 21y-40y, and 5% 40y or older. Among non-LC-FAOD panel genes there were 29 positive, 18 potential positive, and 6 uncertain (VUS only) genetic diagnoses. Of 76 patients with a positive confirmatory acylcarnitine test, 18 had a positive/potential positive LC-FAOD diagnosis; of 197 patients with an inconclusive confirmatory test, 12 had a positive/potential positive LC-FAOD diagnosis. **Conclusion:** These results demonstrate a diverse composition of gene variants in patients with suspected LC-FAOD and suggest the need for further approaches to resolve VUS and identify previously undetected variants. The higher LC-FAOD genetic diagnosis rates in adolescents/adults suggest these patients may have missed newborn screening, consistent with the timing of LC-FAOD addition to the Recommended Uniform Screening Panel in 2008.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2639. Mapping the effects of genomic deletions and duplications on cognitive ability across cortex

Authors:

S. Kazem¹, K. Kumar¹, G. HUGUET¹, J. Kopal¹, E. Douard², M. Jean-Louis¹, Z. Saci¹, L. Almasy³, D. Glahn⁴, G. Dumas², S. Jacquemont¹; ¹CHU Sainte-Justine, Montreal, QC, Canada, ²Université de Montréal, Montreal, QC, Canada, ³Children`s Hosp. of Philadelphia, Philadelphia, PA, ⁴IOL & Yale, Hartford, CT

Abstract Body:

Genomic deletions or duplications, copy number variants (CNVs), increase the risk for psychiatric disorders. Recent studies based on genetic constraint score (LOEUF) estimate that close to 50% of coding genes show measurable effects on cognitive ability when deleted or duplicated. However, functional prioritization is problematic because this broad spectrum of genes impacting cognition involves most annotated GO-term categories. In addition, cognition is not the direct product of these microscale functions; it emerges from large scale brain networks that structure the interaction of molecular and cellular functions. We hypothesize that spatial patterns of gene expression across large scale brain networks - that can be seen as macro-scale meta functions - will shed light on how such a broad group of genes impact cognitive ability when altered by CNVs. Our aim is to provide a brain map for the effect sizes of genes on cognitive ability when deleted or duplicated. First, we identified all CNVs larger than 50 kilobases in 265000 individuals with assessments of cognitive ability. Second, to assign genes to large scale brain networks, we partitioned coding genes into 180 sets based on whether they were highly expressed (data from Allen Human Brain Atlas) in one of the 180 cortical regions of the Glasser parcellation. Third, we fitted 180 linear models to compute the mean effect size on cognitive ability of genes highly expressed in each cortical region, for deletions and duplication separately. Finally we projected these effect sizes on the Glasser brain atlas and compared their cortical distribution with established cortical hierarchies. 155 and 138 out of 180 brain regions showed significant negative effects on cognitive ability with largest effect sizes observed in sensorimotor and association cortical regions, for deletions and duplications respectively. The cortical effect size maps for deletions and duplications were positively and negatively correlated to the gene-expression, anatomical, and functional cortical hierarchies. Stratifying genes using LOEUF and extensive sensitivity analyses did not change these results. Although all brain networks are involved, our results suggest that effect sizes of deletions and duplications have opposing distributions across the cortical gradient. For deletions, the largest effects on cognition are observed in genes highly expressed in the unimodal regions involved in action and perception. Duplications affect cognition most through integrative regions on the other end of the cortical gradient; key brain structures in human evolution that are involved in perceptually decoupled cognition and internal thought.

Molecular Effects of Genetic Variation Posters - Thursday

PB2640*. Massively parallel reporter assays with multi-layer annotations identified cell-type-specific functional variants and genes associated with melanoma

Authors:

E. Long¹, J. Yin², K. Funderburk³, M. Xu⁴, J. Feng¹, A. Kane¹, T. Zhang⁵, T. Myers⁶, A. Golden¹, R. Thakur⁷, H. KONG¹, L. Jessop⁸, E. Kim⁹, K. Jones¹, M. Machiela¹, K. Yu¹⁰, M. Iles¹¹, M. Landi⁵, M. Law¹², S. Chanock¹³, K. Brown¹, J. Choi¹; ¹Natl. Cancer Inst., Rockville, MD, ²NCI/NIH, Rockville, MD, ³Natl. Cancer Institute, Germantown, MD, ⁴Natl. Cancer Inst., Gaithersburg, MD, ⁵Natl. Cancer Inst., Bethesda, MD, ⁶NIH, Bethesda, MD, ⁷NCI-NIH, Bethesda, MD, ⁸NCI, Bethesda, MD, ⁹Yonsei Univ., Seoul, Korea, Republic of, ¹⁰NIH, Bethesda, MD, ¹¹Leeds Inst. of Data Analytics, Leeds, United Kingdom, ¹²QIMR Berghofer Med. Res. Inst., Brisbane, Australia, ¹³Natl. Cancer Inst, Rockville, MD

Abstract Body:

Melanoma is the deadliest form of skin cancer, with ~54 melanoma-associated loci identified by genome-wide association studies (GWAS). Most melanoma-associated loci contain multiple variants that are statistically indistinguishable due to linkage disequilibrium (LD); these are hypothesized to regulate gene expression in *cis* in a cell-type specific manner. Given these challenges, functional variants and their target genes for most melanoma-risk loci have not been characterized. Here, we performed massively parallel reporter assays (MPRA) of 1,992 variants selected from the most recent melanoma GWAS by LD ($R^2 > 0.8$) or GWAS statistics (log likelihood ratio $< 1:1000$) to test their potential as transcriptional enhancers in luciferase constructs via transfections into malignant melanoma ($n = 8$) and normal melanocyte cells ($n = 5$). Of 1,992 tested variants, we identified 285 from 42 loci (78% of the 54 loci) displaying significant allelic transcriptional activities in either cell type ($FDR < 1\%$). We further characterized MPRA-significant variants by motif prediction, epigenome annotation, and statistical/functional fine-mapping to create an integrative variant scoring system. This system efficiently prioritized a median of two plausible candidate variants per locus and nominated a single variant for 18 loci. Among them were well-characterized functional variants including two variants with the highest scores, which supported the validity of our system for variant prioritization. Overlaying the MPRA-significant variants with genome-wide significant expression or methylation quantitative trait loci (QTLs) from melanocytes ($n = 106$) or melanomas ($n = 444$) identified candidate susceptibility genes for 60% of them (172 of 285 variants). CRISPRi of top-scoring variants using the dCAS9-ZIM3 system in the melanoma cell line validated their effect on the eQTL target genes, *MAFF* (22q13.1) and *GPRC5A* (12p13.1). Finally, we identified 36 melanoma-specific and 45 melanocyte-specific MPRA-significant variants, a subset of which are linked to cell-type-specific target genes. Analyses of transcription factor availability in MPRA datasets and variant-transcription factor interaction in eQTL datasets highlighted the roles of transcription factors in cell-type-specific variant functionality. In conclusion, MPRA along with multi-layer annotations effectively prioritized plausible candidates from most melanoma GWAS loci and highlighted cellular contexts where the susceptibility variants and genes are functional; this approach is applicable to other complex diseases.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2641. Massively parallel saturation genome editing of an essential mitochondrial targeting sequence.

Authors:

M. Forrest¹, X. Jia², J. Kitzman³, A. Antonellis¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Univ Michigan, Ann Arbor, MI, ³Univ MICHIGAN, Ann Arbor, MI

Abstract Body:

Aminoacyl-tRNA synthetases (ARSs) are ubiquitously expressed, essential genes that conjugate amino acids to cognate tRNAs for protein synthesis in the cytoplasm and mitochondria. Pathogenic variants in all loci encoding mitochondrial ARSs cause diverse mitochondrial disease phenotypes. Homozygous or compound heterozygous loss-of-function missense variants in the gene encoding the bifunctional lysyl-tRNA synthetase (*KARS1*) cause recessive sensorineural hearing loss, vision loss, developmental delay, and mitochondrial dysfunction. The mitochondrial isoform of KARS1 (mtKARS1) is encoded by *KARS1* transcripts that include exon 2, which encodes 49 distinct N-terminal amino-acids to allow mitochondrial localization. To date, no disease variants have been reported in the mtKARS1-exclusive *KARS1* exon 2 sequence; further, the critical amino-acid residues required for KARS1 mitochondrial localization are incompletely understood.

To assess the functional consequences of synonymous, missense, and nonsense variants across *KARS1* exon 2, we employed saturation genome editing (SGE), which combines CRISPR-Cas9 and homology repair plasmid template libraries to derive pools of knock-in human cell lines bearing all possible single-residue variants across *KARS1* exon 2 and flanking proximal introns. Variant frequency at two subsequent passages versus frequency in the template library was quantitated by deep DNA sequencing of fixed-amplicon libraries, with variants that are depleted by cell death across passages indicating KARS1 loss-of-function. Preliminary results showed recovery of 3596/3816 (94%) variants at passage 1, indicating efficient library knock-in. Nonsense variants were reduced overall by passage 2; however, nonsense variants in several codons adjacent to the cytoplasmic KARS1 termination codon at the 5'-end of *KARS1* exon 2 showed relatively neutral effects. Further, reduced tolerance for amino-acid substitution was noted for codons proximal to a predicted amphipathic mitochondrial localization motif. Here, we will present all of our unpublished functional data on the KARS1 MTS. The results of this study will define the amino-acid residues required for KARS1 mitochondrial localization and will serve as a database to aid in functional assessment of *KARS1* variants identified in patient populations.

Molecular Effects of Genetic Variation Posters - Thursday

PB2642. Maternally Inherited novel *SUMO2* variant associated with a complex phenotype

Authors:

D. Otohinoi¹, C. Hicks², R. Zambrano^{3,4}; ¹Louisiana State Univ. Hlth.Sci. Ctr., New Orleans, LA, ²Louisiana State Univ. Hlth.Sci. Ctr., New Orleans, LA, ³Louisiana State Univ Hlth.Sci Ctr, New Orleans, LA, ⁴Dept. of Pediatrics, Children's Hosp., New Orleans, LA

Abstract Body:

Background: Post-translational modifications (PTMs) remains a key player in nearly all aspects of biological processes by regulating protein functions. They also play a significant role in embryo development by participating in DNA damage response, chromosome condensation, and cytoskeletal organization. Among various PTMs involved in embryogenesis, SUMOylation on lysine residue, has a crucial role in gene expression, DNA damage responses, chromosomal stability, and DNA integrity. Amongst the isoforms of small ubiquitin-like modifier (SUMO) involved in SUMOylation, SUMO2, has been reported to be the most predominant and important. Here, we report a novel *SUMO2* variant, associated with a complex phenotype in both proband and his mother.

Case report: Proband is a 9-year-old male of African American ancestry delivered at 33 weeks gestation via emergency cesarean due to fetal bradycardia. Pregnancy was complicated by spotting and abnormal ultrasounds (lower extremity abnormalities). The proband spent 5 weeks in the neonatal intensive care unit and multiple congenital abnormalities were observed including a right femoral distal bifurcation, an absent tibia, syndactyly of 2-3-4th toes and an absent 1st toe, left metatarsus adductus, patent ductus arteriosus with hypoplastic arch, and dysmorphic facial features. His mother presented with similar phenotype including: partial tibial agenesis, clubfoot, and toe syndactyly. Whole exome sequencing identified a maternally inherited heterozygous variant in exon 1 of *SUMO2* (NM_006937.4; P61956) c.2 T>A: p.Met1?. This variant is not reported in population databases nor literature. In silico analysis including protein-protein interaction and impact prediction following ACMG guidelines suggests this variant is disease causing, hence, we propose a pathogenic classification.

Conclusion: We report a novel and likely pathogenic variant in *SUMO2*. In animal studies knockdown of *SUMO2* altered the expression of *CDX2*, *OCT4*, and *NANOG*; important genes in embryonic development, and result in embryonic lethality. We propose pathogenic variants in *SUMO2* are responsible for the phenotype in this family. Our findings set the stage to characterize the phenotype associated with *SUMO2* start codon pathogenic variants, a clinically heterogeneous presentation that well illustrates the importance of disruption of SUMOylation.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2643. Mechanisms of adaptation in high-altitude pregnancy: association of genotype with oxygen delivery and placental metabolism

Authors:

K. O'Brien¹, W. Gu², J. A. Houck³, L. M. W. Holzner⁴, J. L. Armstrong⁴, A. P. Sowton⁴, P. M. Darwin⁴, L. Toledo-Jaldin⁵, L. G. Moore⁶, A. J. Murray⁴, T. S. Simonson⁷, C. G. Julian³; ¹Univ. of California San Diego, San Diego, CA, ²Univ. of California, San Diego, La Jolla, CA, ³Univ. of Colorado, Denver, CO, ⁴Univ. of Cambridge, Cambridge, United Kingdom, ⁵Hosp. Materno-Infantil, Bolivia, Bolivia, Plurinational State of, ⁶Univ. of Colorado, Denver, CA, ⁷UC San Diego, San Diego, CA

Abstract Body:

The hypobaric hypoxia of high altitude (HA) residence challenges tissue oxygen homeostasis and metabolism. HA exposure during pregnancy reduces birth weight and alters placental metabolism. HA ancestry protects against altitude-associated reductions in birth weight, indicating hypoxia tolerance. Such protection appears to be genetic in nature, yet the underlying mechanisms are unknown. In a cohort of 69 pregnant Andeans (18-45y) living in La Paz, Bolivia (3600 - 4100m) and delivering by unlabored Cesarean section, we performed selection-nominated association studies to explore relationships between putatively adaptive variation and phenotypes related to oxygen delivery in pregnancy or placental metabolism. Maternal genotyping was performed using the 1.8 million SNP Multiethnic Genotyping Array (Illumina). Placental mitochondrial function was assessed in cryopreserved villous biopsies using high-resolution respirometry (Oxygraph-2k, Oroboros). Maternal and umbilical venous plasma was obtained for the measurement of circulating protein levels by ELISA. Using within-population selection tests (iHS) to detect signatures of natural selection, putatively adaptive haplotypes (iHS>3) were identified; those overlapping with an *a priori* cellular hypoxic signaling and metabolism gene list were prioritized for association analysis. Linear regression modeling revealed associations between prioritized haplotypes and key outcome measures at an FDR corrected level of $p < 0.05$. Haplotypes within 200kb of *CPT2* (iHS 5.38) and both *POMC* and *DNMT3* (iHS 3.28) associated with lower maternal plasma erythropoietin ($p=0.02$ and $p=0.01$, respectively). A haplotype within 200kb of *TBX5* associated with lower protein levels of the angiogenic factor VEGF (iHS 3.65, $p=0.04$) in umbilical venous blood. Finally, a haplotype within *PTPRD* (iHS 3.31) associated with lower placental respiratory capacity ($p=0.002$). Our results reveal novel associations between putatively adaptive gene regions and phenotypes linked to oxygen carriage and delivery, as well as placental mitochondrial respiratory capacity. Together, these findings provide new insights into mechanism of high-altitude adaptation at different points of the oxygen cascade in the context of reproduction.

Molecular Effects of Genetic Variation Posters - Thursday

PB2644. Mendelian randomization reveals the causal interaction of *MICA* transcription levels with the risk of Graves' disease

Authors:

Y. Sutoh¹, S. Komaki¹, T. Yamaji², S. Nakano², R. Katagiri², N. Sawada², K. Ono¹, H. Ohmomo¹, T. Hachiya¹, Y. Otsuka-Yamasaki¹, A. Takashima¹, S. Umekage¹, M. Iwasaki², A. Shimizu¹; ¹Iwate Med. Univ., Yahaba, Japan, ²Natl. Cancer Ctr. Inst. for Cancer Control, Tokyo, Japan

Abstract Body:

Exposure to environmental stress factors activates the cellular stress response. Stressed cells express specific molecules, so-called “danger signals”, which activate the immune response, leading to the elimination of these stressed cells. Natural killer group 2 member D (NKG2D) ligands (NKG2DLs) are expressed on the surface of stressed cells as a danger signal. NKG2DL⁺ cells are eliminated by cytotoxic cells to maintain tissue homeostasis. Interestingly, NKG2DL expression is also detected in activated immune cells, such as the macrophages infiltrating atherosclerotic plaques. This expression is thought to be associated with immune regulation; however, there are limited studies on humans. Here, we provide evidence for a phenotype associated with the NKG2DL expression level using an epidemiological approach based on the genotype data coupled with self-reported medical histories of 44,739 Japanese residents collected in the Tohoku Medical Megabank (TMM) project. First, we conducted a phenome-wide association study (PheWAS) of the expression quantitative-trait loci (eQTL) of NKG2DL transcripts. Following Bonferroni correction, we identified 10 or 13 significant outcomes for 58 available phenotypes associated with *MICA* or *MICB*, predominant NKG2DLs in various human tissues; 8 were common in both genes, including chronic hepatitis B (CHB), Graves' disease (GD), and allergic conjunctivitis. Brain tumors and rheumatoid arthritis were associated with only *MICA* eQTLs. Lung cancer, hypertension, and cavity were associated with only *MICB* eQTLs. Next, we performed Mendelian randomization (MR) with NKG2DL expression as exposure and identified the causal effect of the *MICA* expression level on the risk of GD ($p = 4.1 \times 10^{-3}$), allergic conjunctivitis ($p = 1.8 \times 10^{-2}$), and CHB ($p = 4.6 \times 10^{-2}$) using a weighted median estimator. Similarly, we observed the causal effect of *MICB* expression on the risk of CHB ($p = 2.9 \times 10^{-2}$) and cavity ($p = 3.4 \times 10^{-2}$). No significant pleiotropy was found in these associations ($p > 0.05$). A negative causal effect on the risk of GD (odds ratio (OR) 0.983; [95% confidence interval (CI) 0.971-0.995]) per standard deviation increase in *MICA* expression was observed. Finally, we performed replication analysis with different datasets and several sensitivity analyses with different estimators, such as the MR-Egger and inverse-variance weighted (IVW) method, and confirmed the robustness of the conclusion. These results suggest that low *MICA* transcription levels in immune cells are associated with a higher risk of Graves' disease. Our strategy might be useful for investigating a gene function based on a human dataset.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2645. Metabolomics to clarify the mechanism of *ALDH2* polymorphism-induced alcoholic liver injury.

Authors:

S. Harada^{1,2}, R. Toki¹, A. Hirata¹, N. Miyagawa¹, M. Iida¹, A. Hirayama², M. Sugimoto², T. Soga², M. Tomita², T. Okamura¹, T. Takebayashi^{1,2}; ¹Keio Univ. Sch. of Med., Tokyo, Japan, ²Inst. for Advanced BioSci.s, Keio Univ., Tsuruoka, Japan

Abstract Body:

Background: Alcoholic liver injury is associated with *ALDH2* (rs671) polymorphisms, particularly in Asians, but the pathogenic mechanism remains unclear. This study investigated how *ALDH2* polymorphisms and drinking habits are associated with liver function decline via metabolomic changes in a general Japanese population.

Methods: In the Tsuruoka Metabolomics Cohort Study conducted in Japan, 10,993 Japanese were genotyped, profiled for metabolites, and assessed for health status comprehensively. Absolute quantification of 94 plasma metabolites has been performed using the CE-MS method. This study included 3,200 participants who were genotyped for *ALDH2* (rs671) using Fluidigm SNP Type™ Assays (192.24 Dynamic Array™ integrated fluidic circuits) to examine the mechanisms underlying the development of alcoholic liver injury using mediation analysis. First, mediation analysis was conducted using *ALDH2* (rs671) polymorphism as an exposure, daily ethanol intake as a mediator, plasma metabolite level as an outcome, and age and sex as confounders to examine metabolites strongly associated with *ALDH2* polymorphism and alcohol consumption. Then, we conducted a mediation analysis with alcohol intake as an exposure, metabolite level as a mediator, and liver injury (serum AST level) as an outcome to determine how strongly a metabolite explained the pathogenesis of alcoholic liver injury.

Results: *ALDH2* (rs671) alleles were AA in 3.5%, AG in 29.5%, and GG in 67.0%. All but one of the AA carriers with inactive *ALDH2* were non-drinkers; therefore, AA carriers were excluded from the analysis. The 3,079 participants (62.9 ± 7.4 years old, 1,375 men and 1,704 women) with a complete dataset were included in the analysis. 1520 (49.3%) were daily drinkers. *ALDH2* (rs671) polymorphism was associated with plasma Gln level (-0.10 SD (95% CI: -0.17, -0.02) change for GG compared to AG). Of these, 76.5% were mediated by alcohol intake. Then, we conducted a mediation analysis with alcohol intake as an exposure, Gln as a mediator, and AST as an outcome. We found that AST increased by 0.05 (95%CI: 0.03, 0.07) per 1 g increase in daily alcohol intake, and 15% of this increase were mediated by plasma Gln level.

Conclusions: In the general Japanese population, *ALDH2* (rs671) polymorphism was strongly associated with plasma Gln level, and most of the associations were explained by alcohol intake. Gln level explained 15% of AST elevation due to alcohol, suggesting that Gln metabolism plays an essential role in the mechanism of *ALDH2* polymorphism-induced alcoholic liver injury. A follow-up study is needed to clarify the causal relationship.

Molecular Effects of Genetic Variation Posters - Thursday

PB2646. Mimicking NDD phenotypes and functional characterization of *PPP1R9A*^{+/-} mutation using iPSCs- derived neurons and single-cell transcriptomics

Authors:

M. Uddin^{1,2}, B. Zehra¹, N. Mohamed¹, N. Nassir¹, A. Ahmed¹, R. Khalil³, F. Ahmad³, N. Jezawi⁴, B. Berdiev¹; ¹Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates, ²GenomeArc Inc., Toronto, ON, Canada, ³American Univ. of Sharjah, Sharjah, United Arab Emirates, ⁴Princess Margaret Res. Ctr., Toronto, ON, Canada

Abstract Body:

PPP1R9A is an imprinted gene that codes Neurabin I and is located on chromosome 7q12. Neurabin I is a protein phosphatase-1 regulatory subunit that primarily interacts with F-actin, protein phosphatase 1 (PP1), neurabin-2 and p70-S6K; proteins that are directly involved in neurite formation and establishing synaptic connections. Genetic mutations impacting *PPP1R9A* are classified as a variant of uncertain significance in several neurodevelopmental problems and are recently reported as a candidate gene underlying the pathophysiological features of autism spectrum disorder (ASD). However, the functional and molecular role of *PPP1R9A* in NDD is still inconclusive. Therefore, this study aimed to functionally characterize the CRISPR/Cas9 mediated heterozygous knockout (KO) model for the *PPP1R9A*^{+/-} gene in hiPSCs to identify its role in NDD. These KO hiPSCs were then differentiated into neurons to recapitulate the condition *in vitro* and explore morphogenesis and differentiation patterns in the brain through morphometric analysis of dendritic spines, immunofluorescence, and protein expression of biomarkers. Furthermore, using an overexpression vector, a wild-type copy of the *PPP1R9A* gene was introduced in the KO *PPP1R9A*^{+/-} cells to rescue the mutation effect. The whole transcriptomics analysis of control, KO *PPP1R9A*^{+/-}, and over expressed *PPP1R9A* in iPSCs-derived neurons revealed the set of differentially expressed genes that were significantly enriched for functions of neural development, synapse signaling, and morphogenesis. Also, the expression for upregulated genes in KO *PPP1R9A*^{+/-} iPSCs-derived neurons was reverted to stabilized expression in *PPP1R9A* over expression neurons. Currently, we are conducting a single cell (SC) RNASeq experiment for the high-resolution detection of cell type-specific phenotypes through gene dysregulation that resulted in ASD-related neurodevelopmental phenotype in *PPP1R9A*^{+/-} iPSCs-derived neurons. Since we observed a dominant negative effect in the *PPP1R9A* overexpressed neurons, therefore, we are currently developing a gene therapy approach using antisense oligonucleotides (ASO) technology to reverse the KO effect and restore phenotypes comparable to the wild type. In summary, by integrating CRISPR/Cas9 technology in iPSCs, single-cell RNA seq, and other cutting-edge techniques, this study established the role of Neurabin I mutation in the pathogenesis of NDD.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2647. Misexpression of inactive genes is associated with nearby rare structural variants.

Authors:

T. Vanderstichele¹, K. L. Burnham¹, N. de Klein¹, B. Howell¹, K. Walter¹, K. Kundu^{1,2}, A. Nath^{2,3}, E. Persyn², J. Marten², INTERVAL Study, D. Roberts⁴, E. Di Angelantonio², J. Danesh², A. Berton⁵, A. Platt⁵, A. Butterworth², N. Soranzo¹, L. Parts¹, M. Inouye^{2,3}, D. Paul^{2,6}, E. E. Davenport¹; ¹Wellcome Sanger Inst., Hinxton, United Kingdom, ²Univ. of Cambridge, Cambridge, United Kingdom, ³Baker Heart and Diabetes Inst., Melbourne, Australia, ⁴Univ. of Oxford, Oxford, United Kingdom, ⁵Translational Sci. and Experimental Med., Res. and Early Dev., Respiratory and Immunology, BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom, ⁶Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom

Abstract Body:

In human genetics, it remains challenging to identify rare genetic variants that influence human phenotypes and understand the underlying mechanism of their effect. In previous studies, gene expression outliers from population-scale RNA sequencing data have been used to identify functional rare variants with large effects. These analyses have focused on genes that are expressed across many individuals in the studied tissue. However, evidence from Mendelian diseases shows that genetic variants can cause misexpression of inactive genes in a specific tissue. It is currently unknown whether this phenomenon is widespread in human populations, by what mechanisms it occurs and the phenotypic impact it has. Using 4,542 whole blood RNA sequencing samples from a prospective cohort of apparently healthy European blood donors recruited to the INTERVAL study, we identified a set of genes that were not expressed in at least 95% of individuals, but were highly expressed in at least one other tissue according to the Genotype-Tissue Expression (GTEx) project (n = 6,401). We assessed whether rare variants in cis, identified for 2,804 samples with whole-genome sequencing data, were associated with misexpression of these normally inactive genes, and found that rare structural variants (MAF < 1%) were enriched within a +/- 200 kb window around the gene body. This enrichment increased with the magnitude of misexpression from a risk ratio of 1.61 for misexpression events greater than 1 TPM (p-value < 2.49x10⁻⁵) to 11.08 (p-value < 2.76x10⁻¹⁰) for events greater than 5 TPM. Common structural variants (MAF ≥ 1%) were not enriched among misexpression events. Rare duplications were most enriched among different structural variants types, while rare indels and single nucleotide variants were not enriched even within a +/- 10 kb window. Overall, we present an alternative approach for identifying rare genetic variants associated with disruption of normal gene expression in healthy human populations. Future analysis of the properties of the structural variants associated with misexpression will be useful for improving our understanding of the regulatory principles of gene expression and three-dimensional genome organisation.

Molecular Effects of Genetic Variation Posters - Thursday

PB2648. Mitochondrial Complex III (CIII) deficiency case study: Analysis of a likely pathogenic variant in *UQCRC2* gene as the etiology of intellectual disabilities and obesity.

Authors:

G. Arroyo Figueroa¹, A. Lebrón Ilarraza¹, G. Serrano Rodríguez², C. Velez Delgado³, F. Velez Bartolomei⁴, E. Albino⁵, S. Carlo⁶, A. Cornier⁷; ¹Univ. of Puerto Rico- Río Piedras Campus, San Juan, PR, ²Sch. of Med., Ponce Hlth.Sci. Univ., Ponce, PR, ³Genetic Diagnostic Group, San Juan, PR, ⁴Genetics Section San Jorge Children's & Women's Hosp., San Juan, PR, ⁵Sch. of Hlth.Professions, Med. Sci. Campus, Univ. of Puerto Rico, San Juan, PR, ⁶Ponce Hlth.Sci. Univ., Cabo Rojo, PR, ⁷San Jorge Children s Hosp./ Ponce Hlth.Sci. Univ., San Juan, PR

Abstract Body:

Understanding the genetic factors that lead to rare diseases phenotypes is of critical importance, more so in genetically underrepresented populations such as Puerto Ricans. We are reporting a case involving *UQCRC2* gene, which encodes for protein 2, one of the 11 structural subunits of the mitochondrial complex III (CIII). CIII deficiency is a rare genetic condition caused by mitochondrial dysfunction. This multisystem disorder's phenotype includes; hypotonia, failure to thrive, encephalopathy, and delayed psychomotor development. Abnormalities of some nuclear genes that encode mitochondrial assembly factors are well established, but many CIII deficiency disorders remain ambiguous. In this case, we are presenting an exome analysis with a new heterozygous frameshift mutation, c.753delC(p.Asn252Thrfs*21), in *UQCRC2*. This deletion of a base pair results in reading frameshift at position 252 and the introduction of a premature termination codon 20 residues downstream. This is a 14-year-old female who's previous genetic evaluations at other mainland Children's Hospital suggested a possible clinical diagnosis of Prader-Willi syndrome. Genetic testing demonstrated a normal methylation 15q11 through 13 as well as no cytogenetic evidence of a microdeletion, rendering the diagnosis no longer valid. A whole exome sequencing documented the *UQCRC2* gene likely pathogenic variant. Phenotypical findings in this patient includes development delay, obesity, autism spectrum disorder (ASD), speech delay, behavioral problems, hypotonia, incontinence issues and fine motor deficits. Obesity has not been reported associated to mutations in *UQCRC2* gene. This patient does not have family history of obesity or other epigenetic factors that may be associated to obesity in her or her family. This case provides clinical evidence of an expansion of the phenotype associate to *UQCRC2* and/or Complex III phenotypes and should be consider a molecular etiology associate to obesity in the presence of developmental delay and ASD.

Molecular Effects of Genetic Variation Posters - Thursday

PB2649*. Mitotic recombination is a common mechanism of cellular mosaicism in Fanconi anemia.

Authors:

F. X. Donovan¹, R. Ramanagoudr-Bhojappa¹, F. P. Lach², D. C. Kimble¹, S. Soma¹, A-M. Gonzalez², R. O. Rosti², A. D. Auerbach³, A. Smogorzewska², S. C. Chandrasekharappa¹; ¹Cancer Genetics and Comparative Genomics Branch, NHGRI, NIH, Bethesda, MD, ²Lab. of Genome Maintenance, The Rockefeller Univ., New York, NY, ³Human Genetics and Hematology Program, The Rockefeller Univ., New York, NY

Abstract Body:

Objective: The aim of this study is to understand the mechanism of somatic mosaicism and its influence on patient phenotype among individuals with Fanconi anemia (FA).

Methods: We evaluated DNA from 742 individuals enrolled in the International Fanconi Anemia Registry (IFAR) by targeted sequencing with a custom capture kit designed for the entire length of a panel of 180 genes, including all known FA genes, and for 356 individuals we employed genome-wide SNP array (~1M SNPs). For variants that exhibited skewed allele frequency we performed deep sequencing using MiSeq. To confirm pathogenicity, we implemented targeted RNAseq or RT-PCR for splice variants, and cell cycle arrest, cell survival, and FANCD2 ubiquitination assays for missense variants.

Results: We identified evidence of somatic mosaicism in 32 individuals from 30 families. The advent of a *de novo* compensatory variant or the reversal of a pathogenic variant was observed in five and 13 individuals, respectively. Isodisomy of an FA gene resulting from mitotic recombination (MR) was exhibited by 14 individuals. In two families there were siblings that each exhibited mosaicism via different mechanisms. The isodisomy extended to the chromosome terminus in each case, influencing the allelic expression in two different ways: six cases exhibited the breakpoint junction within the FA gene, between two biallelic pathogenic variants, resulting in the replacement of the pathogenic variant by a wildtype sequence; eight cases exhibited isodisomy of the entire FA gene resulting in a loss of one variant and duplication of the second. Of the eight cases with isodisomy of the entire gene we performed functional analysis on 5/7 variants, as one variant occurred in two cases, and in each case the duplicated variant exhibited hypomorphic function. Neither isodisomy event appear to influence the age at onset of hematologic disease, but the cases with an intragenic isodisomy junction showed longer survival, indicating that the loss of variant and subsequent expansion of modified cells promoted a positive effect on the hematologic disease course.

Conclusion: We identified 30 families exhibiting somatic mosaicism via three distinct mechanisms: compensatory *de novo* variant, variant reversion, and isodisomy via MR. The MR group is further classified by whether the isodisomy includes the entire gene or its junction occurs within a gene. These natural corrections improved hematologic outcomes underscoring the importance of knowing how cells naturally revert pathogenic variants and how these reversions affect patient outcomes, which can influence the design of clinical interventions including gene therapy trials.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2650. Modeling single-cell activation states enhances power to identify ex vivo stimulation response eQTLs.

Authors:

C. Valencia^{1,2,3}, A. Nathan^{1,3,2}, J. Kang^{2,1,3}, I. Rumker^{1,2,3}, S. Raychaudhuri^{1,2,3}; ¹Brigham & Women's Hosp., Boston, MA, ²Harvard Med. Sch., Boston, MA, ³Broad Inst. of MIT and Harvard, Boston, MA

Abstract Body:

Genome-wide association studies (GWAS) of complex human traits such as susceptibility to infection implicate noncoding variation, where associations may be mediated through gene expression regulation. Studying expression quantitative trait loci (eQTL) aims to improve the interpretation of GWAS results by linking genetic variants to expression levels of specific genes. However, in most cases this aim remains elusive, in part because eQTLs are context-sensitive, and a complete characterization of eQTLs in the context of environmental stimuli and across cell types remains incomplete. Cell perturbations such as viral infection activate gene programs like interferon response. Previous studies used ex-vivo stimulation to identify “response” eQTLs, i.e. eQTLs that are specific to an infected cell state. However, these studies did not consider variation in cell response to the stimulus. Now single cell data reveal cell-cell heterogeneity in the stimulus response of individual cells. Effective modeling of per-cell activation state (and other cell-level variables) may enhance power to detect response eQTLs. Here, we used published data (Randolph, et al. 2022 Science) to study the effect of 12 hours of ex-vivo stimulation with either influenza virus (Flu) or mock (Control) in 208,223 peripheral blood mononuclear cells (PBMCs) of 45 African-American and 45 European men using single cell RNA-seq. First, we prioritized 1,955 eQTLs with robust effect across all stimuli and cell types. Then, we modeled the expression of these genes in single cells as a function of genotype and an interaction with cell state accounting for confounders and batch using the Poisson Mixed effects Model (PMM) described by Nathan et al 2022, Nature. Because the second expression principal component (PC-2) positioned cells along a stimulus gradient, and was enriched for interferon genes, PC-2 was used as a continuous activation state of stimulus response per cell. We found that around 10% of the 1955 eQTLs were heterogeneous along the continuous activation state based on the significance of the interaction term (response eQTL). In contrast, modeling activation state as binary in single cells or pseudobulk expression profiles found only 6.5% or 4.29% of the eQTLs to be response eQTLs, respectively. In addition, using the PMM with continuous single-cell activation state, we found that almost 66% of the response eQTLs showed heterogeneous effects across the 5 major cell types (Monocytes, NK, CD4-T, CD8-T, and B). In summary, our results demonstrated the benefit of studying response eQTLs at single cell resolution in heterogeneous samples.

Molecular Effects of Genetic Variation Posters - Thursday

PB2651. Modeling the effects of *KCNQ1* SNPs on type 2 diabetes risk in iPSC-derived pancreatic islets identifies an early role of methylation that alters islet composition and beta-cell mass.

Authors:

A. Nair, E. Grellinger, Y. Muller, M. Traurig, R. Nelson, C. Bogardus, L. Baier; NIDDK/NIH, Phoenix, AZ

Abstract Body:

The strongest association for type 2 diabetes (T2D) in Southwestern American Indians maps to intron 15 of *KCNQ1* where the lead SNP (rs2299620) is in a region of strong linkage disequilibrium (LD). We previously mapped a functional region in this intron to an ~400 bp fragment containing rs2299620 and 3 additional SNPs and demonstrated functionality of this fragment using American Indian iPSC-derived pancreatic islets. In the current study, we used CRISPR-CAS9 to generate iPSCs with targeted edits at these 4 SNPs, and 3 additional SNPs in strong LD, and used these isogenic iPSC-derived pancreatic islets as a model system to study the effect of the variants. Flow cytometry of these islets identified a lower percentage of beta-like cells (cells co-expressing *INS* and *NKX6-1*) in islets generated using iPSCs with the risk haplotype (35.6% vs 47.4%, n=3, P=0.03), suggesting an effect of these T2D associated SNPs on beta cell mass. Stage specific gene expression analyses identified a significant difference in the fold increase in *INS* gene expression during the endocrine progenitor (EPs) stage (755.5-fold increase in expression in risk vs 4403-fold increase in expression in non-risk, relative to pancreatic progenitors, n=3, P=0.0002). As endocrine cell type commitment occurs during EP stage and increase in *INS* expression correlates with beta cell commitment, the lower increase in insulin expression during this stage with the risk haplotype could in part explain the effect of the variants on mature islet composition. In EPs with the risk haplotype, we also identified increased expression (fold difference=1.7, n=3, P=0.06) of *H19* which maps upstream of the SNPs and is known to affect beta-cell mass.

Methylation at enhancers regulates gene expression. Therefore, we assessed stage specific methylation at all 13 CpG sites (*CpG-func*) that encompass the functional region (between rs2299620 and rs74046911) and 3 CpG sites (*CpG-ctrl*) immediately upstream of rs2299620. We identified dynamic methylation changes at *CpG-func* whereas no changes at *CpG-ctrl* (remained hypermethylated, average methylation >90%) during different stages of pancreatic islet differentiation. *CpG-func* were hypermethylated (86.71%) during the iPSC stage but became hypomethylated (46.2%) during the EP stage. Interestingly, a strong difference in methylation was seen in *CpG-func* only during the EP stage when compared between the risk (30.6%) and non-risk haplotypes (61.8%). In summary, the T2D SNPs at *KCNQ1* affect islet composition (beta-cell mass) and this effect is manifested early during development, likely via an effect on CpG methylation and gene regulation.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2652. Mosaicism of a *de novo* *TP53* mutation in a patient with breast cancer: value of two germ layers study

Authors:

D. Aguilar¹, A. Aranda-Gutierrez¹, L. Flores-Lagunes², C. Alaez-Verson²; ¹Hosp. Zambrano Hellion Tec Salud, San Pedro Garza García, Mexico, ²INMEGEN, CdMx, Mexico

Abstract Body:

Introduction: Li-Fraumeni syndrome (LFS) is associated with a high lifetime risk of developing breast cancer (BC). A subset of patients can present in the form of mosaicism with uncertain clinical implications. Whether hereditary or *de novo*, the presence of LFS has clinical implications. Our objective is to report a case of BC diagnosed in a patient with mosaicism of a *de novo* *TP53* likely pathogenic variant (LPV) and discuss the literature regarding this entity. **Case description:** 41-year-old woman with stage IA hormone-receptor negative and HER2-positive BC. She was treated with a simple mastectomy and adjuvant therapy. She did not undergo radiotherapy. Germline genetic testing identified a *TP53* LPV (NM_000546:c.523C>T (p.Arg175Cys)) in a saliva sample. Peripheral blood and formaldehyde fixed-paraffin embedded tumor tissue from her pre-adjuvant mastectomy were further studied at Genomic Diagnostic Laboratory of the National Institute of Genomic Medicine (INMEGEN), confirming *TP53* mosaicism. Biological relationship between the patient and her parents was confirmed, with neither being carriers of the *TP53* LPV. Therefore, the patient's LPV was confirmed to be *de novo*. The patient has remained asymptomatic and without evidence of BC recurrence or any other malignancy. **Discussion:** Mutations occurring at post-zygotic stage are responsible for mosaic *TP53*, which accounts for 5% of LFS phenotypes. Mosaicism should be suspected in any patient with allele frequencies that differ significantly from 50% in NGS-based genetic testing. In the present case, suspicions of mosaicism were raised from the identification of a LPV in *TP53* (NM_000546:c.523C>T (p.Arg175Cys)) at a variant fraction of 11.9% in an ectoderm-derived sample (saliva sample), at a variant fraction of 29% in blood sample and 40% in the tumor DNA. Of note, this was documented before the patient was started on adjuvant chemotherapy. In addition, the presence of the mosaic *TP53*, a second somatic *TP53* mutation were also demonstrated in the tumor, supporting the involvement of both variants in the development of the patient's BC. Currently, information concerning the optimal treatment and survival outcomes of BC patients with *TP53* mosaicism is lacking. Notwithstanding, recommendations do exist in literature for the management of BC in patients with LFS. **Conclusion:** This case highlights the importance of distinguishing between germline and somatic mosaic *TP53* mutations, as this knowledge can guide appropriate cancer surveillance protocols and genetic counseling. Future studies aimed at identifying the optimal cancer treatment and follow-up schemes in patients with *TP53* mosaicism are warranted.

Molecular Effects of Genetic Variation Posters - Thursday

PB2653. Multiomics identifies dysregulation of Wnt-signaling through epigenetic gene regulation across tissues in Bohring-Opitz Syndrome

Authors:

I. Lin; UCLA, Los Angeles, CA

Abstract Body:

ASXL1 (*Additional sex combs-like 1*) plays a key role in epigenetic regulation of early developmental processes and is ubiquitously expressed across tissues. *De novo* autosomal truncating mutations of *ASXL1* cause Bohring-Opitz Syndrome (BOS), a rare genetic disorder characterized by profound intellectual disability, developmental delay, seizures, heart defects, and 'BOS posture'. **We hypothesize that highly penetrant and high-effect mutations in *ASXL1* dysregulate common gene regulatory mechanisms across cell types.** To examine this, we collected blood and skin punch biopsies from 17 individuals with BOS and 27 controls. We generated RNAseq data and found that differentially expressed genes (DEGs) were more upregulated in both BOS fibroblasts (129/177 DEGs, 72.9%) and blood (590/1097 DEGs, 53.8%) compared to controls. DEGs were dysregulated in the same direction across tissue types and top DEGs included *GREM2*, *GRIK5*, and *VANGL2*. In addition, gene ontology enrichments were identified in Wnt signaling in blood, and neuron projection development in both tissue types, driven by key early morphogenesis genes. Since *ASXL1* mutations alter histone modifications, we performed ATACseq to examine whether transcriptional dysregulation acts through chromatin accessibility. ATACseq of fibroblasts with truncating *ASXL1* variants identified, globally, more open chromatin regions (3036/4336 differential peaks $p_{adj} < 0.05$, 70.0%) compared to controls. Intersection of ATAC- and RNAseq DEGs (37/177, 31.6%) revealed a strong positive correlation ($r(69) = 0.64$, $p = 2.39 \times 10^{-9}$) between increased chromatin accessibility and increased gene expression, suggesting that *ASXL1* mutations directly influence chromatin accessibility resulting in transcriptional dysregulation. To examine epigenomic changes that may drive this dysregulation, we performed CUT&RUN against H3K4me3 and H3K27me3 marks, known to be regulated by *ASXL1*, and genome-wide DNA methylation profiling. CUT&RUN showed differential histone methylation at transcriptional start sites of key genes of interest, such as *VANGL2*. Genome-wide DNA methylation data identified 5121 CpG sites (FDR < 0.05) associated with an ENSEMBL gene, with 609 sites mapping to blood RNAseq DEGs, including *GREM2*, *GRIK5*, and *VANGL2*. Across our datasets, genes such as *VANGL2*, a planar cell polarity pathway protein that acts through Wnt signaling to direct embryonic nervous system development, are consistently dysregulated. Collectively, our data further elucidates the impact of truncating *ASXL1* mutations on gene regulatory mechanisms and pathogenesis in Bohring-Opitz Syndrome using a multiomics approach across tissues.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2654. Mutant *Wac* mice exhibit phenotypes relevant to DeSanto-Shinawi Syndrome and provide initial insights into molecular and developmental mechanisms.

Authors:

N. Seban¹, A. M. Stafford², C. Canales¹, S. Lozano¹, J. Zhu¹, R. Ortiz¹, M. Corea¹, D. Rahbarian¹, A. Nord¹, D. Vogt²; ¹Univ. of California Davis, Davis, CA, ²Dept. of Pediatrics and Human Dev., Michigan State Univ., East Lansing, MI

Abstract Body:

Rare monogenic syndromes are emerging as a collectively common cause of neurodevelopmental disorders associated with cognitive, social, behavioral, and neurological impacts, including autism, attention deficit hyperactivity disorder (ADHD) and seizures. One such example, DeSanto-Shinawi Syndrome (DESSH), is caused by mutations in the *WAC* gene. Individuals diagnosed with DESSH present craniofacial alterations as well as cognitive symptoms that include autism, ADHD and seizures. *WAC* encodes a protein that contains WW and coiled coil domains, motifs important for protein/protein interactions. Previous reports have demonstrated a multitude of cellular roles for the *WAC* protein, including positive regulation of mammalian target of rapamycin (MTOR), mitosis, transcriptional initiation and autophagy. Comprehensive studies from a mammalian model to understand how these changes occur and whether the same interactions or unique events occur in the mammalian brain are needed. We generated a constitutive murine *Wac* mutant mice and assessed behavioral, anatomical, and cellular phenotypes that are relevant to humans diagnosed with DESSH, then evaluated transcriptional profiling at single cell level on postnatal day (PND) 2. *Wac* heterozygous (Het) mice exhibited dysmorphic craniofacial features, decreased expression of GABAergic interneuron proteins, susceptibility to seizures, and behaviors implicating decreased learning and memory. The phenotypes seen in the murine model indicate some relevance to the underlying biology that causes symptoms of DESSH. Single nuclei transcriptomic profiling of mutant and wild-type cerebral cortex at postnatal day (PND) 2 identified differential regulation across numerous cell populations, including evidence for perturbed GABAergic signaling in line with interneuron histological changes. Validation of these findings is ongoing. Overall, our data support that this new mouse model recapitulates core symptoms of DESSH, and our results provide novel insights regarding molecular and cellular pathologies and neuroanatomical impacts associated with *Wac* haploinsufficiency. Future work is needed to understand how cell-type specific perturbations might drive pathology and to understand the molecular mechanisms of *WAC* protein that are required in the brain. Our results suggest initial therapeutics for DESSH and will spark further exploration of the biological mechanisms underlying this rare syndrome.

Molecular Effects of Genetic Variation Posters - Thursday

PB2655*. Mutational scanning to determine pathogenicity of Variants of Uncertain Significance in genes in the Sonic Hedgehog Pathway

Authors:

D. Baldrige, J. Shepherdson, B. Cohen; Washington Univ. in St. Louis, St. Louis, MO

Abstract Body:

Sequencing of patients with suspected monogenic disorders is a powerful approach for identifying molecular diagnoses, although the expected diagnostic yield continues to be around 30-40%. One major limitation is the abundance of Variants of Uncertain Significance (VUS) due to insufficient evidence for determining if variants are pathogenic or benign. High-throughput functional assessment of variant effects offers a scalable solution to this critical problem in genomic medicine via mutational scanning. This approach involves the simultaneous assessment of thousands of variants in a protein of interest, using a reliable, carefully calibrated, cell-based assay. To demonstrate the feasibility of establishing mutational scans for transcription factors, we have established an assay for the gene, *GLI2*, part of the sonic hedgehog signaling pathway. Pathogenic variants in *GLI2* cause Culler-Jones syndrome, involving endocrine and skeletal patterning abnormalities. Using cells engineered with a GFP reporter and transduced with lentivirus expressing *GLI2* variants, we conducted a SortSeq experiment, coupling fluorescence-activated cell sorting (FACS) to sequencing, to demonstrate the functional effects of these variants. Our results show perfect discrimination of known pathogenic and benign variants in *GLI2*. We use this system to assay all 170 missense variants in *GLI2* that have been submitted to ClinVar, including about 100 VUS. The end result is a “look-up table” that gives clinicians and diagnostic laboratories high confidence functional evidence for reclassification of coding variants in *GLI2*. We also demonstrate that this same cell-based assay can be used to test variants in additional genes in the sonic hedgehog pathway, including *SMO*, suggesting a pathway-based approach for scaling mutational scanning to many genes in the genome.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2656. NASP contributes to autism via epigenetic dysregulation of neural and immune signaling pathways

Authors:

J. Li¹, S. Zhang², J. Yang², J. Glessner¹, H-Q. Qu¹, H. Hakonarson³; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Tianjin Med. Univ., Tianjin, China, ³Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Epigenetics make substantial contribution to the etiology of autism spectrum disorder (ASD) and epigenetic regulators have been found to harbor causal mutations for ASD, but few studies have focused on the contribution of epigenetic regulators to the etiology of ASD and its immune comorbidity. Trio-based whole exome sequencing (WES) was conducted on a simplex family including one proband with autism and comorbidity of atopic disease. Genome editing technique was used to specifically knockout (KO) the tNASP (Nuclear Autoantigenic Sperm Protein) gene to understand the function of the potential causal gene tNASP. ATAC-seq, ChIP-seq and RNA-seq were performed to investigate impact of tNASP KO on chromatin accessibility, active chromatin state and gene expression profiling. Stable cell lines overexpressing wild type tNASP or the mutant tNASP at the tNASP KO background were constructed. ATAC-seq and RNA-seq were similarly conducted to examine the effect of tNASP mutation identified in the family. Screening in additional ASD subjects were carried out. Genetic analysis identified tNASP(Q289X) mutation as the potential causal variant for this family, which is a de novo likely gene disruptive (LGD) variant. Another ASD patient carrying copy number variant covering gene NASP was also identified. ATAC-seq and ChIP-seq showed that tNASP KO increases chromatin accessibility, promotes the active promoter state of genes enriched in synaptic signaling. RNA-seq further showed that tNASP KO leads to upregulated expression of genes functioning in the neural signaling and immune signaling pathways. The expression of tNASP(Q289X) is subjected to nonsense mediated decay. Comparing to WT tNASP, tNASP(Q289X) enhances chromatin accessibility of genes having enriched expression pattern in brain. RNA-seq revealed that genes involved in the neural signaling and the immune signaling are affected by the tNASP mutation. In summary, our study identifies a novel ASD gene tNASP, a histone binding protein with high expression in embryonic cells. Our data suggest that the de novo LGD variants of tNASP contribute to the pathogenesis of ASD and its immune comorbidity by regulating the chromatin accessibility, active promoter state of neural genes and further leading to the expression level change of immune genes.

Molecular Effects of Genetic Variation Posters - Thursday

PB2657*. Natural variants in *SEL1L* modify lethality, ERAD, and proteasome function in a model of NGLY1 deficiency.

Authors:

T. Tu'ifua, C. Y. Chow; Univ. of Utah, Salt Lake City, UT

Abstract Body:

N-glycanase 1 (NGLY1) deficiency is a rare disease caused by autosomal recessive loss of function mutations in the *NGLY1* gene. Patients suffer from movement disorder, developmental delay, liver dysfunction, and alacrima. NGLY1 removes N-linked glycans from glycoproteins in the cytoplasm and is thought to help clear misfolded glycoproteins from the endoplasmic reticulum (ER) through the ER associated degradation (ERAD) pathway. Despite this, the physiological significance of NGLY1 in ERAD is not understood. The best characterized substrate of NGLY1 is NRF1, a transcription factor that upregulates proteasome expression and the proteasome bounce back response. Our lab created a *Drosophila* model of NGLY1 deficiency that faithfully recapitulates several disease phenotypes observed in patients, including movement disorder, seizures, and lethality. We performed a *Drosophila* natural variation modifier screen using this model of NGLY1 deficiency and identified a number of modifiers that reduce the lethality of the model. In particular, we identified two putative loss of function variants in *SEL1L*: S780P and Δ 806-809. Both variants are localized in the SEL1L cytoplasmic tail, an uncharacterized domain of the protein. SEL1L is a known component of the ERAD complex that retrotranslocates misfolded proteins from the ER to the cytoplasm for degradation. To test the interaction between these *SEL1L* variants and *NGLY1*, we created CRISPR mutant fly lines that carry these *SEL1L* variants in a common genetic background and tested them with our model of NGLY1 deficiency. Validating our screen, the *SEL1L*^{P780} variant decreases the lethality observed in the NGLY1 deficiency model, as compared to the *SEL1L*^{S780} variant. Further, we found that, as compared to *SEL1L*^{S780}, *SEL1L*^{P780} reduces the NGLY1 model sensitivity to proteasome inhibition, a known defect in NGLY1 deficiency due to the misregulation of NRF1. We also find that these variants modify general ERAD function. We hypothesize that these *SEL1L* variants modify NGLY1 deficiency through NRF1 signaling and ERAD. These results will provide new insights into the role of NGLY1 in ERAD and the etiology of NGLY1 deficiency. *SEL1L* is a strong candidate modifier gene in patients, where variability in presentation is common.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2658. Novel actionable targets discovered by proteomic genome-wide association of 1715 unstudied proteins

Authors:

S. Kuliesius, L. Klaric, P. R. Timmers, P. Navarro, J. F. Wilson; Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract Body:

Genome-wide association study (GWAS) power lies in testing hundreds of thousands of single nucleotide polymorphisms (SNPs) across many genomes for associations with a trait. Recent technological developments in high-throughput proteomic assays allow simultaneous quantitative measurement of thousands of proteins in a single sample. Blood plasma proteins effectively provide a glimpse into multiple biological systems. Combining broad-capture proteomics with genomic variation across a population is a valuable resource with implications in elucidating complex traits and disease, drug development or repurposing, and precision medicine. Here, we investigated 6434 plasma proteins using the Somalogic aptamer-based technology in samples from the Viking Health Study - Shetland. 456 significant protein quantitative trait loci (pQTL) were found for 363 proteins in plasma (118 cis ($P < 5e-8$), 338 trans ($P < 6.6e-12$)). Of these, 73 pQTL were for previously unstudied proteins. We leveraged this new resource to perform causal inference using Mendelian randomization against complex traits of biomedical importance.

Molecular Effects of Genetic Variation Posters - Thursday

PB2660. Novel non-coding variant in *NMNAT1* reduces the transcription of *NMNAT1*, which leads to Leber Congenital amaurosis

Authors:

J. Han¹, J. Lee¹, S-T. Lee¹, J. Choi²; ¹Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of, ²YONSEI Univ. Coll. OF MEDICINE, Seoul, Korea, Republic of

Abstract Body:

NMNAT1 defects lead to early-onset retinal degeneration characterized by macular coloboma-like degeneration in early infancy. This gene encodes nicotinamide mononucleotide adenylyl-transferase 1 that synthesize from NAD precursor to NAD⁺. Interestingly, the dysfunction in this ubiquitous enzyme usually affect eye without syndromic features. In our cohort with inherited retinal diseases, two unrelated patients with Leber congenital amaurosis had only one pathogenic c.709C>T:p.(Arg237Cys) variant in *NMNAT1*. These 2 patients had infantile onset nystagmus, macular coloboma like degeneration or diffuse ellipsoid zone degeneration since early childhood, which was compatible with *NMNAT1* Leber congenital amaurosis. Therefore, we conducted genome sequencing in 2 unrelated patients, and identified novel non-coding c.-57+21C>T variant in *NMNAT1* in 3 affected patients from 2 unrelated family. This non-coding variant was existed in trans with known pathogenic c.709C>T:p.(Arg237Cys) missense variant. No copy number variation, structural variant, or mobile element insertion was detected by cn.Mops, Manta, and MELT algorithms. The minor allele frequency of c.-57+21C>T variant (hg38: chr1:g.9943536C>T) was 1/152182 and 1/263690 in gnomAD and BRAVO, respectively. Although CADD score was considered as benign (Phred: 4.389), regulatory variant specific in silico predictions showed high impact in DeepSEA (2.26), Eigen-PC (1.880), ncBoost (0.990), ncER (94.579) and ReMM (0.794). Using NCBI open reading frame (ORF) finder, this non-coding variant does not create new uORF. In quadruplex forming G-Rich Sequences (QGRS) mapper, this variant did not alter the G-quadruplexes scores. Because the relevant retinal tissues could not be biopsied, blood RNA-sequencing was done. RNA-sequencing in affected families confirmed decreased expression of *NMNAT1* level compared to controls (DESeq2, 146.4±24.0 vs 253.4±18.0, *P* =0.05). Although its precise mechanism should be elucidated, c.-57+21C>T variant located in the 5' untranslated region may affect the expression of *NMNAT1* in the transcription level. Thus, we suggest that c.-57+21C>T regulatory variant may have potential role in *NMNAT1* expression.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2661. Novel plasma and brain proteins that are implicated in multiple sclerosis risk

Authors:

X. Lin; Menzies Inst. for Med. Res., Hobart, Australia

Abstract Body:

Understanding how variations in the plasma and brain proteome contribute to multiple sclerosis susceptibility can provide important insights to guide drug repurposing and therapeutic development for multiple sclerosis. However, the role of genetically predicted protein abundance in multiple sclerosis remains largely unknown. Integrating plasma proteomics (n=3,301) and brain proteomics (n=376 discovery; n=152 replication) into multiple sclerosis genome-wide association studies (n=14,802 cases and 26,703 controls), we employed summary-based methods to identify candidate proteins involved in multiple sclerosis susceptibility. Next, we evaluated associations of the corresponding genes with multiple sclerosis at tissue-level using large gene expression quantitative trait data from whole-blood (n=31,684) and brain (n=1,194) tissue. Further, to assess transcriptional profiles for candidate proteins at cell-level, we examined gene expression patterns in immune cell types (dataset 1: n=73 cases and 97 controls; dataset 2: n=31 cases and 31 controls) for identified plasma proteins, and in brain cell types (dataset 1: n=4 cases and 5 controls; dataset 2: n=5 cases and 3 controls) for identified brain proteins. We identified 39 novel proteins associated with multiple sclerosis risk. Based on five identified plasma proteins, four available corresponding gene candidates showed consistent associations with multiple sclerosis risk in whole-blood, and we found *TAPBPL* up-regulation in multiple sclerosis B cells, CD8+ T cells and natural killer cells compared to controls. Among the 34 candidate brain proteins, 18 were replicated in a smaller cohort and 14 of 21 available corresponding gene candidates showed consistent associations with multiple sclerosis risk in brain tissue. In cell-specific analysis, six identified brain candidates showed consistent differential gene expression in neuron and oligodendrocyte cell clusters. The associated proteins present a robust set of candidate biomarkers and potential therapeutic targets for multiple sclerosis, reinforced by high concordance in downstream transcriptomics findings at tissue-level. This study also highlighted the heterogeneity of cell-specific transcriptional profiles for the identified proteins. Together, these findings can serve as an important anchor for future studies of disease mechanisms and therapeutic development.

Molecular Effects of Genetic Variation Posters - Thursday

PB2662. Novel splice variant in *PRUNE1* leads to aberrant neurodevelopment and neurodegeneration.

Authors:

A. Srivastava¹, A. Katiyar¹, N. Ghali², A. Moccia³, M. Wheeler⁴, D. Nickerson⁵, M. Bamshad⁴, S. Phadke⁶, K. Srivastava⁷, S. Bielas³; ¹Sanjay Gandhi PostGraduate Inst. of Med. Sci., Lucknow, India, ²Univ. of Michigan Med. Sch., Ypsilanti, MI, ³Univ MICHIGAN, Ann Arbor, MI, ⁴Univ. of Washington, Seattle, WA, ⁵Univ Washington Sch Med, Seattle, WA, ⁶Sanjay Gandhi PGIMS, Lucknow, India, ⁷CSIR-Central Drug Res. Inst., Lucknow, India

Abstract Body:

Recessively inherited variants in *PRUNE1* are the genetic basis of early onset neurodevelopmental and neurodegenerative disease with a range of phenotypes termed NMIHBA (neurodevelopmental disorder with microcephaly, hypotonia and variable brain anomalies). *PRUNE1* is a member of the DHH phosphoesterase family that has exopolyphosphatase and phosphodiesterase activity predicted to alter cellular motility through interaction with GSK-3 β and Wnt/ β -catenin signaling pathway. This study presents two similarly affected siblings from a non-consanguineous Indian family as an additional case of NMIHBA. Both individuals present with severe developmental delay, primary microcephaly, and spastic quadriparesis. Quad - exome and Sanger sequencing identified a novel homozygous *PRUNE1* splice acceptor c.119-2A>G variant shared by both siblings. Exon 2 skipping and stable expression of mutant protein was confirmed by minigene reporter assay and Western blot. Consistent with this observation, *In silico* protein modelling reveals a mechanism for loss of exopolyphosphatase activity in the case, further confirming a critical role in NMIHBA. We present a possible mechanism for *PRUNE* neuropathology, acting on microtubule dynamics through its downstream effects on Wnt/ β -catenin signaling. This novel splice variant exhibits features of a founder variant within the Indian population, proposing future implications to the field of population genetics and highlighting the importance of functional studies.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2663. Nuclear abnormalities in novel LMNA variant segregating with LMNA-associated cardiocutaneous progeria syndrome

Authors:

M. Vernet Machado Bressan Wilke¹, T. Schwab¹, M. Wick^{2,3}, E. Klee^{1,3,4}; ¹Ctr. for Individualized Med., Mayo Clinic, Rochester, MN, ²Dept. of Obstetrics & Gynecology Mayo Clinic, Rochester, MN, ³Dept. of Clinical Genomics, Mayo Clinic, Rochester, MN, ⁴Dept. of Quantitative Hlth.Sci., Mayo Clinic, Rochester, MN

Abstract Body:

The *LMNA* gene encodes both the lamin A and lamin C structural components of the nuclear envelope, associated with progeria syndromes and with forms of cardiomyopathy, muscular dystrophy, lipodystrophy, and neuropathy. Pathogenic variants resulting in changes of the nuclear structure and chromatin architecture may lead to altered gene expression in affected tissues. A group of *LMNA*-linked progerias, has been classified as "atypical" for not altering lamin A processing and for having an older age of onset. One example is the LMNA-associated cardiocutaneous progeria syndrome (LCPS) described only in one family with prominent cardiovascular and cutaneous manifestations segregating with the p.Asp300Gly variant. **METHODS:** This is a case report describing a family with the LCPS phenotype. Informed consent for publication was obtained. Functional testing was performed to assess nuclear morphology by transfection and overexpression of LMNA expressing constructs tagged with RFP for both wild type LMNA and p.(Glu2Lys) LMNA. Cells were imaged on a confocal microscope four days post transfection to assess the effect of overexpression of LMNA on nuclear morphology. **RESULTS:** The proband is a 36-year-old female with personal and family history of severe calcific aortic stenosis along with a calcified mitral valve with insufficiency. The proband reported greying hair in her 20s. She presently has no evidence of lipodystrophy or solar keratoses. Exome sequencing reported a VUS in LMNA c. 4G>A p.(Glu2Lys) (NM_170707.2) located in a conserved amino acid and absent in the gnomAD database. The variant was upgraded to Likely Pathogenic after segregation studies demonstrated variant presence in three other affected relatives. Indirect immunofluorescence analysis of nuclei showed abnormal morphology including lobulation and occasional ringed nuclei as similarly previously described for p.Asp300Gly variant. **CONCLUSION:** Here we report a second family with a novel variant in *LMNA*, with a phenotype of prominent cardiovascular and cutaneous manifestations similar to what is described for LCPS. Preliminary findings show suggest this variant affects nuclear morphology in a manner comparable to what was described in the first LCPS family. Additional functional studies are planned after generation of stable cell lines expressing p.(Glu2Lys) *LMNA*; *it is anticipated that these future studies will help to elucidate* the effect of this variant on nuclear structure. Characterizing the atypical forms of LMNA-linked progerias is important to understand the clinical course of these patients.

Molecular Effects of Genetic Variation Posters - Thursday

PB2664*. Organ-specific prioritization of non-coding regulatory variants with stacking generalization.

Authors:

N. Zhao, S. Dong, A. Boyle; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Understanding the functional consequences of genetic variation in the non-coding regions of the human genome remains a challenge. Recently, we have introduced a computational tool, TURF, to prioritize regulatory variants with tissue-specific functions by leveraging evidence from functional genomics experiments, including over three thousand functional genomics datasets in Hg19 from the ENCODE project provided in our RegulomeDB database. However, TURF was optimized to predict well-studied cell lines/tissues, which lacks the power to generalize to unseen cell lines apart from the training cell lines or at a higher organ-specific level. In addition, new experimental and deep learning-predicted features are available through RegulomeDB in Hg38 which, for example, contains ~ 5X more TF ChIP-seq datasets than in Hg19. Here, we present T-Land, a flexible model architecture to predict non-coding regulatory variants in a tissue or organ-specific manner with the stacking generalization algorithm. New features in Hg38 were bagged into biological meaningful feature subspaces; stacking the classifiers which individually takes input as those feature subspaces prevents overfitting and makes robust predictions. Tissue-specific T-Land outperformed TURF and other top-performing models when evaluating in hold-on cell lines, and on independent MPRA datasets. We developed our organ-specific T-Land by integrating newly designed organ-specific features, which outperformed tissue-specific T-Land even on tissue-specific evaluation datasets. To our best knowledge, it is the first model to prioritize non-coding regulatory variants for 51 organs available on ENCODE. Organ-specific T-Land prioritized relevant organs of various GWAS traits by predicting higher enrichment scores for trait-related SNPs. Interpreting organ-specific T-Land through SHAP value analysis portrays the cross-prediction relationships among five histone marks across organs. An organ map integrated with the new organ-specific scores in RegulomeDB would provide a valuable source for the initial screen of variants of interest for future studies.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2665. Parallel functional screening of human duplicated genes in neurodevelopment at single-cell resolution using zebrafish

Authors:

J. Uribe-Salazar, G. Kaya, N. Mariano, C. Ingamells, M. Y. Dennis; Univ. of California, Davis, Davis, CA

Abstract Body:

Gene duplication is a fundamental source of innovation in the evolution of phenotypic features. In humans, while several notable human-specific duplicated (HSD) genes have been implicated in key neurodevelopmental processes, most genes remain uncharacterized. We evaluated ten HSD genes (including previously characterized *SRGAP2* and *ARHGAP11*) using zebrafish, a model that allows for robust parallel tests of function. Generating mosaic knockout or knockdown embryos for each zebrafish ortholog, we assayed morphometry at 3 days post-fertilization (dpf) and observed significant differences in measured features for numerous crispant mutants (*ptpn20*, *pdzk1*, *hydin*, *gpr89*, *srgap2*, and *fam72*) and a morphant (*arhgap11*) versus controls, including increased head width in *hydin* crispants matching results in knockout mice exhibiting hydrocephalus. We also performed single-cell (sc)RNA-Seq using combinatorial indexing of ~130,000 total cells prepared from dissociated heads of crispants/morphants at 3 dpf. For *srgap2* crispants, we observed significant differences in abundances of various cell types (~2,400 cells per biological replicate, n=3), including optic progenitors as well as excitatory (E; Glut+) and inhibitory (I; GABA+) neurons versus scrambled-injection controls; the latter results were subsequently validated as a shift in E:I ratio in transgenic fluorescent reporter lines and an increased incidence of high-speed motion events following drug-induced seizures. Interestingly, these same results were observed for larvae injected with human *SRGAP2C* mRNA, indicating that the HSD paralog antagonizes the conserved zebrafish ortholog, as also reported in mice. Work is ongoing to characterize the additional nine HSD crispant/morphant larvae, for which we have successfully generated scRNA-seq data, as well as “humanized” fish injected with mRNA of the human paralogs. Overall, this study represents the first to systematically test the functions of multiple HSD genes in parallel, with the ultimate aim to discover novel genes contributing to neurological traits and disorders in humans. Further, our combined use of gene editing coupled with scRNA-seq at scale to identify changes in the brain can be broadly applied to test functions of any gene important in neurodevelopment.

Molecular Effects of Genetic Variation Posters - Thursday

PB2666. Pathogenic mechanisms associated with a recurrent *CUX2* missense variant in epilepsy.

Authors:

J. Cheng, K-M. Lamar, A. Ramamurthy, J. Calhoun, G. Carvill; Northwestern Univ., Chicago, IL

Abstract Body:

Developmental and epileptic encephalopathies (DEEs) are a group of severe, pediatric epilepsies characterized by refractory seizures and associated intellectual disability. DEEs have a heterogeneous genetic etiology; pathogenic variants have been found in genes involved in chromatin remodeling, ion channel activity, and neurotransmitter release. However, the role of transcription factors (TFs) in epilepsy is less well understood. We focus on a recurrent missense variant (c.1768G>A, p.Glu590Lys) in the TF, *CUX2*, that we identified as a novel cause of DEE. *CUX2* has three CUT domains and a homeodomain which confer specificity for binding DNA to regulate gene expression during cortical neurodevelopment. The c.1768G>A variant is a heterozygous *de novo* variant that changes a glutamate in the first CUT domain to a lysine. This residue is highly conserved, and *in silico* analysis predicts the variant as damaging. Given that DNA is negatively charged, it is likely that the positively-charged lysine variant enhances *CUX2* binding to DNA in a gain-of-function manner, altering gene expression critical for proper neurodevelopment. To test this, we use a patient-specific induced pluripotent stem cell (iPSC) derived neuronal model, as well as an isogenic *CUX2* knockout and sex-matched, wild-type (WT) iPSC lines. We differentiated these lines to neural progenitor cells (NPCs). We assessed *CUX2* genome occupancy in the WT NPCs using CUT&RUN and determined that *CUX2* binds primarily to promoter regions related to genes involved in DNA and RNA maintenance processes. Additionally, we conducted RNA-seq on the variant, knockout, and WT NPC lines. Differential gene expression analysis between the variant and WT NPCs revealed the upregulation of genes significantly enriched in GABAergic interneuron development and neural morphogenesis. In the next few months, I will integrate the results of variant *CUX2* occupancy with the differentially expressed genes, thus identifying direct *CUX2* target genes and biological pathways. Given our preliminary results and the previously established role of *CUX2* in neurodevelopmental processes, we will assess the effects of the variant on neuronal differentiation, survival, and synapse abundance over the next year. Combined, these studies will reveal the pathogenic mechanism of this missense variant and lay the foundation for studying other TFs with missense variants in patients with neurodevelopmental disorders. Overall, our results will expand our knowledge of the biological mechanisms that underpin DEEs and may reveal pathways and genes that could provide novel targets for the development of therapeutics against these disorders.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2667*. Pathogenic variants in *PLEKHO2* predispose to heritable thoracic aortic disease

Authors:

X. Duan¹, D-C. Guo¹, A. Pinard¹, I. C. Marin¹, G. Zhu², H. Yang², L. Zhang³, E. M. Hostetler¹, M. J. Bamshad⁴, D. A. Nickerson⁴, Z. Zhou², D. Milewicz¹; ¹Dept. of Internal Med., The Univ. of Texas Hlth.Sci. Ctr., Houston, TX, ²Fuwai Hosp., Chinese Academy of Med. Sci. and Peking Union Med. Coll., Beijing, China, ³The Fifth Med. Ctr., Chinese PLA Gen. Hosp., Beijing, China, ⁴Dept. of Pediatrics and Dept. of Genome Sci., Univ. of Washington, Seattle, WA

Abstract Body:

While 11 genes have been identified for heritable thoracic aortic disease (HTAD), the majority of HTAD families have yet to be solved. A rare heterozygous variant, p.Gly392Glu, in *PLEKHO2* that was present in 1/125,000 anomAD v2.1.1 exomes and is predicted to be damaging, segregated with HTAD in a large family with an LOD score of 3.26. Analysis of exome sequencing data from affected probands of 392 unsolved HTAD families and from 545 unrelated individuals with early-onset sporadic thoracic aortic dissections \leq 60 years of age identified four additional rare and predicted damaging *PLEKHO2* missense variants. *PLEKHO2* encodes a protein with no established function. Smooth muscle cells (SMCs) explanted from the aorta of a person with *PLEKHO2* p.Gly392Glu showed significantly increased expression and protein levels of SMC contractile genes and reduced proliferation compared to that of wild type (WT) SMCs. Similar changes were observed in SMCs with *PLEKHO2* knockdown using shRNAs. *Plekho2*^{-/-} mouse aortas had increased expression and protein levels of SMC contractile genes and thickened medial layers than WT aortas. In *PLEKHO2* knockdown SMCs, there were no alterations in phosphorylation of SMAD2/3 with TGF- β 1 exposure. Instead, non-canonical TGF- β signaling was constitutively activated in the absence of TGF- β 1 exposure based on increased phosphorylation of TAK1 and p38 MAPK, but not JNK or p65. Inhibitors of TAK1 and p38 MAPK, along with knockdown of TRAF6 using shRNAs, decreased p21 and SMC contractile gene expression in *PLEKHO2* knockdown SMCs. Co-immunoprecipitation assays indicate that PLEKHO2 binds to TRAF6 and TAK1 and controls TAK1 activation. Our results indicate that loss of PLEKHO2 function leads to constitutive activation of TAK1-P38 MAPK signaling in SMCs, which ultimately predisposes to HTAD. Moreover, they emphasize the role of activation of TRAF6-TAK1-p38 MAPK in the molecular pathogenesis of thoracic aortic disease.

Molecular Effects of Genetic Variation Posters - Thursday

PB2668*. Patient-derived iPSC-CMs from TANGO2-deficient disorder revealed multi-channel defects, functional association between ERK1/2 and TANGO2, and folate as a potential therapeutic agent to mitigate arrhythmias in patients

Authors:

W. Xu¹, Y. Cao¹, A. Yu¹, S. Lalani², C. Miyake¹, L. Zhang¹; ¹Baylor Coll. of Med., Houston, TX, ²Baylor Coll. Med., Houston, TX

Abstract Body:

TANGO2-deficient disorder (TDD) is a rare recessive multiorgan genetic disease caused by biallelic loss-of-function (LOF) mutations in *TANGO2* (Transport and Golgi Organization protein 2) gene. The exons 3-9 deletion ($\Delta E3-9$) and G154R are two common mutations found in patients with European and Hispanic/Latino ancestry, resulting in no and reduced protein expression, respectively. Despite normal cardiac function at baseline, metabolic stresses such as fasting can cause cardiac arrhythmias, which is the leading cause of death in TDD due to poor response to standard antiarrhythmic medications. The molecular function of TANGO2 is poorly understood, and a cardiac model for studying cardiac crisis is lacking, which largely hinders the research for TDD. We generated multiple independent TDD patient-derived iPSC-CMs lines, along with WT isogenic controls by adenoviral expression or CRISPR editing. The TANGO2-deficient iPSC-CMs lines recapitulated key arrhythmic phenotypes typically seen in TANGO2 patients, which were fully reversed by ectopic expressing WT-TANGO2. First, a reduction in field potential (FP) spike amplitude that can be exacerbated by fasting, which resembles Brugada-like EKG indicating LOF in cardiac Na⁺ channel SCN5A. Second, prolonged FP duration which represents prolonged QT interval indicating LOF in K⁺ channel or gain of function in Na⁺ channel. Third, premature ventricular contractions (PVCs) caused by early afterdepolarization, which is primarily due to Ca²⁺ channel or Ca²⁺ handling defect. Collectively, our data suggest that TANGO2 deficiency leads to multi-channel defects in cardiomyocytes that results in lethal arrhythmias in TDD patients. Mechanistically, we detected reduced membrane fraction of SCN5A and KCNH2 in TANGO2-deficient iPSC-CMs, which explains the Brugada and prolonged QT EKG patterns, respectively. Moreover, ERK1/2 inhibition led to a TANGO2-dependent increase in FP spike amplitude which plateaued at 24hr, while causing PVCs in iPSC-CMs with $\Delta E3-9$ -TANGO2. These data suggest ERK1/2 signaling pathway is associated with TANGO2 function. Interestingly, ERK inhibition also increased FP spike amplitude in iPSC-CMs expressing G154R-TANGO2 but with a delayed rising time and plateaued at 36hr. We speculate that ERK inhibition may increase the stability of G154R-TANGO2 to restore its function. Therapeutically, we observed Vitamin B supplement dramatically reduced metabolic crises in TANGO2 patients. We show that folic acid (B9) abolished the PVCs and intermittent long pauses in TANGO2-deficient iPSC-CMs. Therefore, folate may be a potential therapeutic agent to mitigate cardiac arrhythmias for TDD patients.

Molecular Effects of Genetic Variation Posters - Thursday

PB2669*. Phenotypic and functional genomic impact of complex structural variation at the 17q21.31 locus.

Authors:

P. Zhang¹, R. Handsaker², L. Hernandez³, C. Wen¹, C. Jops¹, M. Kim⁴, M. Margolis¹, T. Werge⁵, W. Thompson⁶, S. McCarroll⁷, M. Gandal¹; ¹UCLA, Los Angeles, CA, ²Broad Inst., Charlemon, MA, ³Univ. of California, Los Angeles, Los Angeles, CA, ⁴David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ⁵Lundbeck Fndn. Initiative for Integrative Psychiatric Res., Copenhagen, Denmark, ⁶Laureate Inst. for Brain Res., Tulsa, OK, ⁷Harvard Med. Sch., Boston, MA

Abstract Body:

Post-GWAS fine-mapping efforts mainly focused on prioritizing among multiple small variants, ignoring potential contributions from other forms of genetic variation, such as complex structural variants (SVs). Although SVs have well established contributions to complex traits, their contribution to GWAS results remains underexplored, due to the limitations in ascertainment and a lack of good methods for imputing these variants at large scale. As a result, some significant SNPs identified from GWAS may fail to be correctly fine mapped to nearby SVs, especially at complex genomic loci with extended linkage disequilibrium (LD). The 17q21.31 locus contains an inversion polymorphism, present in ~20% of Europeans, surrounded by several multiallelic copy number variation (mCNV) regions. SNPs within 17q21.31 locus have been found to be associated with a variety of traits and diseases, including neuroticism, body fat and height measures and blood cell counts, but whether these SNPs are tagging nearby SVs remains unknown. To evaluate the contribution of SVs at 17q21.31 locus to phenotypic associations, we imputed the inversion configuration and mCNV copy numbers into the UK Biobank (UKBB) and the Adolescent Brain Cognitive Development (ABCD) study, and performed phenome-wide association studies (PheWAS) taking into account of both SNPs and SVs at this locus. We found that SVs account for about half of the observed GWAS hits at this locus, with distinct SV-phenotype associations observed across a host of traits. The inversion captured the majority of trait associations - including neuroticism and blood cell counts. However, controlling for the inversion, mCNVs also served as lead variants for several traits including body fat and reaction time measurements. To further investigate the molecular mechanisms by which SVs regulates traits/diseases, we imputed SVs into several functional genomics reference panels, including PsychENCODE, GTEx and AMP-AD cohorts and analyzed the impact of SVs at this locus on chromatin accessibility, gene expression, local splicing and protein abundance across a host of tissues including the human brain. We found that 17q21.31 SVs had broad functional impact at different molecular levels, both *in-cis* and *in-trans*. Through colocalization and Mendelian randomization analyses, we pinpointed the potential molecular cascade through which 17q21.31 SVs regulate different phenotypes. Altogether, our results highlight the importance of accounting for complex structural variation when interpreting GWAS signals at a complex locus and identify new insights into the biological mechanisms underlying the pleiotropic effects at 17q21.31.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2670. PIEZO2 as a molecular candidate of respiratory phenotype induced by Congenital Central Hypoventilation Syndrome mutations in *PHOX2B*.

Authors:

T. Hedgecock¹, K. Victor¹, L. O. Romero¹, J. Cordero-Morales¹, C. M. Rand², D. E. Weese-Meyer², L. T. Reiter¹; ¹UTHSC, Memphis, TN, ²Ann and Robert H. Lurie Children's Hosp., Chicago, IL

Abstract Body:

Congenital central hypoventilation syndrome (CCHS) is an autosomal dominant disease that affects the autonomic nervous system (ANS). Individuals with CCHS present with a variety of clinical manifestations including impaired respiration, cardiac sinus pauses, neural crest tumors, gastrointestinal dysmotility, and in some cases, Hirschsprung disease. CCHS is caused by polyalanine repeat mutations (PARMs) and non-PARMs (NPARMs) in the gene *PHOX2B*. PARMs can range in alanine repeat number from less severe (20/24 genotype) to the need for continuous ventilation for 27-33 alanines on the affected allele (20/27-20/33 genotype). The *PHOX2B* gene encodes a transcription factor that regulates ANS neural development. To date, there are no direct molecular ties between *PHOX2B* mutations and the clinical ventilatory slope. CCHS respiratory defects observed in humans and in animal models are reminiscent of the respiratory phenotype found in mice with a knockout (KO) for the *Piezo2* gene. PIEZO2, a mechanosensitive cation channel, plays a known role in maintaining baroreflex and respiratory homeostasis. Using neurons differentiated from neurotypical and CCHS dental pulp stem cells (DPSC), *PIEZO2* and *PHOX2B* mRNA transcripts were analyzed using qPCR to determine if there is a relationship between *PHOX2B* and *PIEZO2* transcription, as well as whether PARM number influences *PHOX2B* and *PIEZO2* transcription. Additionally, PIEZO2 currents were examined to determine if PARM number impacts PIEZO2 activity in CCHS patients compared to neurotypical individuals. We found that PIEZO2 levels are significantly decreased in CCHS neurons versus neurotypical neurons. These findings will provide new possibilities for the treatment of CCHS by restoring PIEZO2 function.

Molecular Effects of Genetic Variation Posters - Thursday

PB2671. Pooled RNA-IP approach to investigate variant effects on RBP binding and splicing.

Authors:

M. Schertzer^{1,2}, P. Halmos¹, K. Dobrindt³, K. Brennand³, D. Knowles^{1,2}; ¹New York Genome Ctr., New York, NY, ²Columbia Univ., New York, NY, ³Yale Univ., New Haven, CT

Abstract Body:

Alternative splicing is a fundamental cellular process that regulates 95% of multi-exon genes to diversify protein output and define cell-type specific functions. Both constitutive and alternative splicing are controlled by combinations of cis-acting pre-messenger RNA sequences and trans-acting RNA-binding proteins (RBPs), including splice factors. Therefore, variation in splicing regulatory elements or RBPs can be highly disruptive to basic cellular activities and often lead to disease, including neurological and muscular disorders as well as various cancers. We and others have identified thousands of genetic variants that are associated with splicing changes: splicing quantitative trait loci (sQTLs). For many complex traits, sQTLs appear as enriched as expression QTLs. The causal variants for most sQTLs and their intermediate molecular effects are largely unclear. Although it is expected that many sQTLs function by altering binding of one or more RBPs, there is very limited experimental data on how genetic variants alter RBP binding to assess this hypothesis. The goal of this project was to identify binding QTLs for a subset of RBPs and, then, to mechanistically link genetic variants with splicing outcome by overlapping binding and splicing QTLs. To identify sites of allele-specific binding (ASB), where a variant increases or decreases RBP binding, in a high throughput manner, we generated data using a pooled cell line approach and developed a corresponding analysis pipeline. First, we pooled 9 induced pluripotent stem cell lines (iPSCs) and performed RNA immunoprecipitations (RNA-IPs) for three splicing factors, HNRNPK, RBFOX2, and HNRNPA1. Next, to estimate cell line proportions in the pooled data, we imputed genotypes for the 9 cell lines and used constrained linear regression to estimate cell line proportions. Finally, we developed a Bayesian probabilistic model to distinguish sites of allele-specific binding from allele-specific expression via analysis of pre-IP and post-IP RNA allelic counts. At a 10% false discovery rate, we identified thousands of variants that exhibit ASB (binding QTLs). In support of a mechanistic link between these binding QTLs and splicing, we detected a strong overlap with fine-mapped sQTLs from 1,367 iPSC lines. Collectively, our work provides a novel high-throughput method for identifying binding QTLs which will significantly improve interpretation of splicing QTLs and disease-associated variants.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2672. Population genetic study for 16 autosomal STR loci in unrelated 15,546 Korean individuals

Authors:

H. Park, Y-H. Sohn; Seegene Med. Fndn., Seoul, Korea, Republic of

Abstract Body:

Genotyping of autosomal short tandem repeat (STR) markers is widely used for the genetic identification of individuals in paternity disputes, acquirement of nationality of refugees, forensic DNA analyses in Korea. Our institute has performed genetic identification of individuals over than 15 years. Hereupon we analyze allele frequencies of STR of large Korean population. DNA typing was performed on 31,092 Koreans submitted to our institute during 2009-2022. Every one sample from submitted pairs was excluded for avoid any selection bias, so 15,546 results were used for analysis. Allele frequencies of 16 autosomal STR loci (*D8S1179*, *D21S11*, *D7S820*, *CSF1PO*, *D3S1358*, *TH01*, *D13S317*, *D16S539*, *D2S1338*, *AMEL*, *D5S818*, *FGA*, *D19S433*, *vWA*, *TPOX*, *D18S51*) were obtained using the AmpF/STR Identifier PCR Amplification kit. For all 16 autosomal STR markers, no deviations from the Hardy-Weinberg equilibrium were observed. In addition, power of discrimination, mean exclusion chance, polymorphism information content of each locus were calculated using Arlequin 3.5 software (Excoffier & Lischer, 2010). This is the largest STR profile dataset will be a useful tool for forensic identification & parental testing in Korea.

Molecular Effects of Genetic Variation Posters - Thursday

PB2673. Population-specific non-coding and coding putative causal variants shape quantitative traits

Authors:

X. Liu¹, S. Koyama^{1,2,3}, Y. Koike^{1,4}, K. Tomizuka¹, K. Hikino¹, M. Koido⁵, S. Ito^{1,6}, N. Otomo¹, H. Suetsugu^{1,7}, S. Yoshino^{1,7}, M. Akiyama⁸, C. Benner⁹, P. Natarajan^{2,3}, P. Ellinor^{2,3}, T. Mushiroda¹, M. Horikoshi¹⁰, M. Ikeda¹¹, N. Iwata¹¹, The Biobank Japan Project, S. Niida¹², K. Ozaki¹², Y. Kamatani¹³, Y. Momozawa¹, S. Ikegawa¹, K. Ito¹, C. TERAO^{1,14,15}, ¹RIKEN, Yokohama, Japan, ²Massachusetts Gen. Hosp., Boston, MA, ³Broad Inst., Boston, MA, ⁴Hokkaido Univ., Sapporo, Japan, ⁵The Univ. of Tokyo, Tokyo, Japan, ⁶Shimane Univ., Izumo, Japan, ⁷Kyushu Univ., Fukuoka, Japan, ⁸Kyushu Univ., Fukuoka, Japan, ⁹Univ. of Helsinki, Helsinki, Finland, ¹⁰RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, ¹¹Fujita Hlth.Univ., Toyoake, Japan, ¹²Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, ¹³The Univ. of Tokyo, Minato-ku, Japan, ¹⁴Shizuoka Gen. Hosp., Shizuoka, Japan, ¹⁵Univ. of Shizuoka, Shizuoka, Japan

Abstract Body:

Human genetic variants associate with numerous human traits through alterations in gene function and regulation. Determining these functional variants in a large-scale genetic study may help elucidate the causes of disease or traits. In this study, combining Japanese-specific genotype imputation with a novel reference panel and statistical fine-mapping methods, we identified 826 and 9,406 putatively causal genetic associations (posterior probability of inclusion >90% and >10%) in 3,309 loci across 63 quantitative traits among 203,216 Japanese individuals. These putative causal variants are not always the lead variants of the loci, but tend to be rare, functional coding or non-coding variants in medically relevant genes/tissues and specific to the Japanese population. These variants include known and unrevealed pathogenic variants in monogenic disorders shaping the relevant quantitative traits. A series of putative causal non-coding variants were supported by state-of-art in-silico functional assays and had comparable effect sizes to coding variants. Namely, a Japanese-specific cryptic splicing variant in the intron of *FLT3* had a large impact on various immunological traits. Our study provides a list of fine-mapped causal variants to be tested for functionality and provides insights into the strategy to detect causal variants in the associated loci. Further, our approach underscores the importance of sequencing/genotyping and associating efforts in diverse populations to discover causal variants for diseases/traits, which can be a point of intervention in global populations.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2674. Population-specific RNA-seq data empowers the functional follow-up of disease-associations from FinnGen

Authors:

C. Benoit-Pilven¹, E. Vartiainen¹, S. F. Qadri², K. M. Donner³, K. Tuomainen³, M. Daly¹, H. Yki-Järvinen², T. Tukiainen¹; ¹Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, ²Dept. of Internal Med., Helsinki Univ., Helsinki, Finland, ³FIMM genotyping services, Inst. for Molecular Med. Finland, Helsinki, Finland

Abstract Body:

The Finnish population is unique in its genetic characteristics compared to other European populations. This property, together with the longitudinal health records available in Finland, are utilized in the FinnGen project to discover disease-associated genetic variation that can point to new therapeutic avenues. As expected, many FinnGen findings indicate highly Finnish-enriched variants which are typically absent or present in very low frequencies in datasets of broader European origin. To facilitate the functional interpretation of FinnGen discoveries, we generated a dataset combining liver RNA-seq and genotypes from 127 Finnish individuals. To allow for the examination of coding variant effects on the gene transcripts, we carried out an allele-specific expression (ASE) analysis on this high-quality data and found on average, 12200 informative heterozygous sites per sample (with a minimum coverage of 20 reads). Out of the 14876 coding variants associated ($p < 1e-5$) to a FinnGen (release 9) health registry endpoint, 2760 single-nucleotide variants had reliable ASE data from at least two samples. In aggregate, stop-gain variants ($N=18$) showed most pronounced allelic imbalance (AI) with an average of 63% bias for the reference allele ($p=0.02$), in line with these variants inducing nonsense mediated decay. As an example, our data confirms the loss of function impact of a low frequency variant (rs2270416) in CDH15 associated with varicose veins ($AI=0.34$). However, some variants predicted as stop-gain variant did not show the expected allelic imbalance (e.g. stop-gain variant (rs41257904) in CFHR2 associated to Age-related macular degeneration with $AI=0.03$), pointing to potential tissue-specific mechanisms. Focusing on the low frequency ($AF < 10\%$) Finnish-enriched (>2 -fold enrichment) missense variants ($N=425$), 77 show significant (proportion test $qvalue < 0.01$) allelic imbalance (>0.15), indicating impacts on the transcript abundance. Further, some of these missense variants show only expression from the reference allele, suggesting that they are partial loss-of-function variants. For example, a low frequency missense variant (rs142351376) in STAB2 gene associated with venous thromboembolism has an AI of 0.38 and another rare missense variant (rs150414818) in ALDH16A1 gene associated with gout has an AI of 0.34. Overall, we show that a Finnish RNA-seq dataset empowers us to define and confirm the functional consequences of disease-associated Finnish-enriched variants, that are too rare to be studied in datasets of broader European origin. By refining the variant consequences, these results can help to inform therapeutic development.

Molecular Effects of Genetic Variation Posters - Thursday

PB2675. Possible complex Di-Genetic inheritance in Primary Congenital Glaucoma

Authors:

D. Bercovich; Tel Hai Academic Coll., Tel Hai, Israel

Abstract Body:

Primary Congenital Glaucoma (PCG) generally have strong genetic contributions and many disease-causing mutations that lead to pathogenesis have been identified like the CYP11B1, LTBP2, MYOC, FOXC1, GPATCH3, and TEK genes and in patients with early-onset secondary glaucoma PAX6, PITX2, PITX3, FOXC1, FOXE3, EYA1, LMX1B, and MAF. To between 10-20% of patients, no mutations are found in these genes. Because glaucoma is a complex, heterogeneous disease likely to be the consequence of the interaction of multiple genes. This underlying complexity may hinder efforts to identify glaucoma-associated genes (or related mutations) and to uncover their pathogenic mechanisms. Identification of new glaucoma inheritance models, may assist in addressing both of these issues. To this end, screening the full Exom of PCG patients and family members can help to identified a glaucoma models caused by a mutation's new genes. Whole Exom sequencing on 26 PCG patients and family members (a total of 72 participants), with no mutations in the genes list above, reviled possible complex Di-Genetic inheritance in 9 families. Of the 18 pathogenic heterozygous mutations, 9 are nonsense mutations, 3 mutations are Indel (small deletions), 3 are splicing mutations & 3 are rare missense mutations. These genes can be found in the literature or data bases as been related to the possible development of glaucoma like the progressive death of retinal ganglion cells and atrophic excavation of the optic nerve, Elevated intraocular pressure, defects in outflow of aqueous humor aqueous humor secretion, by the ciliary body, and drainage through the trabecular meshwork, a porous tissue located in the iridocorneal angle. Finding new genes and inheritance models in PCG may contributed to our knowledge of glaucoma and can help clinicians to preform prenatal screening.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2676. Predicted nasal epithelial transcriptome-wide association study in African-ancestry populations

Authors:

R. K. Johnson¹, E. Esquinca¹, M. Boorgula¹, B. Szczesny², A. Romero¹, M. Campbell¹, S. Chavan¹, N. Rafaels¹, I. Ruczinski², K. Kammers³, H. Watson⁴, R. C. Landis⁴, M. A. Taub⁵, M. Daya¹, N. Hansel⁶, C. N. Rotimi⁷, C. O. Olopade⁸, C. A. Figueiredo⁹, C. Ober⁸, A. H. Liu¹⁰, E. Kenny¹¹, R. A. Mathias¹², K. C. Barnes¹, Barbados Asthma Genetics Study (BAGS), Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA); ¹Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ²Johns Hopkins Univ., Baltimore, MD, ³Johns Hopkins Univ Sch. of Med., Baltimore, MD, ⁴Faculty of Med. Sci., The Univ. of the West Indies, Queen Elizabeth Hosp., St. Michael, Barbados, ⁵Johns Hopkins, Baltimore, MD, ⁶Johns Hopkins Med., Baltimore, MD, ⁷NIH, Woodstock, MD, ⁸Univ. of Chicago, Chicago, IL, ⁹Federal Univ. of Bahia and Fundação Program for Control of Asthma in Bahia (ProAR), Salvador, Brazil, ¹⁰Children's Hosp. Colorado, Aurora, CO, ¹¹Icahn Sch. of Med. at Mt Sinai, New York, NY, ¹²Johns Hopkins Univ, Baltimore, MD

Abstract Body:

The nasal epithelium plays a central role in modulating asthma, but this tissue is poorly represented in repositories such as GTEx, particularly for African-ancestry populations that are disproportionately affected by asthma. Using African-ancestry populations, we built and applied predictive models to estimate genetically driven gene expression in the nasal epithelium, and identified associations between predicted gene expression and asthma. We trained gene expression prediction models using RNA-seq data generated from nasal epithelial samples and MEGA genotypes from 536 participants (253 asthma cases, 283 controls) of phase 2 of the *Consortium on Asthma among African-ancestry Populations in the Americas* (CAAPA). Using cross-validated elastic nets, we predicted gene expression from SNPs within a 1Mb window of each gene's start and stop position, adjusting for sex, asthma, ancestry PC1, and 60 PEER factors. We used these models to predict nasal epithelial gene expression from TOPMed WGS available among 920 participants (426 asthma cases, 494 controls) of the Barbados Asthma Genetics Study (BAGS) and discovered transcriptome-wide associations (TWAS) with asthma using linear mixed models adjusted for sex, age, PC1, and kinship. Out of 21,144 autosomal genes, 9,407 were significantly predicted from cis-SNPs (model $R^2 > 0.01$ and $P < 0.05$). From TWAS in BAGS, six genes were associated with asthma (q -value < 0.1). Asthma cases had lower nasal epithelial expression of *SCG3* ($P = 6E-5$), *UGGT1* ($P = 4E-5$), *TRIO* ($P = 1E-5$), and *KPNA5* ($P = 2E-5$), and higher expression of *SORBS1* ($P = 5E-5$) and *UFD1* ($P = 2E-5$). *SCG3* is located in the 15q21 locus near asthma candidate genes *RORA* and *SMAD3*. None of the other five genes were identified in prior asthma GWAS; *KPNA5* and *SORBS1* were implicated in prior asthma expression studies. While *SORBS1* encodes an adaptor protein primarily involved in insulin signaling, *SORBS1* transcript usage has been shown to be higher in lung tissue of COPD cases compared to controls. The *UGGT1*-encoded protein glucosyltransferase 1 performs quality control of misfolded proteins in the endoplasmic reticulum (ER); its increased expression in the nasal airway may reflect ongoing ER stress characteristic of asthma. Identification of expression signatures associated with asthma in the nasal epithelium recapitulated candidate asthma genes and identified novel candidate genes with biologically plausible mechanisms of action, providing new insight into the molecular mechanisms driving dysfunction in asthma.

Molecular Effects of Genetic Variation Posters - Thursday

PB2677. Predicting functional and fitness effect of missense variants

Authors:

Y. Zhao, Y. Shen; Columbia Univ., New York, NY

Abstract Body:

Missense variants are the most common type of single nucleotide variants in protein coding regions, but most of them have uncertain functional effect. Accurate prediction of functional effect of missense variants is critically important to improving statistical power in new risk genes discovery of human diseases and diagnosis in clinical testing. Most of existing computational methods rely on pathogenicity labels from curated databases in supervised training. This leads to bias in performance since curated databases are noisy and over-represented with well-studied genes. Moreover, pathogenicity is dependent on implicated diseases and conditions. It is conceptually ambiguous as a generic score in the absence of disease or conditions. Here, we propose a probabilistic method which jointly and explicitly estimates the effects on protein function (damageness, d) and the fitness effect in human population (selection coefficient, s). We argue that d and s have more direct damaging variant in a protein that is implicated in a disease is pathogenic to the disease, and the aggregated fitness effect of all diseases or conditions determine the selection coefficient in human population. We show that in each gene, pairs of variants with strong coevolution strength are more likely to have similar selection coefficient, and such information could be used to aggregate similar variants for estimating their selection coefficients s . We encoded the distribution of damageness d by a dense neural network given the gMVP latent embedding vectors. Further, we modeled the selection coefficient s is a sigmoid function over d , while scaled by gene-specific parameters. We selected C-to-T variants in 943 highly constrained genes and trained this model by maximizing the probability of observed allele counts in UKBB (2e5 samples). Prior to training, the probability of allele counts was already approximated as Inverse Gaussian distribution with parameters optimized to simulated European population as functions of mutation rate and heterozygous selection coefficient s . We found in genes TP53, BRCA1 and PTEN, the estimated mean of d given the protein context embedding, correlates better with the functional scores derived from deep mutation scanning experiments than most existing methods. The distribution of d for benign variants is also more consistent across genes. Finally, the estimated selection coefficient s predicts allele frequency in a second population better than the baseline models without information on damageness.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2678*. Profiling Local Ancestry Effects on Gene Regulation in An Admixed Population

Authors:

H-H. Chen¹, W. Zhu², H. Polikowsky², L. Petty², S. Fisher-Hoch³, J. McCormick³, K. North⁴, E. Gamazon¹, J. Below¹; ¹Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Vanderbilt Univ., Nashville, TN, ³The Univ. Of Texas Hlth.Sci. Ctr. at Houston, Brownsville, TX, ⁴Univ North Carolina, Chapel Hill, NC

Abstract Body:

Background: Although transcriptome regulation in humans has been investigated and used successfully to interpret genetic findings and map disease genes, most expression quantitative trait loci (eQTL) were estimated in European populations. While ancestry-specific effects on gene expression have been reported, environmental differences among populations confound efforts to identify the ancestry-specific eQTL effects. In this study, we leveraged haplotypic patterns resulting from recent admixture in Hispanics to identify eQTL and estimate their local-ancestry effect.

Methods: We sequenced 645 whole blood RNA samples from the Cameron County Hispanic Cohort. These individuals were genotyped by Illumina MEGA array and imputation was performed to the TOPMed reference. Global genetic ancestry analysis identified an average of 49.0% Amerindian (AMR), 45.7% European (EUR), and 4.6% West African in our samples, estimated by Admixture with 1000 Genomes enriched with 248 participants from the Hispanic Community Health Study / Study of Latinos as a reference panel. Local ancestry was detected by RFMix2 with the same reference. Haplotype-specific expression was estimated by Phaser. Then, both total aligned read counts and haplotype-specific aligned reads were used to identify ancestry-specific cis-eQTL (within ± 1 Mb of transcription start site) and estimate their effects. DAP-G was performed for cis-eQTL fine-mapping.

Results: Among the 16,093 genes that passed quality control, 2,565 genes are significantly differentially expressed between EUR haplotypes and AMR haplotypes after Bonferroni correction (p -value $< 3.1 \times 10^{-6}$) and covariate adjustment, including sex, age, five genetic principal components, RNA-seq batch, and 60 transcriptome PEER factors. We identified over 5.6 significant variants for 9,916 genes (FDR-adjusted p -value < 0.05), of which 7,463 have at least one local ancestry-specific eQTL. Among 78,169 ancestry-specific causal eQTL which passed fine-mapping (posterior inclusion probabilities > 0.1), 32,541 variants are only significant in AMR haplotypes (p -value in EUR > 0.1 and FDR-adjusted p -value in AMR < 0.05). In addition, 9,964 eQTL have opposite effects on gene expression in two ancestral haplotypes (both FDR-adjusted p -value < 0.05).

Conclusions: We generated the first profile of an ancestry-specific regulatory landscape in Hispanic/Latinos, a recently admixed population. These effects may contribute to inconsistent GWAS and TWAS findings across populations and emphasize the importance of diversity in genetic research.

Molecular Effects of Genetic Variation Posters - Thursday

PB2679*. Quantifying negative selection in human 3'UTRs uncovers deleterious non-coding genetic variation.

Authors:

S. Findlay¹, L. Romo², C. Burge³; ¹MIT, Cambridge, MA, ²Boston Children's Hosp., Boston, MA, ³Massachusetts Inst. of Technology, Cambridge, MA

Abstract Body:

Interpretation of non-coding genetic variants is a fundamental challenge in contemporary human genetics. In processed mRNA transcripts, the majority of non-coding variants are found in 3'untranslated regions (3'UTRs), which play integral roles in post-transcriptional regulation of gene expression. Most of this regulation is mediated by interactions between RNA-binding proteins (RBPs) and regulatory elements in the mRNAs they bind.

We used allele frequency data from gnomAD to identify classes of genetic variants in 3'UTRs under negative selection in humans. Since mutability also influences allele frequency spectra, estimates of negative selection have typically been calibrated using a set of "neutral" synonymous coding variants matched for dinucleotide context. However, synonymous variants are: 1) not truly neutral, and 2) routinely found within the regulatory elements we sought to examine. Thus, we developed intergenic MAPS (iMAPS), an enhanced version of the previously described mutability-adjusted proportion singleton (MAPS) metric used to quantify negative selection. By calibrating to a much larger set of intergenic variants, iMAPS captures the impact of extended nucleotide contexts known to affect mutability. This resulted in improved calibration for the majority (54%) of 3'UTR variants. This is especially important for regulatory elements in 3'UTRs which are known to converge on similar motifs of low sequence complexity.

We derived a set of high-confidence RBP binding sites across a diverse set of RBPs in 3'UTRs by intersecting RBP binding data obtained from cells (eCLIP) and in vitro (Bind-n-Seq). Variants in our high-confidence sites were enriched for: 1) variants altering steady state transcript levels in a 3'UTR MPRA, and 2) high-PIP eQTLs.

Variants that disrupted these high-confidence RBP binding sites in 3'UTRs were under levels of negative selection similar to (or even exceeding) missense coding variation. In contrast, variants preserving RBP-RNA interactions at the same sites experienced lower negative selection similar to synonymous variants. Variants disrupting a set of affinity-matched sites within the same 3'UTRs were under minimal selection, suggesting that sequence motifs alone are insufficient to identify RBP binding sites where disruption is likely to be deleterious and thus pathogenic.

We then tested variants from the high-confidence sites for their ability to modulate transcript levels in a 3'UTR MPRA in cells, directly identifying variants with regulatory activity.

Collectively, this work has nominated thousands of likely deleterious genetic variants that act by disrupting RBP-RNA interactions in human 3'UTRs.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2680. Rare protein-coding variants in the insulin receptor gene (*INSR*) identified in women with polycystic ovary syndrome (PCOS).

Authors:

R. Bauer¹, L. Gorsic^{2,1}, R. Legro³, G. Hayes¹, M. Urbanek¹; ¹Northwestern Univ., Chicago, IL, ²GeneDx, Gaithersburg, MD, ³MS Hershey Med. Ctr, Hershey, PA

Abstract Body:

Polycystic ovary syndrome (PCOS) is the most common form of anovulatory infertility among reproductive age women. In addition to reproductive symptoms, women with PCOS are at elevated risk of developing obesity, insulin resistance (IR), and type 2 diabetes. Severe Mendelian diseases of IR include Donohue Syndrome and Rabson-Mendenhall Syndrome, two disorders caused by recessive missense alleles in the insulin receptor gene (*INSR*). Women with these disorders of IR also experience symptoms of PCOS such as amenorrhea and hyperandrogenism. We therefore hypothesize that genetic variation in *INSR* also contributes to PCOS, which is a common form of IR. To test this hypothesis, we comprehensively screened 602 women with PCOS and 125 reproductively normal women for genetic variation in *INSR*. We identified 11 missense variants in *INSR* in 22 cases and 1 control subject. One of these variants, A119V, has previously been identified in Donohue Syndrome subjects. Another variant, V1012M, found in 9 PCOS subjects and zero controls, has been identified in insulin resistant diabetes mellitus with acanthosis nigricans. 10 out of 11 *INSR* variants are likely to be deleterious (CADD score \geq 15). There was no significant evidence for association between the set of all 11 *INSR* variants in the PCOS cohort relative to our phenotyped controls ($X^2 = 2.7$; p-value = 0.10; odds ratio (OR) = 4.61) or relative to the gnomAD non-Finnish European cohort ($X^2 = 1.03$; p-value = 0.31; OR = 1.25). In order to determine which of these variants are rare polymorphisms and which truly impact protein function and thus likely have a role in the development of PCOS, we have pursued cell-based studies to assess the effect of these variants on the insulin receptor's ability to achieve insulin-dependent signaling. Here we show preliminary investigation into the effects of the identified variants on the insulin signaling cascade in response to insulin stimulation.

Molecular Effects of Genetic Variation Posters - Thursday

PB2681. Regulation of stress granules by VCP in health and genetic degenerative diseases

Authors:

N. Helton, B. Dodd, S. Moon; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

The integrated stress response (ISR) is an evolutionarily conserved mechanism in eukaryotes that induces transcriptional and translational reprogramming in response to various stressors. These stressors can include nutrient deprivation, oxidative stress, buildup of unfolded proteins, or viral infection. During the ISR, global canonical translation is inhibited and RNAs are sequestered in mRNP condensates known as stress granules. Chronic ISR activation and aberrant stress granule-like aggregates are prevalent in tissues of patients with degenerative diseases of the muscular, skeletal, and nervous systems including amyotrophic lateral sclerosis and frontotemporal dementia. Valosin-containing protein (VCP), a hexameric AAA+ ATPase that plays many roles in proteostasis, is an important stress granule regulator. Mutant alleles of VCP associated with neurodegeneration impair stress granule composition and disassembly. These alleles correspond to progressive autosomal dominant adult late-onset multisystem diseases that are ultimately lethal after degeneration of muscle, bone, and brain. Therefore, understanding the molecular mechanism by which VCP regulates stress granules is critical for understanding its role in disease. We discovered that VCP indirectly regulates stress granules by mediating translation in a co-translational quality control pathway. We are using live cell fluorescence microscopy and ribosome profiling to elucidate the molecular mechanisms behind VCP-mediated stress granule regulation. Given that VCP is involved in many cellular processes, defining the mechanisms by which it regulates stress granules will advance our understanding of how stress granule behavior becomes impaired in common and devastating degenerative conditions.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2682. Regulatory rare variant trait associations in 127,724 UK Biobank genomes

Authors:

D. Ribeiro^{1,2}, D. Avalos^{3,2}, C. Ziyani¹, R. Hofmeister^{1,2}, O. Delaneau^{1,2}; ¹Univ. of Lausanne, Lausanne, Switzerland, ²Swiss Inst. of Bioinformatics, Lausanne, Switzerland, ³Univ. of Geneva, Geneva, Switzerland

Abstract Body:

A central goal of genetics is to shed light on the **genetic basis of polygenic human diseases and traits**. Genome-wide association studies (GWAS) have associated tens of thousands of common variants with traits, but these fail to account for the entire trait heritability. Due to their predicted stronger effects on phenotypes, the study of rare variation promises to be complementary to GWAS in linking genotype to phenotype. Indeed, the study of rare variation in exonic regions has proved instrumental in associating hundreds of genes to human traits. Yet, the impact of **rare variation in gene regulatory regions** - where most GWAS hits are found - has not been fully explored. The recent availability of the **UK Biobank whole genome sequences** for >125,000 individuals thus offers an unprecedented opportunity to assess the impact of rare genetic variation, particularly in the regulatory regions of the genome.

Here, we present a **novel approach to determine the effect of rare variation in gene regulatory regions** in multiple complex traits using 125,363 UK Biobank British genomes. To evaluate the use of different regulatory region annotations, we performed association testing between several **blood cell count and biomarker phenotypes** and rare variation (MAF<1%) present in three state-of-the-art gene regulatory annotations in blood cell types: (1) regulatory regions from **promoter capture Hi-C data**, (2) cis-regulatory domain predictions based on **ChIP-seq data** and (3) gene-enhancer associations from **multimodal single-cell data**. Applying association testing methods specialised for rare variation - including burden tests and SAIGE-GENE - on these three regulatory annotations, **we identified 251 gene-trait associations** (189 distinct genes, 125 loci) across the 12 blood traits tested. The three regulatory annotations shown to be complementary in finding associations. Importantly, **~40% of the gene-trait associations were replicated** in rare variant exome studies or GWAS studies. Moreover, >80% of the associations were kept when conditioning for known fine-mapped common variants, indicating that rare variants reveal **independent signals**. The discovery of hundreds of gene-trait associations - including many novel ones - at a level similar to what we found in exonic regions demonstrate that **gene regulatory regions can be highly informative of genotype-to-phenotype signals**.

Our results warrant further exploration of regulatory regions **across human cell types** as well as across the **thousands of phenotypes** available in biobanks. We expect these analyses to improve our understanding of complex human traits and ultimately aid the development of pertinent therapeutic agents.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2683. Role of SNPs of Leptin gene in Etiology of Venous Thromboembolism among Females contraceptive users in Pakistan Population

Authors:

Z. Agha, M. azam, A. Hashmi, K. Khalil; COMSATS Univ. ISLAMABAD, ISLAMABAD, Pakistan

Abstract Body:

Venous thromboembolism (VTE) is critical multifactorial disorder caused by various acquired and genetic factors. Combined hormonal contraceptives (CHCs) including estrogen and progestin, are one of the highly recommended methods of contraception for women and play critical role in increasing the risk for the females already susceptible to VTE genetically. It has been suggested that raised leptin levels may increase the risk of venous thromboembolism (VTE) in animal studies. However, clinical studies in this field are still largely unexplored. Thus the following case control study was aimed to analyze single nucleotide polymorphisms (SNPs) in leptin gene in elevating the risk of VTE in females contraceptive users. It has been hypothesized that the SNP in LEPR (Gln223Arg) gene was associated with susceptibility to increase the risk for developing VTE in female contraceptive users. Genotyping of SNP were done by restriction fragment length polymorphism assay and the data was analyzed using various statistical tools. It has been observed that the SNP in LEPR rs1137101 G>A is significantly associated with VTE in accordance with both genetic models (DM; OR=9.270, CI (4.977-17.264), p=0.0000, RM; OR=19.70, CI (2.658-146.11), p=0.000). The frequency of risk allele G (52%) also shows significant association with the disease phenotype (OR=6.63, CI (3.910-0.1345), p=0.0001). Other risk factors including body mass index (BMI), body weight and ABO blood types were also analyzed for their association with VTE. Pearson correlation and regression analysis was done to further analyze the data. The correlation between body weight and duration of contraceptives usage was found to be 0.152 (p=0.020). Regression analysis was done to find the relationship between body weight (dependent variable) and duration of contraceptive usage (independent variable). The results show significant relationship (p=0.000) where 1 fold increase in duration (time in years) results in 0.4 folds increase in body weight (Kg) of females using contraceptives. Here, the current study concluded that CHCs could be one of the critical risk factor which increase the chances of developing VTE in genetically susceptible females in Pakistani population.

Molecular Effects of Genetic Variation Posters - Thursday

PB2684. Scalable mQTL analysis of biochemical pathways with the UK Biobank using REVEAL:Biobank

Authors:

Z. Pitluk, M. Peterson, U. Mudgal, A. Poliakov, S. Sarangi; Paradigm4, Waltham, MA

Abstract Body:

Metabolite levels in different bodily fluids such as blood, serum, and urine offer an invaluable window into understanding the physiological state of the body. As precursors, intermediates, and end products in biochemical pathways, linking metabolite analysis to underlying genetic variations can be crucial in developing systems-wide models of complex diseases and disorders by highlighting variants with metabolic consequences. Metabolite quantitative trait loci (mQTL) analysis can be used to identify variants across multiple genes that are potentially involved in regulation of metabolite concentrations. Due to the multifactorial dependence between metabolites and genetic variants in a pathway, mQTL studies are challenging and involve testing millions of associations. Metabolomics-focused resources such as the KEGG PATHWAY and Human Metabolome Databases, and biobank-scale multiomics datasets such as the UK Biobank are valuable tools to build and test hypotheses and identify biomarkers of interest. However, performing such studies at scale present major challenges such as computational cost and efficiency, and ease of use. REVEAL: Biobank presents a scalable, cost-effective, and user-friendly solution to conduct multiomic data analyses. In this study, we created an example workflow in REVEAL: Biobank querying the KEGG and the Human Metabolome Databases for genes involved in the Bile Biosynthesis pathway. Following that we did a GWAS analysis between variants of the identified genes from the 200K Whole Exome Sequence data and the NMR Metabolomics dataset (data category ID - 220; 169 phenotypes plus metabolite ratios) from the UK Biobank (Application ID: 51518) using PLINK and SAIGE algorithms. We also performed a linkage disequilibrium (LD) analysis combined with a regression burden test between pairs of the gene variants and the metabolomic phenotypes. Finally, we explored clinical data from the UK Biobank for cohorts differentiated by the genotype and metabolite values from the associations identified in the GWAS and LD analyses. Our goals in this study were to create a workflow for conducting cost-effective and efficient mQTL analysis of biochemical networks at scale using external databases and biobank-scale datasets to improve understanding of the complex associations between genetic variations and metabolomics.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2685. Seamless integrative pipeline for QTL datasets enhance the discovery of putative causal variants for Alzheimer's Disease.

Authors:

J. Cifello, P. P. Kuksa, L-S. Wang, Y. Leung; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Colocalization of significant loci between genome-wide association studies (GWAS) and quantitative trait loci (QTLs) have demonstrated tissue-specific associations. Such signals have been useful in understanding the mechanisms of Alzheimer's Disease (AD), utilizing recent data from multiple authors and brain regions. However, transforming these datasets into analysis-ready files is not straight-forward due to differences in files formats, variant nomenclature, how alleles are represented, and variations in provided statistics. For example, some datasets are published with effect sizes reported with respect to the minor alleles while other datasets use the alternative alleles. Here, to facilitate the use of QTL datasets in understanding the regulatory potential of non-coding variants, we developed a scalable, automated pipeline (bitbucket.org/wanlab-upenn/filer-xqtl-pipeline) for expedited harmonization and ingestion of QTL data into the FILER repository (Kuksa et al, 2022), which contains over 60,000 functional genomic and annotation datasets available for query. Using this pipeline, we have integrated over 25 billion QTL records representing 84 different tissues and cell types. We illustrated use of these harmonized QTL data by analyzing 11,449,726 variants (241 linkage-disequilibrium blocks) from an AD GWAS (Kunkle et al. 2019). These were compared to AMP-AD QTL data (Sieberts et al, 2020) from three brain regions (cerebellum (CER), dorsolateral prefrontal cortex (DLPFC), and temporal cortex (TCX)). We used our scalable colocalization pipeline (based on coloc (Giambartolomei et al, 2014)) and found a total of 173 significantly colocalized variants, associated with 24 expression-QTL target genes and 116 GWAS loci. Then, after restricting our search to significant loci by excluding HLA and APOE regions, four GWAS loci remained. We queried these loci using FILER and found they have all been previously reported as significant loci in GTEx splicing-QTLs (Zhang et al, 2020). Three chr8 variants (rs6987305, rs2322599, and rs17057043), colocalizing in CER and TCX, have been shown to be positively correlated with PTK2B in whole blood, but negatively correlated in adipose tissue. These variants, all within 12kb on chr8, may add additional insights into how PTK2B plays a role in AD. Using a scalable pipeline for harmonization and analysis of QTL and GWAS data can help to systematically profile GWAS results for functional effects and identify their tissue and cell type context.

Molecular Effects of Genetic Variation Posters - Thursday

PB2686. SEPT-GD: a decision tree to prioritise potential RNA splice variants in cardiomyopathy genes for functional splicing assays in diagnostics.

Authors:

M. Alimohamed¹, L. Boven², K. van Dijk², Y. Vos³, Y. Hoedemaekers⁴, P. van der Zwaag², R. Sijmons⁵, J. Jongbloed⁶, B. Raddatz², H. Westers⁷; ¹Muhimbili Univ. of Hlth.and Allied Sci., Dar es salaam, Tanzania, United Republic of, ²Univ. Med. Ctr. Groningen, Groningen, Netherlands, ³Univ Med Ctr Groningen, Groningen, Netherlands, ⁴Erasmus Med Ctr, Rotterdam, Netherlands, ⁵Univ. Med. Ctr. Groningen, Groningen, Groningen, Netherlands, ⁶Univ Med Ctr, Groningen, Netherlands, ⁷Univ Med Ctr. Groningen, Groningen, Netherlands

Abstract Body:

Background: Splice prediction algorithms currently used in routine DNA diagnostics have limited sensitivity and specificity, therefore many potential splice variants are classified as variants of uncertain significance (VUSs). However, functional assessment of VUSs to test splicing is labour-intensive and time-consuming. We developed a decision tree to prioritise potential splice variants for functional studies and functionally verified the outcome. **Materials and methods:** We built the decision tree, SEPT-GD, by setting thresholds for the splice prediction programs implemented in Alamut. A set of 343 variants with known effects on splicing was used as control for sensitivity and specificity. We tested SEPT-GD using variants from a Dutch cardiomyopathy cohort of 2002 patients that were previously classified as VUS and predicted to have a splice effect according to diagnostic rules. We then selected 12 VUSs ranked by SEPT-GD to functionally verify the predicted effect on splicing using a minigene assay: 10 variants predicted to have a strong effect and 2 with a weak effect. RT-PCR was performed for nine variants. Variant classification was re-evaluated based on the functional test outcome. **Results:** Compared to similar individually tested algorithms, SEPT-GD shows higher sensitivity (91%) and comparable specificity (88%) for both consensus and non-consensus splice-site variants. Using clinical diagnostic criteria, 1295 unique variants in our cardiomyopathy cohort had originally been classified as VUSs, with 57 predicted by Alamut to have an effect on splicing. Using SEPT-GD, we prioritised 31 variants in 40 patients. In the minigene assay, 12 variants showed results concordant with SEPT-GD predictions. RT-PCR confirmed the minigene results for two variants, *TMEM43* c.1000+5G>T and *TTN* c.25922-6T>G. Based on these outcomes, the *SGCD* c.4-1G>A and *CSRP3* c.282-5_285del variants were reclassified as likely pathogenic. **Conclusion:** SEPT-GD outperforms the tools commonly used for RNA splicing prediction and improves prioritisation of variants in cardiomyopathy genes for functional splicing analysis in a diagnostic setting.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2687. Sex differences in gene expression affect host response to influenza A infection

Authors:

A. Jones, B. Schott, L. Wang, D. Ko; Duke Univ., Durham, NC

Abstract Body:

Every year, approximately 30 million people are infected with influenza A virus (IAV) in the United States alone. However, there are diverse outcomes in IAV susceptibility and severity, with some cases presenting as entirely asymptomatic, and others resulting in death. An integral variable in IAV infection is host biological sex. Female individuals of reproductive age have greater morbidity and mortality from IAV infection, while male individuals, particularly children and those >65, are more susceptible to infection. However, despite well-established evidence for sex differences in IAV infection, common genetic variants that contribute to this difference have not been identified. We hypothesize that human genetic variation drives sex differences in gene expression that impact IAV infection susceptibility and severity. To elucidate mechanisms that underly sex differences in IAV infection, we utilized a single-cell RNA-seq model of IAV infection in lymphoblastoid cell lines (48 females, 48 males) to **1**) investigate genes with sex-biased expression associated with IAV phenotypes and **2**) determine human genetic variants that affect sex-biased gene expression and IAV infection. Transcriptomic data revealed widespread sex biases in gene expression during IAV infection with 192 genes showing significant differential expression by sex (FDR < 0.1). In addition, we identified genomic variants that significantly affect gene expression in a sex-dependent manner during infection (n = 40, FDR < 0.05) and are associated with IAV infection phenotypes in human cells. Our study, the first to identify common genetic variants affecting sex-biased response to IAV infection, confirms that biological sex is a critical variable in IAV pathogenesis while identifying targets for future development of diagnostic and therapeutic approaches.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2688. Short chain fatty acid abnormalities: Genotype-Phenotype correlation in thirteen Puerto Rican patients.

Authors:

J. R. Pascual¹, O. Quincoces¹, C. Hernandez², S. Carlo³, A. Cornier⁴; ¹Ponce Hlth.Sci. Univ., Ponce, PR, ²Natl. Children's Hosp. Pediatrics, Washington, DC, DC, ³Ponce Hlth.Sci. Univ., Cabo Rojo, PR, ⁴San Jorge Children s Hosp./ Ponce Hlth.Sci. Univ., San Juan, PR

Abstract Body:

We are presenting thirteen patients of Puerto Rican Descent that were diagnosed with mitochondrial fatty acid oxidation disorder (mFAO). Two patients are brother and sister while the rest are unrelated. The patients were referred to the genetics clinic at different ages ranging from seven months to nine years. The reasons for the referral included the broad phenotype of hypotonia, behavioral problems, speech, and global developmental delay, and sensory disorder. Microarray studies were unremarkable. Metabolic laboratory workup showed one or various elevated metabolites. These include lactate, ethylmalonic acid (EMA), methylmalonic acid (MA), dicarboxylic acids (DA), butyryl glycine and C3-C5 acylcarnitine species by-products. Due to these results, a Fatty Acid Oxidation Defects gene panel was ordered. Results showed all patients had genetic within the fatty acid oxidation disorders genes. The patients were identified to be heterozygous, compound heterozygous or homozygous for the genetic changes. The most prevalent genetic change was c. 625G>A or the c. 511 C >7 variant in the ACADS gene. These variants have been identified to cause enzyme susceptibility in previous in vivo and cell culture studies. Other identified variants in genes associated with mFAO include ACADVL, ACADM, *SLC22A5* and *HADHB*. The laboratory and clinical findings, and identified variants are suggestive of short chain fatty acid abnormalities as the genetic etiology of these patients' phenotype. We believe there is a strong genotype-phenotype correlation in this patient cohort.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2689*. Single-cell caQTL reveals cell-state-dependent regulation in immune cells in rheumatoid arthritis.

Authors:

Z. Mu¹, W. Wei², P. Zhu², Y. Li¹; ¹Univ. of Chicago, Chicago, IL, ²Dept. of Clinical Immunology, Xijing Hosp., Xi'an, China

Abstract Body:

For many complex diseases such as rheumatoid arthritis (RA), genome-wide association studies (GWAS) have identified hundreds of independent susceptibility loci (**Ishigaki et al., medRxiv, 2022**). However, mechanistic understanding of these loci remains fundamentally challenging, as the majority of associated variants map to non-coding sequences. Previous studies have shown that only an average of 40% of GWAS loci colocalize with molecular quantitative trait loci (molQTL) in healthy donors (**Mu et al., Genom. Biol., 2021**), suggesting that the remaining 60% of susceptibility loci may affect disease- and context-specific gene regulation. Here, we applied scATAC-seq to 41,813 peripheral blood mononuclear cells (PBMC) from 5 RA patients and 13 healthy donors, and identified 268,112 accessible peaks. We applied topic modeling based on Poisson Non-negative Matrix Factorization (NMF) to peak count data (**Dey et al., PLoS Genet., 2017**), revealing a continuum of cell states in multiple immune cell lineages. Using these continuous cell states, we identified 62,413 state-specific peaks, 27,207 (43.6%) of which were not captured in cluster-level differential analysis. We developed a framework to integrate fine-mapped disease GWAS with the topic model to nominate causal SNPs and genes in a cell-state-specific fashion. In general, we show that topic modeling is a powerful alternative to cluster analysis for scATAC-seq data. Finally, we performed chromatin QTL (caQTL) mapping in our dataset followed by meta-analysis with two additional published datasets. In total, we discovered ~8,000 caQTLs at 10% FDR across various cell types. By modeling continuous cell states from the topic model in caQTL mapping, we identified a continuous relationship between caQTL effect sizes and cell states, further refining the definition of cell-type-specific caQTLs. We estimated that about 5%-20% of peaks with significant caQTL show evidence of genotype-by-cell-state interaction. Finally, we highlighted several examples where caQTLs could help explain previously uncolocalized GWAS loci. To summarize, our study provides an atlas of chromatin accessibility within RA and healthy PBMCs, establishes a framework for studying dynamic gene regulation in disease contexts, and offers novel insight in the genetic architecture of RA.

Molecular Effects of Genetic Variation Posters - Thursday

PB2690. Single-nucleus transcriptome analysis of 424 aging brains identifies cell type and subtype specific *cis*-eQTL and their relation to neurodegenerative disease susceptibility.

Authors:

M. Fujita¹, Z. Gao¹, L. Zeng¹, C. McCabe², C. White¹, B. Ng³, G. Green⁴, B. Vardarajan¹, H-U. Klein¹, G. Wang¹, N. Habib⁴, J. A. Schneider⁵, T. Young-Pearse⁶, S. Mostafavi⁷, D. A. Bennett⁵, V. Menon¹, P. L. De Jager¹; ¹Columbia Univ., New York, NY, ²The Broad Inst., Cambridge, MA, ³Univ. of British Columbia, Vancouver, BC, Canada, ⁴The Hebrew Univ. of Jerusalem, Jerusalem, Israel, ⁵Rush Univ., Chicago, IL, ⁶Harvard Med. Sch., Boston, MA, ⁷Univ. of Washington, Seattle, WA

Abstract Body:

Genome-wide association studies have identified genetic variation associated with complex traits of human brains, such as susceptibility to Alzheimer's disease. However, the functional consequences of these variants remain largely unclear, particularly at the level of cell type specificity. Recent advances in single-cell technology opened new opportunities to explore genotype-expression relationship at cellular resolution. In this study, we performed single-nucleus RNAseq of dorsolateral prefrontal cortex material from 424 individuals of the ROS/MAP cohorts of cognitive aging. More than 1.7 million cells were analyzed and classified into seven cell types and further subclassified into 81 cell subtypes. Cell-level expression data were aggregated into individual-level "pseudobulk" expression data. This pseudobulk expression was used to map cell type-specific *cis*-eQTL by correlating it with genotype of matched individuals. In total, 10,004 unique eGenes were detected at the cell type level, and 8,138 unique eGenes were detected at the cell subtype level. Statistical power to detect eGenes was strongly correlated with the frequency of cell types. Subtype analysis boosted eQTL discovery by identifying eGenes that were not detected at the cell type analysis. Most eGenes were expressed in more than one cell types but had eQTL association only in one cell type, demonstrating context specificity of brain eQTL. As an example, we detected a SNP that is a microglia-specific eQTL of *APOE* expression level. The SNPs had significant association with risk of Alzheimer's disease and cerebral amyloid angiopathy, independently from *APOE* ϵ 4 haplotype. To explore association with disease susceptibility, we performed colocalization analysis between GWAS risk SNPs and cell type-specific eQTL. Risk SNPs of Alzheimer's disease, Parkinson's disease, and schizophrenia colocalized with eQTL, revealing potential cell types responsible for the altered susceptibility. Similar analyses were performed for brain-related traits, including educational attainment and regional brain volumes based on MRI. We further utilized our data to perform cell type-specific transcriptome wide association study, which revealed additional novel candidates of causal genes of neurodegenerative diseases. Overall, this large dataset of single nucleus data is a valuable resource to explore cell type-specificity of the genetic architecture of gene expression in the aging human brains and to prioritize the gene(s) and cell subtype(s) involved in the propagation of functional consequences of susceptibility variants for neuropsychiatric diseases.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2691*. SMN modulates adverse effects associated with GEMIN5-mediated neurodevelopmental syndrome

Authors:

U. Pandey¹, S. Kour¹, A. Chimata², E. Anderson¹, C. Ward¹, O. Chauhan¹, C. O'Brian¹, D. Rajasundaram¹, D. Rajan¹, A. Singh², T. Fortuna¹; ¹Univ. of Pittsburgh Med. Ctr., Pittsburgh, PA, ²Univ. of Dayton, Dayton, OH

Abstract Body:

GEMIN5, an RNA-binding protein, plays an important role in the core assembly of small nuclear Ribonucleoproteins (snRNPs), the building blocks of spliceosome formation. We identified novel autosomal recessive variants among 35 patients presenting with developmental delay, ataxia, motor dysfunction, and cerebellar atrophy using whole-exome sequencing approach. GEMIN5 variants perturb the subcellular distribution, stability, and expression of GEMIN5 protein and its interacting partners in iPSC neurons, indicating a potential loss-of-function mechanism. Interestingly, GEMIN5 mutations disrupt snRNP complex assembly formation in patient iPSC neurons. Molecular determinants of GEMIN5-mediated disease are yet not known. While doing a genetic screen, we identified SMN, a component of the snRNP complex, as a genetic modifier of GEMIN5 phenotypes in vivo. We found that GEMIN5 levels are upregulated in response to genetic expression of SMN in iPSC neurons and in vivo. Interestingly, treatment of the SMN2 therapeutic antisense oligonucleotide, ISS-N1 (nusinersen), significantly upregulated GEMIN5 protein levels in iPSC neurons, suggesting a functional interaction between both proteins. SMN expression restored the defective snRNP components of the SMN complex in GEMIN5 patient neurons. In addition, we show that SMN expression increases the number of nuclear bodies which are otherwise lost in GEMIN5 patient neurons. Our findings indicate that SMN is a novel regulator of GEMIN5 expression and neuropathologies.

Molecular Effects of Genetic Variation Posters - Thursday

PB2692. Spectrum of pathogenic ATM variants in Indian ataxia telangiectasia patients.

Authors:

R. Shukla¹, A. Sankar¹, R. Gatti²; ¹All India Inst. of Med. Sci., New Delhi, India, ²UCLA Sch. of Med., Los Angeles, CA

Abstract Body:

Pathogenic variants in the *ATM* gene are responsible for the autosomal recessive neurodegenerative disorder, ataxia telangiectasia (A-T). Mutations in various ethnic groups have been described and there seem to be no hotspots. In the present study, A-T patients from 41 Indian families were assessed for their clinical phenotypes and *ATM* haplotype analysis, and screened for *ATM* mutations. Thirty seven distinct haplotypes were observed in 41 unrelated families comprising of 21 homozygous and 16 heterozygous haplotypes. Pathogenic *ATM* variants were identified in all 41 families, Eleven of these were aberrant splicing, 5 truncations, 13 frameshifts, 2 missense, and one large genomic deletion spanning exon 17-63 and the other a deletion of 193nt of exon 11. Recurring haplotype and associated variant c.5631_5635delinsA was observed in 7 families of North Indian origin, suggestive of a possible founder effect in Indian A-T patients. Of the 42 *ATM* variants identified in the present study, 29 were novel unreported variants, thereby suggesting that the profile of mutations in the India subcontinent is unique. The present data adds to the multitude of unique *ATM* mutations identified worldwide.

Molecular Effects of Genetic Variation Posters - Thursday

PB2693. Structural and functional characterization of missense variants in *NPR2* augments genetic associations, phenotypic prediction, and mechanistic understanding of skeletal development.

Authors:

S. Covarrubias, C. Bauer, D. Shanghavi, Y-S. Tseng, A. Luu, K. Estrada, S. Froelich; Biomarin Pharmaceuticals, San Rafael, CA

Abstract Body:

Associations between rare coding variants and the human phenome can provide a rough draft of the biological mechanisms underlying disease, but these efforts are currently limited by our inability to classify the effects of novel missense variants. We have conducted cell-based functional screens on hundreds of missense variants in the human Natriuretic peptide receptor B (*NPR2*) derived from patients and the general population. Our data not only allows us to classify dozens of variants as loss of function as well as several as gain of function, but we can quantitatively score variants with intermediate activity levels. These variant scores enhance prediction of adult height when combined with polygenic scores, augment statistical power to detect phenotypic associations, and provide insights into the biological mechanisms of C-Type Natriuretic Peptide (CNP) signaling through *NPR2* in skeletal development. *NPR2* is a key regulator of skeletal growth. Biallelic loss of function (LoF) mutations in *NPR2* can cause severe disproportionate dwarfism while some rare missense variants have been linked to overgrowth and extremely tall stature. Our results demonstrate that variants implicated in overgrowth are often substitutions of basic residues near the kinase homology domain that increase cellular cGMP production. Non-neutral variants in the guanylyl cyclase domain are exclusively linked to reduced signaling while we observe a range of effects caused by mutations in the ligand binding domain. We also show that incorporation of the functional activity score for each variant predicts human phenotype. Across all carriers of a single LoF allele, we can explain 34% of the residual variance in height after adjusting for their polygenic scores. Using variant activity scores in association tests improves power in association tests for height and allows us to detect other disease associations involving cardiovascular and mental health traits. Finally, we compare our results to computationally predicted variant effects as well as predicted protein folding changes induced by these substitutions providing a path toward missense variant scoring for other genes where cell-based screening is less tractable. The classification of missense variants in *NPR2* more than doubles the number of LoF variants for this gene, indicating an overall frequency of more than 1 in 1500. A CNP analog, vosoritide, has recently been demonstrated to enhance growth velocity in children with achondroplasia. Our data suggests that some of these individuals may benefit from this therapeutic targeted to the causal mutations of their growth deficits.

Molecular Effects of Genetic Variation Posters - Thursday

PB2694. Systematic analysis of the impact of short tandem repeats on gene expression

Authors:

S. Benton, L. Zhang, M. Maksimov, S. Shleizer-Burko, Q. Gong, C. Wang, M. Lamkin, M. Gymrek, A. Goren; Univ. of California San Diego, La Jolla, CA

Abstract Body:

Short tandem repeats (STRs) represent an intriguing type of genetic variation given the high level of complexity they can provide as well as their high rate of mutability. The repeat units of STRs vary in properties such as repeat unit length (ranging from 1-6 base pairs), composition (e.g., (A)_n, (AC)_n or (CCG)_n), and sequence variations within the STR (i.e., a mix of perfect and non-perfect repeat units). We previously identified ~28,000 STRs that show an association between the number of repeat units and expression of nearby genes and proposed that they can impact gene expression via multiple mechanisms. Here, to systematically evaluate the involvement of STRs in proximal regulation of gene expression, we established a massively parallel reporter assay (MPRA) built on existing MPRA frameworks but optimized to handle complex and error-prone repetitive sequences. For each STR of interest, we test 230bp fragments including the STR and surrounding genomic context, and test for three separate repeat lengths. In total, our array includes 100,000 different STR sequences spanning approximately 33,000 unique STRs found in promoter regions of human genes. Our system employs random barcodes, enabling up to dozens of measurements per sequence in a single experiment.

In our initial evaluation, we generated libraries capturing >44% of the sequences in our pool. Notably, longer STR sequences and those with high GC content are more likely to drop out of our pool, and thus, incomplete library coverage is expected. Following transfection of these libraries to the HEK293 human cell line, we observed a strong correlation of normalized expression levels between replicates (R=0.60). Further, an STR nearby the POMC gene, a previously identified STR associated with transcription regulation, was among the STRs in our MPRA that showed a positive impact on the expression of the reporter gene, and a linear trend with increased repeat number.

Together, our study provides a system for high-throughput interrogation of the involvement of STRs in modulating gene expression by directly perturbing repeat length and sequence. This system is likely to provide new insights into causal mechanisms by which STRs influence gene regulation. It will also serve as valuable resource for evaluating STR associations with expression identified in large population-based cohorts. While our initial experiment has focused on promoter STRs, future extensions of our approach to study STRs in other genomic regions (e.g., intergenic or intragenic) that potentially may be impacting transcription by other mechanisms.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2695. Targeted-NGS across the β -globin gene cluster identifies multiple linkage disequilibrium patterns in north Indian non-transfusion dependent β -thalassemia patients

Authors:

N. Singh, P. Sharma, M. Jamwal, R. Sharma, R. Das, A. Khadwal, A. Trehan; Post Graduate Inst. of Med. Ed. & Res., Chandigarh, India

Abstract Body:

Introduction: Non-transfusion-dependent β -thalassemia (NTD β T) is an enigmatic disorder influenced by *HBB* and α -globin genotypes and fetal hemoglobin (HbF) levels. Haplotyping of the β -globin gene cluster's (β -LCR's) Locus Control Region is often attempted in order to discover single nucleotide polymorphisms (SNPs) with high linkage disequilibrium (LD) that may modulate the HbF-to-HbA switch and thereby, the severity of thalassemia. Tag-SNPs are representative polymorphisms that "tag" a particular haplotype. They are useful to test associations of a marker locus with qualitative/quantitative traits, without having to genotype all of its SNPs. **Objective:** To identify different haplotypes and their associated Tag-SNPs across the β -globin gene cluster including the β -LCR in genetically similar north Indian NTD β T patients. **Methods:** NTD β T patients (n=51) harboring *HBB*:c.92+5G>C and β^0/β^+ variant underwent targeted NGS of the β -globin gene cluster using Agilent's HaloPlexHS™ target-enrichment kit, with variant calling using Agilent's SureCall™ pipeline. Haplotypes were constructed using HaploView™ v4.2 (Broad Institute). The patients were divided into mild (n=28) and moderate (n=23) severity using the Mahidol scoring system. **Results:** 135 SNPs were delineated spanning an over 95 kbp region. Haplotypes were constructed using common SNPs (MAF >0.05, Hardy-Weinberg p-value >0.01). Among mild NTD β T, 94 qualifying polymorphisms meeting the above criteria were segregated by HaploView™ into 8 LD-blocks. HaploView's Tagger function used with LD-threshold $r^2 > 0.8$ and with aggressive tagging with 2- and 3-marker haplotype and an LD cut-off 3.0 yielded 30 Tag-SNPs. The same approach applied among moderate NTD β T generated five LD blocks containing 62 SNPs and 27 tag-SNPs. In the mild NTD β T group LD block 2 was the largest (an over 22 kbp region with 14 tag SNPs) while in the moderate group LD block 1 represented a 21 kbp region with 13 tag SNPs. These two largest blocks covered the ϵ , $\zeta\gamma$, $A\gamma$, δ , and β globin genes. This region had 5 tag SNPs that were common to both the severity groups. The *HBG1-HBD* ($A\gamma$ - δ) intergenic region contains the non-coding gene *BGLT3*. This region contained two tag-SNPs: rs7483789 in mild and rs7480197 in moderate NTD β T. **Conclusion:** Haploview™ analysis using targeted-NGS of the β -globin cluster identifies multiple LD patterns in genetically similar north Indian NTD β T patients with varying degree of severity. This suggests a diversified genomic structure in this group and should aid future comparisons of NTD β T patients with divergent phenotypes across various regions.

Molecular Effects of Genetic Variation Posters - Thursday

PB2696. Testing the association of Grantham score with functional impact for clinical variant interpretation

Authors:

K. Owens, G. A. Maston; Quest Diagnostics, Seacaucus, NJ

Abstract Body:

Background: The Grantham score (GS) is a calculation that compares the different properties of any 2 amino acids to predict the impact of a substitution on protein function. Previous research has explored the use of GS in the clinical interpretation of missense variants; however, no systematic testing has been done to confirm the accuracy of GS for predicting pathogenicity for all types of amino acid changes. The present study takes advantage of previously published high-throughput *BRCA1* functional data (Findlay GM, Daza RM, Martin B, et al. *Nature*. 2018;562[7726]:217-222) to address how well GS correlates with loss of function (LOF) for different types of amino acid changes. We hypothesized that greater GS would be associated with a higher percentage of loss of function results (% LOF) for any given amino acid substitution. **Methods:** Data from Findlay, et al regarding the functional impact of *BRCA1* variants in the RING and BRCT domains were stratified by reference amino acid and tallied to determine how many times each specific amino acid substitution resulted in a loss of protein function based on the functional assay performed by Findlay, et al. These results were converted to a percentage (% LOF) and plotted against the GS for each amino acid substitution on a graph with a best-of-fit line. These values were also compared to Align GVGD scores for each given amino acid position. **Results:** A polynomial equation provided the best-of-fit line for each reference amino acid chart, although most of the reference amino acids did not reflect a strong correlation between % LOF and GS. Five amino acids exhibited a correlation (R^2) of greater than 0.70: cysteine (Cys), isoleucine (Ile), methionine (Met), tyrosine (Tyr), and valine (Val). Cys, Ile, and Val substitutions displayed a positive correlation between % LOF and GS, as anticipated. Met substitutions appeared to have a threshold for GS, where the % LOF was stable until after a score of 80, at which point the % LOF increased dramatically. Tyr substitutions displayed a parabolic inverse correlation between % LOF and GS. **Limitations:** This analysis only examines codons in known functional domains of *BRCA1*; additional testing is needed to determine if these correlations hold true for amino acid changes in different protein domains or in different genes. **Conclusions:** The results of this study suggest that GS for amino acid substitutions do not always correlate with functional evidence and should not be used in isolation to interpret missense variants. It may be necessary to develop gene- or domain-specific rules on the proper use of GS in clinical variant interpretation.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2697. The contribution of mitochondrial DNA variation to disease risk in a diverse cohort

Authors:

A. Zaidi, C. Morse, A. Verma, M. Ritchie, Regeneron Genetics Center, Penn Medicine Biobank, I. Mathieson; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Little is known about how variation in mitochondrial DNA or mtDNA copy number (mtCN) contributes to phenotypic variation. There is extensive interaction between proteins encoded by the mitochondrial and nuclear genomes, but the degree to which these interactions contribute to disease risk is unknown. To address these questions, we analyzed data from the Penn Medicine Biobank, a large (N = 43,000) diverse cohort with genetic and electronic health record data. We estimated mtCN from exome sequence data and replicated associations between mtCN and predictors such as sex, age, and blood cell counts. We also carried out a genome-wide association study for mtCN, replicating previously discovered genetic variants with consistent effects across individuals of European and African ancestry. Mitochondrial copy number was associated with 14 out of 1,131 health-related phenotypes at a false discovery rate of 0.05. These included diabetes, hypertension, epilepsy, liver and pulmonary disease, prostate cancer, and chronic kidney disease (CKD). The association with CKD was supported by consistent associations with related biomarkers such as serum creatinine and serum albumin levels and was significant in both European ($P = 1.2 \times 10^{-5}$) and African ancestry ($P = 5.7 \times 10^{-5}$) individuals.

Disease associations for specific mitochondrial haplogroups have been reported but are hard to confirm because geographic variation in haplogroups is confounded with environmental variation and nuclear ancestry. To overcome this, we analyzed data from individuals with mixed African and European ancestry (N = 9,037) where we can correct for the effects of nuclear ancestry. We show that mitochondrial haplogroups are not significantly associated with any phenotype, suggesting that haplogroup-specific variants do not contribute substantially to disease risk variation across ancestries. In admixed individuals, the interaction between mitochondrial and nuclear ancestry is also not significantly associated with mtCN or any other phenotype, despite sufficient power to detect moderate to large effects. This suggests that mito-nuclear incompatibility does not contribute substantially to health-related variation. Our study confirms the utility of mtCN as a biomarker of disease and shows that mitochondrial variation has limited or no contribution to ancestry-related differences in disease risk.

Molecular Effects of Genetic Variation Posters - Thursday

PB2698. The contribution of short tandem repeats to splicing variation in humans

Authors:

Y. Li, J. Margoliash, Y. Dong, Z. Zhang, A. Goren, M. Gymrek; Univ. of California, San Diego, San Diego, CA

Abstract Body:

Alternative splicing (AS) is a fundamental step that plays a critical role in the regulation of gene expression and protein diversity in eukaryotes. Emerging studies have identified thousands of genetic variants associated with variation in AS of nearby genes. Most of those studies have focused on single nucleotide polymorphisms (SNPs). Short tandem repeats (STRs), consisting of repeated sequences of 1-6bp, have been implicated in multiple splicing-related processes, including forming binding sites for splicing factors such as hnRNP-L and modulating RNA secondary structure. However, their contribution to splicing variation in humans has not been explored on a genome-wide scale.

Here, we use whole-genome sequencing and expression data from 41 tissues in the Genotype-Tissue Expression Project to systematically evaluate the contribution of STRs to AS in humans. We applied HipSTR to perform multi-sample STR genotyping of 1.6 million STRs across 652 individuals. We then tested for associations between the mean repeat length of STRs and the percent spliced in (PSI) of exons within 10kb of those STRs while controlling for population structure and PEER factors.

In total, we identified 7,467 unique associations (spliceSTRs). Cross-tissue analysis suggests the 48.5% of these spliceSTRs are shared across two or more tissues. We applied the FINEMAP algorithm to perform fine-mapping of spliceSTR associations against nearby SNPs. We identified 1,921 unique spliceSTRs with at least 50% posterior inclusion probability and recapitulated multiple previously known spliceSTRs, including the myotonic dystrophies (*DMPK*), suggesting those spliceSTRs are strong candidate causal variants. Overall, our study reveals the regulatory associations of STRs with AS, identifies a set of high confidence spliceSTRs candidates and provides a valuable resource for future studies of complex traits.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2699. The effect of rare protein-coding sequence variants on the human plasma proteome

Authors:

B. Prins, O. Burren, Q. Wang, R. Dhindsa, K. Carss, K. Smith, S. Petrovski; Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom

Abstract Body:

Combining human genomics with proteomics can facilitate therapeutic target identification and validation and shed light on pathogenic processes and associated biomarkers. We integrated the plasma abundances of 1,472 proteins measured on the Olink™ platform with exome sequences from 37,124K UK Biobank participants to assess the influence of rare coding sequence variation on the human plasma proteome. We identified 1,248 significant ($p < 2e-9$) gene-protein associations through a gene-level collapsing analysis of rare variants that included 706 *cis* and 539 *trans* signals. A complementary rare variant (MAF < 0.1%) exome-wide association analysis identified 1,192 non-synonymous variants associated with the plasma levels of 850 proteins, of which 723 were in *cis* and 469 were in *trans*. Phenome-wide analysis identified 24 genes harbouring rare genetic variants that were associated with both plasma protein abundance and a binary clinical endpoint. Variants within the gene *TNFRSF13B*, encoding the transmembrane activator and CAML interactor (TACI), are associated with low immunoglobulin levels and combined variable immune deficiency (CVID)¹. In *cis*, known CVID pathogenic missense variants in *TNFRSF13B* were associated with decreased TACI plasma protein abundance (e.g., 17-16940415-G-T, $P = 2.4 \times 10^{-46}$, $\beta = -0.70$ [95%CI -0.80 — -0.60]). In *trans*, we observed significant associations with increased plasma abundance for 13 proteins including BAFF, the B cell survival factor that signals through TACI (17-16940415-G-T, $P = 1.8 \times 10^{-17}$, $\beta = 0.42$ [95%CI 0.32 — 0.52]). Such associations highlight the complex role of TACI in B-cell regulation and antibody response and suggest additional proteins that may be involved in CVID pathogenesis. Our study shows that rare, non-synonymous variants can modulate the plasma proteome and can be used to explore disease mechanisms and therefore inform therapeutic development. Future work will integrate these results with other clinical and molecular phenotypes available for UK Biobank participants and will be made available through our online PheWAS portal (<https://azpewas.com/>).

¹ Pan-Hammarström, Q. et al. (2007) Nat. Genet., 39, 429-430

Molecular Effects of Genetic Variation Posters - Thursday

PB2700. The genetic architectures of gene expression in individuals of African and European ancestry: results and consequences of eQTL studies

Authors:

K. Fletez-Brant¹, G. Atla², E. Bullis³, T. Cavazos¹, A. Cortes⁴, P. Gormley⁵, B. Hicks³, L. Howe⁶, K. Kukar³, Y. Liang³, S. Micheletti³, R. Mishra⁷, M. Moreno³, P. Nandakumar³, J. O'Connell³, A. Petrakovitz³, S. Pitts¹, R. Sauteraud¹, D. Seaton⁶, K. Song⁷, A. Sugathan¹, C. Wong³, the 23andMe Computational Biology Team, the 23andMe Research Team, V. Vacic¹; ¹23andMe Therapeutics, South San Francisco, CA, ²GlaxoSmithKline, Stevenage, United Kingdom, ³23andMe, Sunnyvale, CA, ⁴GSK, Stevenage, United Kingdom, ⁵GSK, Cambridge, MA, ⁶GlaxoSmithKline, Stevenage, United Kingdom, ⁷GlaxoSmithKline, Upper Providence, PA

Abstract Body:

Underrepresentation of individuals of African ancestry in biomedical studies is an established fact in need of remediation. This is particularly true in the case of genetics and genomics, where nearly all studies are performed in individuals of European descent. To help address this inequity, we have conducted whole genome sequencing (WGS) of DNA from saliva and RNA-sequencing (RNA-seq) of venous blood samples of 1,012 23andMe, Inc consented research participants of African ancestry as part of the Black Representation in Genomic Research (BRGR) study. Additionally, we did RNA-seq of lymphoblastoid cell lines (LCL) from 665 individuals belonging to the African Ancestry superpopulation (AFR) from the 1000 Genomes Project. All biological sample collection was performed in accordance with the terms of informed consents and under an IRB approved protocol. These data were used to conduct expression quantitative trait loci (eQTL) studies in the BRGR cohort (N = 737), and together with published WGS, in the AFR LCL cohort (N = 659). We find 26,877 eQTLs affecting expression of protein coding genes in the BRGR study, and 20,371 in the AFR LCL study. We compare these numbers to previously published eQTL studies in European individuals -- the venous blood study from the Parkinson Progression Marker Initiative (PPMI), and the European cohort of GEUVADIS LCLs -- and discover twice as many or more eQTLs per person in BRGR and AFR LCL as compared to PPMI (N = 1,251; 20,525 eQTLs) and GEUVADIS Europeans (N = 358; 7,171 eQTLs). We show that increased heterozygosity in individuals of African as compared to European ancestry strongly impacts eQTL discovery. Using these African and European datasets, we further (a) show that comparisons of heritability of gene expression are confounded by population substructure; (b) show that local ancestry does not largely affect eQTL discovery; (c) explore strategies for eQTL calling in a multi-ancestry setting; (d) investigate the relationship between genes differentially expressed between cohorts and eQTLs unique to a cohort; and (e) using both African American GWAS from the Million Veterans Project and a GWAS of height in 23andMe research participants of African ancestry, examine results of GWAS variant-to-gene mapping when ancestries are or are not aligned between GWAS and eQTL studies. Consistent with consent provided, we will make these data publicly available by 2023 to all qualified researchers -- academic or industry -- in the hopes that they will accelerate research that reduces health disparities and improves outcomes for individuals of African ancestry.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2701. The human myocardial transcriptomic response to the interaction effect of genetic variants and ischemic stimulus identified by a novel approach.

Authors:

A. Yazdani¹, L. Luigi Pesce², M. Oliva², J. gorham³, C. Seidman³, J. Seidman⁴, S. Body⁵, S. Aranki³, B. Stranger⁶, J. Muehlschlegel¹; ¹Brigham & Women's Hosp. / Harvard Med. Sch., Boston, MA, ²Northwestern Univ., Evanston, IL, ³Harvard Med. Sch., Boston, MA, ⁴Harvard Med Sch, Boston, MA, ⁵Boston Univ., Dover, MA, ⁶Northwestern Univ., Chicago, IL

Abstract Body:

Interindividual gene expression variation is impacted by both genetic variation and environmental exposures. We investigated genetic-transcriptomic relationships in response to myocardial ischemia as an environmental factor to understand mechanisms that underlie the pathological processes of myocardial ischemia in humans. Left ventricular biopsies from 118 patients undergoing aortic valve replacement surgery were obtained at baseline and after an average of 79 minutes of cold cardioplegic arrest/ischemia. In this study, we explored the genetic basis of gene expression through mapping expression quantitative trait loci (eQTL), with a focus on response eQTL (reQTL), defined as the effect of genetic variants on the gene expression response to stimulus, in this case, ischemia. Since genetic variants affect expression level regardless of stimuli, we need to account for baseline expression while estimating the reQTL. While conventional reQTL approaches subtract/divide expression values, we instead proposed to adjust for the “effect” of baseline expression on post expression as a complementary approach. Focusing on cis-QTLs, where gene expression levels are affected by variants in the gene body and 1 megabase up and 1 megabase down the gene, the conventional approaches identified at most 11 reQTL at a false discovery rate (FDR) ≤ 0.25 . The conditional model introduced above identified 530 reQTL at FDR ≤ 0.05 , where 64 were among the post-ischemia-specific eQTL, 35 among baseline-specific eQTL, and 5 were neither baseline nor post-ischemia eQTL, and 426 were eQTL at baseline and post-ischemia. To assess the role of post-ischemia eQTL and reQTL in myocardial ischemic injury, we integrated the summary statistics of the eQTLs with a separate set of 1751 patients undergoing a coronary artery bypass graft surgery and estimated the causal effect of eQTLs on troponin using Mendelian randomization after assessing the lack of pleiotropic action of eQTLs utilized as instrumental variables. Troponin is a complex protein integral to cardiac muscle contraction and elevated levels in the blood after cardiac surgery. In addition, we carried out a colocalization analysis between cis-eQTLs and variants associated with heart diseases and other enrichment analyses including replication in GTEx and in a mouse model. We identified 19 cis-regulatory genes acting as mediators of myocardial infarction effects, such as MRAS (Ras family GTPase), NBR2 (long non-coding RNA), and MED15 (subunit of a transcription co-activator).

Molecular Effects of Genetic Variation Posters - Thursday

PB2702. The importance of the use of genomics in early diagnosis of tuberous sclerosis: case report.

Authors:

A. Lambraño^{1,2}, L. Moreno Giraldo³; ¹1045701331, cali, Colombia, ²Univ. Libre, Cali , Valle del Cauca, Colombia, ³Orphan Diseases Program, New Clinic of Cali., cali, Colombia

Abstract Body:

INTRODUCTION: tuberous sclerosis is a rare disorder with multisystemic clinical manifestations that can compromise vital organs such as the kidney, lung and heart, which requires early diagnosis to provide timely and targeted treatment, improving the prognosis and reducing the morbidity and mortality attributed to these pathologies. **OBJECTIVE:** To establish the importance of the use of genomics and the phenotype-genotype correlation for the diagnosis, treatment, follow-up, prognosis, genetic counseling of tuberous sclerosis **MATERIALS AND METHODS** case report with clinical suspicion of tuberous sclerosis with genetic confirmation by having a Pathogenic heterozygous variant in the TSC2 gene **RESULTS:** A pathogenic heterozygous deletion was found that changes a cytosine at position 2539 of the TSC2 gene cDNA (c.2539delC), leading to a premature stop codon at amino acid 893 (p.Leu847Cysfs*47) into a protein of 1,807 amino acids with pathogenic clinical significance **CONCLUSIONS:** Tuberous sclerosis complex is an orphan disease for Colombia given its low population prevalence, with a high morbidity and mortality burden due to multisystem involvement. Its confirmation is carried out through molecular-genomic methods that allow establishing a phenotype-genotype correlation and searching for possible affected organs or systems in individuals, bringing us closer to personalized and precision medicine

Molecular Effects of Genetic Variation Posters - Wednesday

PB2703. The Multiple *de novo* Copy Number Variant (*MdnCNV*) phenomenon: peri-zygotic DNA mutational signatures and multilocus pathogenic variations

Authors:

H. Du¹, A. Jolly¹, C. Grochowski¹, B. Yuan¹, M. Dawood¹, S. Jhangiani¹, D. Muzny¹, J. Fatih¹, Z. Coban-Akdemir^{1,2}, A. Scheuerle³, K. Witzl⁴, J. Posey¹, M. Pendleton⁵, S. Juul⁵, P. Hastings¹, W. Bi¹, J. Lupski¹, F. Sedlazeck¹, R. Gibbs¹, C. Carvalho⁶, P. Liu¹; ¹Baylor Coll. of Med., Houston, TX, ²The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ³Univ. of Texas, Southwestern Med. Ctr., Dallas, TX, ⁴Univ. of Ljubljana, Ljubljana, Slovenia, ⁵Oxford Nanopore Technologies, New York, NY, ⁶Pacific Northwest Res. Inst., Seattle, WA

Abstract Body:

Background: The multiple *de novo* copy number variant (*MdnCNV*) phenomenon is a rare peri-zygotic mutational event observed once in at least every 12,000 individuals referred for clinical Chromosome Microarray Analysis (CMA). Variable congenital abnormalities have been observed in individuals with the *MdnCNV* phenomenon. We investigated genetic variation-causing genome instability and the underlying phenotypic rare disease trait manifestations in families with *MdnCNV*. **Methods:** Using several genomic analysis techniques, including CMA and short and long-read DNA sequencing, we characterized the *de novo* mutation signatures of a newly identified *MdnCNV* case. Trio-based rare variant analysis and single-based substitutions (SBS) signatures analysis on *dnSNV* was performed on this family and four previously published *MdnCNV* families to infer potential genetic etiologies. Lin semantic similarity scores informed quantitative human phenotype ontology analysis on three *MdnCNV* families to identify gene(s) driving or contributing to the clinical phenotype. **Results:** The newly identified *MdnCNV* family presents with eight non-mosaic *de novo* tandem direct duplications, each ~1 Mb long with microhomology present at 6/8 breakpoint junctions. CNV phasing reveals a 1:1 ratio of *dnCNV* parental origin. Enrichment of *de novo* single nucleotide variants (SNV: 6/79) and *de novo* indels (1/12) was found at and within 4 Mb of the *dnCNV* genomic regions. Characterization of *dnSNV* suggests an elevated post-zygotic SNV mutation rate in *MdnCNV* families. The rare variant analysis identified multiple maternally inherited high-impact variants affecting the DNA replication repair pathway. SBS Mutational signature analysis suggests one *MdnCNV* family shows a potential mutation signature associated with DNA polymerase epsilon. Phenotype analysis suggests that gene(s) within *dnCNV* regions contribute to the observed proband phenotype in 3/3 of cases. CNVs in two cases, one with a contiguous gene duplication of *PMP22* and *RAI1* and another with *NSD1* and *SMARCC2* at separate loci, contribute to phenotype manifestation. **Conclusions:** The characteristics of *dnCNVs* in the newly described case were consistent with a microhomology-mediated break-induced replication (MMBIR) driven mechanism generating multiple *dnCNVs* during the peri-zygotic period. Maternal genetic variants in DNA repair genes could increase peri-zygotic genomic instability, increasing the post-zygotic SNV and CNV mutation rate. The variable phenotypic features observed across *MdnCNV* probands reflect the contribution of more than one locus, supporting *de novo* multilocus pathogenic variation (MPV).

Molecular Effects of Genetic Variation Posters - Thursday

PB2704*. The role of introgressed archaic DNA in the human genome.

Authors:

R. Young¹, J. Kang², A. S. Ramgolam¹; ¹Univ. of Edinburgh, Edinburgh, United Kingdom, ²Zhejiang Univ. - Univ. of Edinburgh Inst., Haining, China

Abstract Body:

Modern humans harbour a substantial component of DNA which has been introduced from archaic hominid species through historical introgression events. Neanderthal DNA is more common within the European population, while the highest percentage of Denisovan DNA is found within Asian genomes. However, we do not yet have a comprehensive understanding of the biological contribution this introgressed, archaic DNA makes to contemporary human population variation. Here, we focus on gene expression regulation as a molecular phenotype to address this question genome-wide across tissues and populations.

We first integrated publicly available data on the locations in the human genome of (1) Neanderthal and Denisovan DNA and (2) various expression quantitative trait loci (eQTLs) datasets from the GTEx project and others. We found that all tissues for which we had eQTL data were enriched for archaic eQTLs (relative to non-regulatory polymorphisms, Fisher's exact tests, $p < 0.05$). These enrichments were particularly prominent for Neanderthals when compared to European polymorphisms and when compared to Asian polymorphisms, i.e. in the populations where each archaic genomic is more commonly found. The effect sizes of Neanderthal and Denisovan eQTLs were less than for tissue-matched eQTLs from the same population (Mann-Whitney tests, $p < 0.01$). We reconstructed the evolutionary trajectory of eQTL alleles across Neanderthal and modern humans where we observed a bias towards down-regulation of gene expression by the Neanderthal allele (0.95-fold, Mann-Whitney test, $p < 2.2 \times 10^{-16}$). We are currently working to replicate this analysis for Denisovan eQTLs.

We next investigated whether those genes whose expression was regulated by archaic DNA have previously been associated with human disease. Using the OMIM database, we found 3,641 unique disease-associated genes which are regulated, in at least one tissue, by Neanderthal or Denisovan DNA. Neanderthal eQTLs are more likely to regulate genes associated with disease (1.08-fold, Fisher's exact test $p = 5.01 \times 10^{-52}$) while Denisovan eQTLs are less likely (0.96-fold, Fisher's exact test $p = 1.10 \times 10^{-8}$). Ongoing work is investigating whether these archaic alleles are protective or drive susceptibility to disease.

Our work suggests that regulatory archaic DNA has preferentially retained its regulatory potential and that this contribution varies across global populations. Regulatory Neanderthal DNA is more commonly associated with disease than Denisovan. These results have important consequences for interpreting the biological significance of evolutionary recent additions to the human genome.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2705. The Unmasking Of 'Mitochondrial Adam' And Structural Variants Larger Than Point Mutations As Stronger Candidates For Traits, Disease Phenotype And Sex Determination

Authors:

A. Singh^{1,2}; ¹Schiller Intl. Univ., Heidelberg, Germany, ²ABioTeQ, Delhi, India

Abstract Body:

Structural Variations, SVs, in a genome can be linked to a disease or characteristic phenotype. The variations come in many types and it is a challenge, not only determining the variations accurately, but also conducting the downstream statistical and analytical procedure. Structural variations, SVs, with size 1 base-pair to 1000s of base-pairs with their precise breakpoints and single-nucleotide polymorphisms, SNPs, were determined for members of a family. The genome was assembled using optimal metrics of ABySS and SOAPdenovo assembly tools using paired-end DNA sequence. An interesting discovery was the mitochondrial DNA could have paternal leakage of inheritance or that the mutations could be high from maternal inheritance. It is also discovered that the mitochondrial DNA is less prone to SVs rearrangements than SNPs, which propose better standards for determining ancestry and divergence between races and species over a long-time frame. Sex determination of an individual is found to be strongly confirmed using calls of nucleotide bases of SVs to the Y chromosome, more strongly determined than SNPs. We note that in general there is a larger variance (and thus the standard deviation) in the sum of SVs nucleotide compared to sum of SNPs of an individual when compared to reference sequence, and thus SVs serve as a stronger means to characterize an individual for a given trait or phenotype or to determine sex. The SVs and SNPs in HLA loci would also serve as a medical transformational method for determining the success of an organ transplant for a patient, and predisposition to diseases a priori. The sample anonymous dataset shows how the de-novo mutation can lead to non-inherited disease risk apart from those which are known to have a disease to mutation association. It is also observed that mtDNA is highly subjected to mutation and thus the factor for a lot of associated maternally inherited diseases. 'mitochondrial Adam' can be a fair reality as certainly the biparental mode of mtDNA puts in question the theory of 'mitochondrial Eve'. SVs would serve as a stronger fingerprint of an individual contributing to his traits, sex determination, and drug responses than SNPs.

Molecular Effects of Genetic Variation Posters - Thursday

PB2706. *TP53* codon 72 polymorphism modulates macrophage polarization through altered PI3K/Akt signaling pathways

Authors:

A. Silwal, B. Reese, N. Lodhi, M. Karbowniczek, M. Markiewski; Texas Tech Univ. Hlth.Sci. Ctr., Abilene, TX

Abstract Body:

The purpose of this study is to determine the role of the most common p53 single nucleotide polymorphism (SNP) at codon 72, which encodes proline (P72) or arginine (R72), in the regulation of macrophage polarization. Bone marrow-derived macrophages (BMDMs) from human p53 knock-in (Hupki) mice, in which exons 4-9 of the endogenous mouse p53 allele were replaced with the homologous human p53 gene sequence, carrying P72 and R72 variants were treated with lipopolysaccharide (LPS) to activate macrophages. Signaling pathways involved in macrophage activation were analyzed by RT-PCR, Western blotting, and immunofluorescence. A highly selective Akt inhibitor MK-2206 was used to determine the effect of Akt blockade in R72 macrophage polarization and signaling. Volumes of tumors generated by subcutaneous injections of tumor cells (TC1), mixtures of tumor cells, and LPS-stimulated P72 (TC1+P72^{LPS}) and R72 (TC1+R72^{LPS}) macrophages were measured. LPS stimulation of BMDMs led to a greater upregulation of M1 genes (*Socs1* & *Nos2*) in P72 compared to R72 macrophages. Further, we examined the activation of PI3K/Akt as this pathway is essential for restricting proinflammatory (M1) and promoting anti-inflammatory (M2) responses in toll-like receptor (TLR4)-stimulated macrophages. The expression of p-Akt (S473) was increased to a greater extent in R72 vs. P72 macrophages upon LPS treatment. Consistent with increased Akt activity in R72 macrophages, the p-FOXO3a (S253) residue was greater in these cells. This phosphorylation inhibits FOXO3a transcriptional activity via nucleus to cytoplasm shuttle. NF-kB activation, nuclear translocation, and subsequent transcriptional regulation is a key to the induction of several proinflammatory genes. Cytoplasmic FOXO3a was demonstrated to inhibit NF-kB via direct binding to NF-kB in the cytoplasm and prevent its nuclear translocation. We observed increased cytoplasmic colocalization of FOXO3a and NF-kB upon LPS stimulation by Immunofluorescence. Western blotting further corroborated these data showing impairment of NF-kB nuclear translocation. Following treatment with MK2206, we observed reduced p-FOXO3a (S253), enhanced NF-kB nuclear translocation, and increased expression of M1 genes in R72 (*Socs1* & *Nos2*) cell. The addition of LPS-stimulated P72 macrophages to tumor cells reduced tumor growth, in contrast to R72 macrophages in a syngeneic model of HPV-induced cancer. Here, we report that macrophages carrying R72 variant are biased toward M2 phenotype through altered NF-kB nuclear translocation which also impacts in vivo macrophage function as R72-LPS stimulated macrophages lose ability to reduce tumor growth.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2707. Two types of variants pointing out a crucial role for RNF216 in healthy and diseased brain.

Authors:

L. Versluys¹, A. Sieben², P. Ervilha Pereira¹, N. Schuermans¹, E. Mondragon Rezola³, P. LeBlanc¹, M. Galas⁴, P. Santens⁵, E. Bogaert¹, B. Dermaut¹; ¹Ghent Univ., Ghent, Belgium, ²Antwerp Univ., Antwerp, Belgium, ³Univ. Hosp. Donostia, Gipuzkoa, Spain, ⁴Univ. Lille, Inserm, CHU Lille, CNRS, Lille NeuroSci. & Cognition, Lille, France, ⁵Ghent Univ. Hosp., Ghent, Belgium

Abstract Body:

Mutations in RNF216, an E3-ubiquitin ligase, have been identified as the genetic cause of Gordon Holmes syndrome (GHS), a rare recessive neurodegenerative disease. RNF216 mutations can be roughly classified into two groups based on their location within the protein. The first group of mutations cluster within the catalytic domain and lead to a loss of ubiquitin ligase activity. In the brain of one patient with a homozygous R751C variant in the catalytic domain, intranuclear ubiquitin-positive inclusions were found (Margolin et al, 2013). The second group of mutations reside in the N-terminal region, outside the catalytic domain with preserved ligase activity. Still, the mechanism by which this second group of mutations is leading to loss-of-function is elusive. Interestingly, in a previously reported patient with a homozygous N-terminal G456E mutation outside the catalytic domain (Santens et al, 2015) we replicated the neuropathological findings observed in the R751C patient, pointing towards a similar loss-of-function mechanism at play. Immunohistochemical analysis shows ubiquitin and p62 immunoreactive intranuclear inclusions which are negative for hyperphosphorylated tau, β -amyloid and TDP-43. Intriguingly, we identified a patient combining variants from both groups: one variant, C752Y, in the catalytic domain and one, W449R, in the N-terminus, suggesting that indeed both variants convergence into a loss-of-function pathway. As the endogenous function of RNF216 is crucial to understand RNF216-mediated neurodegeneration, we want to investigate beyond the catalytic activity, a novel and additional nuclear function for RNF216. This will be enabled by in-vitro experiments and proteomics on human GHS-patient brain tissue and fibroblasts, hiPSC-derived neuronal cultures expressing RNF216 disease-variants, and BioID. The physiological and pathological mechanisms explored in this project hold the potential to greatly improve our understandings of GHS, shed light on novel key players in neurodegeneration, and open new avenues for a novel therapeutic target for this rare neurodegenerative disease.

Molecular Effects of Genetic Variation Posters - Thursday

PB2708. TxEWAS identifies genetic modifiers of drug response and side effects through retrospective gene-environment association studies.

Authors:

M. Sadowski, M. Thompson, J. Mefford, R. Border, S. Sankararaman, N. Zaitlen; Univ. of California, Los Angeles, Los Angeles, CA

Abstract Body:

The impact of environmental exposures on complex traits can be modulated by genetic variation. Such gene-environment interactions (GxE) may underlie population variability in drug response, susceptibility to carcinogens, and resilience in aging. However, interaction effects at individual genetic variants are typically small, and few examples have been discovered to date. Moreover, when identified, associations at individual variants are difficult to interpret functionally.

To address these concerns, we develop a new framework to test for GxE effects in an imputed expression-trait association model—TxEWAS, which extends the TWAS model. TxEWAS trains a linear predictor of the *cis* genetic component of expression in a cohort of unrelated individuals with both genotypes and gene expression measured. It then uses this predictor to impute expression levels and scan for expression-environment interactions in a cohort of phenotyped individuals (without directly measured expression).

Importantly, TxEWAS accounts for the heterogeneity of phenotype variance—a continuously underappreciated source of bias in conventional GxE methods. Realistic biobank scale simulations showed that failing to account for this bias can cause a threefold increase in the false positive rate.

We applied TxEWAS to eight drug exposure-response pairs in the UK Biobank, where responses were measures of target effects or side effects of the tested drugs (statins, warfarin and methotrexate). We identified 191 significant interacting genes, potentially involved in response to those drugs. In particular, our analysis revealed *PCSK9* and *GIPR* as one of the potential modulators of the effect of statin use on LDL cholesterol and HbA1c (a clinical marker of glycemic control) levels respectively. *PCSK9* was previously postulated to modulate efficacy of the statin therapy. *GIPR* was demonstrated by mouse model studies to be involved in regulation of insulin levels in the presence of elevated glucose, and shown to be associated with increased risk of developing hyperinsulinemia under antipsychotic drug treatment of schizophrenia. Together, these results demonstrate the power of the TxEWAS approach, and in particular, its potential to offer an insight into genetic determinants of individual drug response.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2709. Types of cis- and trans-gene regulation of expression quantitative trait loci across human tissues.

Authors:

A. Fu¹, J. Kvamme¹, M. Badsha², E. Martin¹, J. Wu³, M. Megheib¹, X. Wang³; ¹Univ. of Idaho, Moscow, ID, ²St Jude Children's Res. Hosp., Collierville, TN, ³Inst. of Basic Med. Sci., Chinese Academy of Med. Sci., Beijing, China

Abstract Body:

Expression quantitative trait loci (eQTLs) have been identified for most genes in the human genome across tissues and cell types. While most of the eQTLs are near the associated genes, some can be far away or on different chromosomes, with the regulatory mechanisms largely unknown. Although the canonical model is cis-gene mediation (eQTL → cis-gene → trans-gene), other modes of relationships have been observed, although only in a couple of cell lines. There is also experimental evidence supporting functionally important inter-chromosomal interactions.

Here, we study cis- and trans-regulation of eQTLs across nearly 50 tissues and cell types. Specifically, we constructed trios consisting of an eQTL, its cis-gene and trans-gene and inferred the regulatory relationships with causal network inference. Similar to the GTEx consortium, we focus on protein-coding genes and long noncoding RNAs (lncRNAs). Across tissues, the number of trios selected for this analysis ranges from 627 to 4668 (median: 2279), mainly due to the varying sample sizes (median: 235). We performed PEER normalization on the gene expression data for each tissue and identified the principal components (PrCs) that are significantly associated with each trio; these PrCs are potential confounding variables. We then applied our MRPC method (Badsha and Fu, 2019; Badsha et al., 2021) to perform causal network inference on each trio and its associated PrCs, treating each variable as a node in the network. Our extensive simulation studies had demonstrated that MRPC tends to be conservative and achieves better recall, precision and stability across many scenarios on networks of small to moderate sizes.

We identify multiple types of regulatory networks for trios: across all the tissues, above 50% of the tested trios are inferred to be conditionally independent, where the two genes are conditionally independent given the genotype of the eQTL (cis-gene ← eQTL → trans-gene). Around 1.8% of the trios are inferred to be mediation (eQTL → mediator → target), around 0.3% fully connected among the three nodes, and just a handful v-structures (eQTL → gene 1 ← gene 2). The remaining trios are inferred to have just one edge. Unexpectedly, across the tissues, on average over 70% of the mediation trios have the trans-gene as the mediator. Most of the cis- and trans-gene mediators are also tissue specific. To verify the results, we further applied to these trios other methods, such as a less conservative version of MRPC, and the GMAC method (Yang et al., 2017) that detects only mediation. These analyses provided consistent results.

Molecular Effects of Genetic Variation Posters - Thursday

PB2710*. UBA5-related Epilepsy: From Cellular Models to Novel Therapies

Authors:

H. Chen¹, E. Almanza Fuerte¹, H. Mefford²; ¹St Jude Children's Res. Hosp., Memphis, TN, ²St. Jude Children's Res. Hosp., Memphis, TN

Abstract Body:

UBA5 encodes an E1 activating enzyme involved in the ubiquitin-fold modifier 1 (UFM1) pathway, known as ufmylation. Biallelic mutations in *UBA5* cause a neurodevelopmental disorder characterized by epilepsy, severe developmental delay, and growth failure. In ~70% of affected individuals, one of the pathogenic alleles is a missense variant, p.A371T, that produces a hypomorphic protein while the other is either a nonsense or missense variant that encodes a nonfunctional protein. The p.A371T variant is present in the general population at a frequency of 0.19% (gnomAD). Interestingly, at least three healthy adults with homozygous p.A371T variants have been reported. Thus, despite the decreased enzymatic activity of the p.A371T allele, the combination of two hypoactive alleles provides enough enzyme activity to prevent disease outcome. This suggests that augmenting the amount of the p.A371T allele may be an effective therapy. SINEUPs are long noncoding antisense RNA that can increase translation of sense mRNAs of target genes. Synthetic SINEUPs have been used to restore motor deficits in a murine model of Parkinson disease. Thus, we propose that increasing the level of the p.A371T protein using SINEUP to restore UBA5-mediated ufmylation pathway can alleviate symptoms in UBA5 patients. We generated cell lines with pathogenic alleles representing those identified in patients, specifically, UBA5^{A371T/F292X}. We observed an increase of ufmylation pathway proteins by immunoblotting and qRT-PCR analysis. UBA5-mediated ufmylation pathway is involved in ER stress response and apoptosis, thus our analysis will focus on the ER stress, unfolded protein response (UPR) pathways and cell death. We then noted that UBA5^{A371T/F292X} showed increased abundance of ER stress proteins and ER volume, suggesting elevated base-level ER stress compared to UBA5^{Wt/Wt}, UBA5^{Wt/A371T} and UBA5^{Wt/Q303X}. It is unknown whether UBA5^{A371T/A371T} phenocopies UBA5^{Wt/Wt} at the cellular level. We generated a cell line with homozygous p.A371T allele and demonstrated that UBA5^{A371T/A371T} behaves similar to UBA5^{Wt/Wt}, showing no elevated base-level ER stress, unlike UBA5^{A371T/F292X} cells. We designed various UBA5-specific SINEUPs and showed they elevate the levels of p.A371T allele in UBA5^{A371T/F292X} cells. We are now testing whether this increase can rescue UBA5 mutant cellular phenotypes. We demonstrated that UBA5^{A371T/A371T} phenocopies UBA5^{Wt/Wt} at the cellular level, and SINEUPs can be utilized to augment the levels of p.A371T protein in UBA5^{A371T/F292X} cells. These findings further support that restoration of the hypomorphic p.A371T by SINEUP may be a novel therapeutic outlet.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2711. UK Biobank whole genome sequencing and large-scale proteomics identifies novel protein quantitative trait loci (pQTLs) contributing to human disease.

Authors:

L. Hou¹, A. H. Li¹, E. H. Baugh¹, A. Doostparast Torshizi¹, H. A. Hejase¹, T. Sato², H. Bell-Temin³, A. Hart⁴, S. Guo¹, B. A. J. Sarver¹, S. Li¹, M. Black¹; ¹Population Analytics, Data Sci. Analytics & Insights (DSAI), Janssen R&D, Spring House, PA, ²Early Detection & Data Sci., World Without Disease Accelerator, Janssen R&D, Spring House, PA, ³Discovery Proteomics Platforms, Discovery Technologies and Molecular Pharmacology, Janssen R&D, Cambridge, MA, ⁴Immunology Translational Sci., Immunomics, Janssen R&D, Spring House, PA

Abstract Body:

Background: Previous large-scale protein quantitative trait loci (pQTLs) studies focused primarily on common variants based on genotype arrays. The contribution of rare variants to protein levels genome-wide remains largely unknown. Here, we utilized proteomics and whole genome sequencing (WGS) data from the UK Biobank to comprehensively map plasma pQTLs. **Methods:** We identified pQTLs for 1,463 proteins assayed by Olink™ from plasma samples collected from 13,227 UK Biobank European participants with WGS data. We performed variant-level genetic association analyses for ~23M variants with minor allele count ≥ 5 (lowest MAF=0.026%). We also performed gene and exon-level association analyses by collapsing rare functional variants (i.e., clinically classified pathogenic/likely pathogenic, and in silico predicted deleterious and loss of function variants). **Results:** Single-variant analyses identified 2,374 significant sentinel pQTLs for 1,088 proteins ($p < 1.46 \times 10^{-12}$), including 940 cis-pQTLs and 1,434 trans-pQTLs. Of these, 18 loci were not previously identified by the UK Biobank Pharma Proteomics Project (UKB-PPP) consortium-led analysis of SNP array data. Gene-based analyses revealed 258 significant sentinel pQTLs (207 cis-pQTLs and 51 trans-pQTLs, $p < 2.97 \times 10^{-10}$), of which seven loci are novel. Exon-based analysis identified 152 significant sentinel pQTLs (132 cis-pQTLs and 20 trans-pQTLs, $p < 1.69 \times 10^{-10}$), including one novel trans-pQTL. For example, gene-level collapsing analyses yielded significant association between rare (MAF<0.1%) functional variants in *STAB1* and higher levels of NOTCH3 ($p = 3.89 \times 10^{-12}$, $\beta = 0.53$), as well as higher levels of MYOC and LTBP2, all of which constitute novel pQTL signals in trans. Aberrant NOTCH3 signaling and overexpression is known to be oncogenic and associated with resistance to chemotherapy and decreased survival. NOTCH3-targeted antibody drug conjugates have recently been investigated for treatment of solid tumors, and an understanding of the role of STAB1 in upregulating NOTCH3 may inform therapeutic efficacy, safety, and patient stratification. **Conclusion:** Our findings provide a comprehensive genetic atlas of the human plasma proteome, including rare/ultra-rare pQTLs not previously described. Such results may be used to finemap pQTLs and refine our understanding of causal variants toward the goal of improved identification of targets for safe and effective therapies.

Molecular Effects of Genetic Variation Posters - Thursday

PB2712. Unbiased linkage of SNV to expression in single cells for regulatory variant discovery influencing cancer drug resistance

Authors:

I. Salas-Gonzalez, T. Morozova, D. Arvapalli, S. Velivela, J. West, G. Harton, J. Zawistowski, V. Weigman; BioSkryb Genomics, Durham, NC

Abstract Body:

Single-cell RNAseq methodologies have revolutionized the ability to define cell type identity and transcriptional phenotypic diversity within a tumor sample, yet in most studies DNA-level variation contributing to driving this diversity of gene expression remains uncharacterized. We employed ResolveOME chemistry to bridge this gap within an individual cell, by taking inventory of genome-wide single nucleotide variation (SNV) and cross referencing it with full-transcript RNAseq data in an acute myeloid leukemia (AML) cell line model of drug resistance. The catalog of genome-wide SNV in conjunction with transcriptomic data in this model powered an association screening to identify regulatory SNVs with biased prevalence in single cells resistant to the FLT3 inhibitor quizartinib vs treatment-naïve single cells, which then were correlated with differential expression of genes proximal to the variant in the same cell. Initially, we focused on nucleotide variation within core promoter regions and within gene bodies. Fitting a zero-inflated linear model to a matrix of expression and genotype across all single cells revealed an intronic heterozygous variant in *MYC* in parental single cells, while, in contrast, the alternate allele was absent in quizartinib-resistant cells. Upregulation of *MYC* transcript was correlated with the absence of this variant in resistant single cells, suggesting that the single nucleotide change may have intronic enhancer activity as opposed to intronic regulation of splicing. Additionally, the screen uncovered a heterozygous single nucleotide change within 5kb 5' of the transcriptional start site of the mRNA binding factor *PABPC4* as a candidate promoter variant not present in parental single cells yet harbored by 50% of the quizartinib resistant single cells displaying *PABPC4* expression relative to parental cells. Differential expression analysis uncovered the enhancer factor *CEBPA* as upregulated in the resistant single cells, and the genomic component of ResolveOME identified two SNVs 20 kb upstream of the locus representing putative *CEBPA* enhancer variants influencing the differential gene expression. The co-identification of *CEBPA* and *PABPC4* through this multi-omic approach suggest that quizartinib resistance in this model is mediated in part by global gene regulation through enhancer modulation and through mRNA stability/translational regulation, respectively. We are furthering enhancer variant detection in intergenic space by overlaying ChIP-seq, chromatin accessibility, and transcription factor binding site data to prioritize candidates contributing to resistance in this AML drug resistance model.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2713. Uncovering sources of human gene expression variation in a globally diverse cohort.

Authors:

D. Taylor, M. Tassia, R. McCoy; Johns Hopkins Univ., Baltimore, MD

Abstract Body:

Genetic variation influencing gene expression accounts for a large proportion of phenotypic variation within and between species, including humans. While previous large-scale studies of human gene expression have been invaluable for revealing mechanisms of genome function, they have largely focused on individuals of European ancestries. This bias may diminish the generalizability of results, while limiting broader understanding of the diversity and evolution of gene expression across human populations. To address these challenges, we performed RNA sequencing of lymphoblastoid cell lines derived from 731 globally-diverse individuals from the 1000 Genomes Project, evenly drawn from each of the 5 superpopulations and 26 subpopulations. We quantified gene expression in this dataset, and characterized patterns of gene expression across populations. On average, we observed that the vast majority (92%) of gene expression variation is partitioned within as opposed to between populations, mirroring patterns of human diversity at the level of DNA sequence. Moreover, gene expression variation is greatest among African populations and declines as a reflection of the serial founder effects during historical eastward expansions across the globe. By combining the expression data with existing high-coverage whole-genome sequencing data from the same individuals, we mapped associations between genetic variants and the expression of nearby genes (i.e., cis expression quantitative trait loci or "eQTLs"). We identified over 15,000 eGenes (genes with at least 1 eQTL) and over 1.9 million eQTL variants across the entire sample. In addition to replicating the results of previous large-scale eQTL studies (e.g. Geuvadis), we identified over 1.7 million novel eQTL variants not reported by Geuvadis. This includes many variants that are specific to Asian or Admixed American samples, which were absent or under-represented in previous molecular association studies. Such eQTL data could be particularly useful for interpreting the results of past and future GWAS that include individuals of diverse ancestries. Additionally, the diverse history of recombination events in sample breaks up linkage disequilibrium and allows us to narrow in on smaller sets of candidate causal variants to isolate the genetic and epigenetic mechanisms that modulate transcription, RNA stability, and/or degradation. Together, our study expands understanding of gene expression diversity across human populations and offers an inclusive resource for studying the evolution and function of human genomes.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2714. Unsupervised Density Estimation for Noncoding Variant Effect Prediction

Authors:

R. Rastogi, R. Chung, A. Reddy, N. Ioannidis; Univ. of California, Berkeley, Berkeley, CA

Abstract Body:

Cross-species conservation is an integral feature in most previous computational methods that predict the pathogenicity of human noncoding variants. However, the utility of these conservation metrics, which are computed over multiple sequence alignments (MSAs), is hindered by the limited number of species that are closely related to humans and by the sparsity in whole-genome alignments. However, buoyed by the recent 241-way alignment of eutherian mammals produced by the Zoonomia Project, we develop new unsupervised density estimation models to predict noncoding variant effects and learn meaningful representations of regulatory regions using only MSAs, and no other functional annotations, as input. Specifically, we train Potts models, which learn pairwise interactions, and variational autoencoders, which learn higher-order interactions, on noncoding regions of interest. We will also train a single large language model, similar to the MSA Transformer, using masked autoencoding to learn from alignments throughout the genome. Aside from comparing each of these models to existing methods on noncoding pathogenicity prediction, we will also finetune the language model on downstream tasks, such as cell-type specific accessibility prediction, to validate that the model learns meaningful representations of regulatory regions.

Molecular Effects of Genetic Variation Posters - Thursday

PB2715. Use of reference population data for resolving variants of uncertain significance in hematopoietic genes.

Authors:

V. Avramovic, C. Diao, T. Maroilley, M. Tarailo-Graovac; Univ. of Calgary, Calgary, AB, Canada

Abstract Body:

Limited ability to interpret variant effects is one of the major bottlenecks in genome-wide diagnostic approaches. The interpretation of certain variant types, including deep intronic, 5'UTR and deep exonic synonymous variants, is especially difficult.

Here we explore a strategy to improve variant interpretation in hematopoietic genes. Germline variants in many hematopoietic genes have been associated with rare diseases. It has been suggested that somatic variants in dozens of hematopoietic genes implicated in human rare diseases may be present in the blood-based reference population datasets. These variants often drive the process of clonal expansion, by which their carrier cells become more abundant in the blood tissue. Consequently, these variants appear in reference population datasets, where they are often misclassified as germline variants. We proposed that the presence of these somatic variants in reference population databases, such as TOPMed and gnomAD, can be used as an important feature for prediction of their functionality in rare diseases, in germline. To test this, we assessed the variants from one of the hematopoietic genes with known rare disease association, *ASXL1*. The results we obtained with *ASXL1* nonsense variants (severe pediatric rare disease, Bohring-Opitz syndrome is due to *ASXL1* haploinsufficiency) confirmed our basic hypothesis. The majority of the *ASXL1* nonsense variants observed in "healthy" individuals in the TOPMed database show signs of somatic origin, while their assessment with a selected combination of *in silico* pathogenicity predictors returned very high pathogenicity scores. Next, we expanded the assessments to include deep intronic, 5'UTR and deep exonic synonymous variants of uncertain significance (VUS) in the *ASXL1*. We reasoned that similar to nonsense variants, the transcript altering ones will present mosaic effects in population databases. Our analyses revealed a number of variants with driver potential, which we are experimentally validating for their transcript effects. The experimental validations will help support our hypothesis that evidence of somatic mosaicism in blood-derived population databases suggests driver/pathogenic effect of variants affecting hematopoietic genes, yet currently considered as VUS.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2716. Using massively parallel reporter assays to dissect context-specific regulatory grammars in type 2 diabetes

Authors:

A. Tovar, Y. Kyono, M. Bose, A. Varshney, J. Kitzman, S. Parker; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Genome-wide association studies have revealed that ~90% of disease-associated genetic variants are found in non-coding regions, including in type 2 diabetes (T2D), where they are expected to modify gene regulation and downstream disease mechanisms. The framework to translate the impact of non-coding variants on specific transcriptional processes is ill-defined, underscoring the need for a set of regulatory grammars to accelerate efforts to link genetic associations with specific pathogenic consequences. Here, we used a massively parallel reporter assay (MPRA) to screen enhancer activity across a panel of 197-bp fragments spanning >10k T2D- and metabolic trait-associated variants. To measure enhancer activity across several promoter and enhancer contexts in a diabetes-relevant cell model, we cloned these fragments up- or downstream of a reporter gene driven by a synthetic housekeeping promoter (SCP1) or the cell-specific human insulin (INS) promoter, and delivered the library to a pancreatic beta cell line (832/13 rat insulinoma). Next, we examined enhancer activity bias across this library (FDR < 0.05) based on position and linked promoter. Two unique subsets of fragments emerged: one with positional bias (n = 702/11,656) and one with promoter bias (n = 698/11,656). The former set was evenly distributed across up- and downstream activity bias, while a majority of fragments in the latter set had higher enhancer activity when paired with the cell-specific INS promoter (n = 512/698). To identify sequence features associated with promoter preference, we used Lasso regression with 562 genomic annotations and discovered that fragments with INS promoter-biased activity are enriched for HNF1 motifs. HNF1 family members HNF1A and HNF1B are key regulators of glucose metabolism and exonic mutations cause maturity onset diabetes of the young (MODY), suggesting genetic convergence between rare coding variants that cause MODY and common T2D-associated regulatory variants. We designed a follow-up MPRA containing HNF1 motif-enriched fragments to observe consequences of mutating or deleting these motifs on promoter-biased activity, and results from this library will be presented. Together, our study suggests that cell-specific regulatory activity is partially influenced by enhancer-promoter compatibility, and indicates that MPRA design is critical to capture context-specific regulatory processes at disease-associated genetic signals.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2718. Variant-to-gene mapping at the insomnia *WDR90* locus and subsequent luciferase assay analyses implicates *PIGQ* as an effector gene in sleep dysregulation

Authors:

S. Sonti¹, J. Palermo², A. Zimmerman^{3,1}, A. Chesi^{3,1}, C. Lasconi⁴, E. Brown², J. A. Pippin⁵, A. D. Wells^{3,1}, F. Doldur-Balli³, D. R. Mazzotti⁶, A. Pack³, P. Gehrman³, A. C. Keene², S. F. A. Grant^{1,3}; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Texas A&M Univ., College Station, TX, ³Univ. of Pennsylvania, Philadelphia, PA, ⁴CHOP, Philadelphia, PA, ⁵Children's Hosp. of Philadelphia, Philadelphia, PA, ⁶Univ. of Kansas Med. Ctr., Kansas City, KS

Abstract Body:

Sleep disturbances in humans are associated with several physiological complications and diseases. However, the pathophysiology and pathways involved in sleep disturbances are not fully understood. While initial studies in twins identified genetic contributions to variability in sleep patterns, GWAS have greatly assisted in identifying novel genetic variants associated with sleep traits and disturbances such as insomnia. To date, the largest GWAS of insomnia was performed in 1,331,010 individuals and identified 202 novel associated loci (Jansen et al, Nature Genetics 2019). Leveraging the 202 GWAS loci, our ATAC-seq/promoter focused Capture C approach in neural progenitor cells implicated proxy SNPs at 36 insomnia loci that both resided in a region of open chromatin region and contacted one or more open gene promoters. At the *WDR90* locus, our 3D genomic constraints shortlisted three insomnia-associated intronic proxy SNPs (rs3752495, rs8062685 and rs9932282; r^2 with sentinel SNP rs3184470 = ~1), all of which contacted the open promoters of Phosphatidylinositol N-acetylglucosaminyltransferase subunit Q (*PIGQ*) and NHL-Repeat-Containing Protein 4 (*NHLRC4*). The involvement of *PIGQ* in sleep was previously unknown; however, our group has reported that *PIGQ* deficient fruit-flies (*Drosophila melanogaster*) and zebra fish (*Danio rerio*) present with longer sleep phenotypes (Palermo, Chesi, Zimmerman et al, bioRxiv 2021). This warranted further validation of this variant-to-gene contact in a human cell model. Using a luciferase reporter assay, we sought to determine *in vitro* if these proxy SNPs influence promoter activity in SH-SY5Y cells (a *PIGQ* expressing neuronal cell line). One of the three proxy SNPs, rs3752495, revealed a 4-fold increase in luciferase reporter expression ($N=4$), indicating that this variant resides in an enhancer element influencing the expression of *PIGQ*. Our variant-to-gene approach therefore implicates *PIGQ* in the pathogenesis of human insomnia.

Molecular Effects of Genetic Variation Posters - Thursday

PB2719. Variation of 5'untranslated regions by dosage sensitivity reveals differences in post-transcriptional regulation important for interpreting variant effect

Authors:

N. Wieder, A. C. Martin-Geary, F. Lassen, E. N. D'Souza, M. Fernandes, R. Davies, N. Whiffin; Univ. of Oxford, Oxford, United Kingdom

Abstract Body:

Background: Untranslated regions (UTRs) are the regions flanking the protein-coding sequence of genes that form part of the mRNA, but are not translated into protein. UTRs are important mediators of post-transcriptional regulation, controlling mRNA stability, cellular localisation and the rate of translation into protein. UTRs are known to vary substantially across genes, both in terms of length, and the composition of regulatory elements within them, including linear and structural elements which mediate their effects through binding to various proteins and RNA. **Aims:** We investigated the differences in 5'UTR composition across deciles of tolerance to loss-of-function (LoF; LOEUF scores) and between different sets of disease genes (developmental disorder (DD) and cancer (C) genes sets). We investigated 5'UTR length, whether the 5'UTRs undergo splicing (intron number) and the number and type of upstream open reading frames (uORFs) within the 5'UTRs, using both uORF predictions and high-confidence uORFs from ribosome profiling data.

Results: Genes that are more sensitive to changes in dosage or that are linked to dominant disease have more complex 5'UTRs. Whilst 5'UTRs average 193bps in length, with 35% containing uORFs, LoF intolerant and disease gene 5'UTRs are significantly longer (DD: 368bp, $P=1.1 \times 10^{-41}$; C: 251bp, $P=1.8 \times 10^{-5}$; most intolerant LOEUF decile: 273bp, $P=1.1 \times 10^{-98}$) and more often contain uORFs (DD:60%, $P=2.6 \times 10^{-23}$; C:50%, $P=0.0002$; most intolerant LOEUF decile:46%, $P=2.4 \times 10^{-56}$). This is despite an overall depletion of upstream start codons (uAUGs). The uORFs of LoF intolerant genes also have stronger Kozak consensus than LoF tolerant gene uORFs (18.2% vs 14.2% "strong" kozaks) but weaker stop-codons (54.5% vs 48.9% weak stop codon (TGA codon).

Conclusion: Genes that are intolerant to LoF have more complex 5'UTRs and are enriched for cis-acting regulatory elements. The properties of uORFs in LoF intolerant genes point to them being efficiently translated and optimised for downstream re-initiation. These results confirm the important role of 5'UTRs in tight regulation of mRNA and protein levels, particularly for genes where changes in dosage are deleterious and lead to disease. A better understanding of the differences in UTR composition across genes will help us interpret how alterations to regulatory elements within them cause disease.

Molecular Effects of Genetic Variation Posters - Wednesday

PB2720. x-QTL: A mixture-model for identification and characterization of trans-eQTLs

Authors:

C. Wu¹, T. Xu¹, D. Munro^{1,2}, P. Mohammadi², A. Palmer¹, A. Goren¹, M. Gymrek¹; ¹Univ. of California, San Diego, La Jolla, CA, ²Scripps Res., La Jolla, CA

Abstract Body:

Studies of expression quantitative trait loci (eQTLs) aim to discover genetic variants that explain variation in gene expression levels, and may in turn drive associations with complex traits and diseases. While thousands of cis-eQTLs have been reliably identified, consistently replicating trans-eQTL effects proved to be challenging due to insufficient statistical power, and lack of comparable tissues and cohorts. In particular, technical covariates lead to substantial variation in expression datasets and result in false positive eQTL calls. As a result, biological mechanisms and characteristics of trans-eQTLs remain largely unknown.

Here, we introduce x-QTL, a novel trans-eQTL detection method that improves statistical power over methods that test gene-by-variant pairs separately by jointly modeling effects of an individual variant across all genes. The x-QTL package also implements previous trans-eQTL detection methods for comparison. x-QTL's model is similar to that of CPMA, an existing trans-eQTL detection technique, but uses a more biologically plausible mixture model of effect sizes. Importantly, our use of the mixture model infers the total number of target genes of a variant and demonstrates improved power over other tools. Additionally we developed a novel open-source simulation framework to benchmark performance of x-QTL. The simulation framework allows us to evaluate the suitability of datasets and the effects of trans-eQTLs that we can hope to detect.

We test x-QTL by applying it to a publicly available yeast expression dataset with 1,012 meiotic segregants. We identify 367 unique variants acting as trans-eQTLs at FDR 5% and replicate 3 hotspots of trans-eQTLs which x-QTL predicted to affect over half of the expressed genes. Next, we applied x-QTL to a dataset from RatGTEx that consists of five brain regions from 88 outbred rats for which a variety of behavioral and physiological traits have been measured. We identify novel trans-eQTLs across different tissues and observe an interesting trans-eQTL in NR3C1, a transcription factor that regulates genes controlling development, metabolism, and immune response. Finally, we applied x-QTL to expression data for 52 tissues in humans from the GTEx cohort. While this analysis did not return strong novel trans-eQTL candidates, our power analyses suggest existing human datasets are still substantially underpowered to detect even moderately strong trans-eQTLs. Overall, our framework provides important resources for future trans-eQTL studies involving humans and other complex organisms.

The x-QTL package and simulation framework can be found here:

<https://github.com/cynthiaewu/trans-eQTL>

Pharmacogenomics Posters - Wednesday

PB2721. A Pharmacogenetic Integration Solution for Clinical Decision Support Software

Authors:

H. Katzov-Eckert¹, M. Dawes^{1,2}, C. Verzosa¹, B. Esquivel^{1,3}; ¹GenXys Hlth.Care Systems, Vancouver, BC, Canada, ²Univ. of British Columbia, Vancouver, BC, Canada, ³Personalized Med. Latin American Association, Mexico City, Mexico

Abstract Body:

Pharmacogenetics has the potential to improve the effectiveness of medications and reduce the incidence of adverse drug events. One of the main barriers to implementing pharmacogenetics into practice is the lack of tools that simplify the use of pharmacogenetic results into the clinical workflow. In addition, providers are required to interpret the genotype data into phenotypes and clinically actionable information, something that many clinicians find challenging. Clinical decision support systems (CDSS) consolidate complex drug-gene interactions with biophysical patient data and provide safe and effective medication therapy options. We present a novel tool that facilitates the integration of pharmacogenetics to a CDSS. The web-based tool, LabGx offers a comprehensive solution for translating pharmacogenetic test results and embedding those results into a CDSS. The use of CDSS provides clinicians with real-time, multi-factorial medication optimization and personalization options for patients. LabGx was designed for clinical laboratories, irrespective of genotyping testing technology and supports SNP, SNV, CNV, and haplotype data. LabGx incorporates this information into a CDSS and interprets the content using evidence-based guidelines. Due to the complexity of genomic data, a Unified Lab Result (ULR) JSON file was developed as a standard for genotype data input. The genetic data converted into the ULR format streamlines the process of translation into the CDSS for laboratories. LabGx enables the use of pharmacogenetic information at point of care, potentially increasing adoption by healthcare providers, leading to a significant improvement in the population's quality of care and quality of life.

Pharmacogenomics Posters - Thursday

PB2722. A systematic review and comparison of statistical methods for testing genotype-treatment interaction effects in trans-ethnic pharmacogenomics GWAS.

Authors:

W. Zhong, J. Shen; Merck & Co., Inc., Rahway, NJ

Abstract Body:

Pharmacogenomics (PGx) genome-wide association studies (GWAS) aim to discover prognostic and predictive genetic biomarkers associated with drug responses. The typical approach of detecting predictive biomarkers in PGx GWAS analysis is to test the significance of a genotype-treatment interaction (GTI) term in a statistical model. Many statistical methods have been proposed to detect genotype-environment interaction (GEI) effects in disease genetics space, which can be directly applied to testing GTI effects in PGx GWAS. However, it is unclear whether these existing methods can accurately control type I error (or genome inflation) and achieve detecting power for GTI test in *trans-ethnic* PGx GWAS. We systematically reviewed and compared multiple statistical methods for testing GTI effects to find optimal tests and methods for *trans-ethnic* PGx GWAS analysis. Extensive simulations were conducted under various genetic architectures and scenarios to evaluate the genome inflation and power performance of the statistical methods including pooled-analysis approaches (i.e., linear regression-based methods and linear mixed model-based methods) and meta-analysis approaches (i.e., fixed effects, random effects, and meta-regression methods). We simulated different scenarios of distribution of variants' effect sizes, different strengths of genetic main and GTI effects, and different sample sizes across multiple ancestries. These simulation studies showed that meta-analysis-based methods controlled the genomic inflation better and had larger power than alternative methods in most scenarios. Mixed model-based methods demonstrated advantage over meta-analysis-based methods, if the sample size is limited in a subset of ancestries for GWAS analysis. Our systematic review and simulation-based comparison of existing statistical methods provide a clear guidance for predictive biomarker discovery in *trans-ethnic* PGx GWAS.

Pharmacogenomics Posters - Wednesday

PB2723. Accurate CYP2D6 star (*) allele diplotyping for long-read PacBio HiFi sequencing

Authors:

J. Harting¹, N. Gonzaludo¹, L. Zhu¹, X. Chen¹, A. Gaedigk², S. Scott³, P. Baybayan¹; ¹PacBio, Menlo Park, CA, ²Children s Mercy Kansas City, Washington, MO, ³Stanford Univ., Stanford, CA

Abstract Body:

The CYP2D6 gene is widely studied in the field of pharmacogenomics as it is directly involved in the metabolism of ~25% of the most commonly prescribed medications, including antidepressants, opioids, and cancer drugs. However, CYP2D6 is challenging to interrogate by short-read sequencing due to its highly homologous neighboring pseudogenes, and the presence of common structural variation, including complete gene deletions/duplications, tandem alleles, and gene conversions. Given these challenges, it is common to employ multiple technologies to comprehensively characterize the CYP2D6 locus; however, phasing variants and alleles still relies on inference and/or computational predictions. Recently, long-read HiFi amplicon sequencing has demonstrated that CYP2D6 can be fully captured, with direct phasing of haplotypes and characterization of specific duplicated alleles. However, a remaining challenge is the translation of identified variants and haplotypes into star (*) allele diplotypes, as the majority of bioinformatic translation tools were developed for short-read sequencing. Here we present a novel long-read CYP2D6 typer that accurately and unambiguously assigns star (*) allele diplotypes from PacBio HiFi reads. Briefly, long reads are sorted into unique alleles and variants from each are matched to star (*) allele definitions, as defined by PharmVar. Structural breakpoints are identified to characterize tandem duplications, deletions, and hybrid alleles. Phased alleles are then sorted into haplotypes for fully characterized diplotypes.

We demonstrate the use of this caller on Coriell reference samples that had undergone long-read PacBio HiFi amplicon, probe-based targeted enrichment, and/or genome sequencing. Taken together, these results support the use of this method for accurate and unambiguous star (*) allele assignment, enabling an end-to-end solution for CYP2D6 phenotype prediction and the implementation of long-read sequencing for comprehensive CYP2D6 characterization.

Pharmacogenomics Posters - Thursday

PB2724. Allelic diversity of the pharmacogene *CYP2D6* in New Zealand Māori and Pacific peoples.

Authors:

M. Kennedy¹, L. Hitchman¹, A. Faatoese¹, T. Merriman², A. L. Miller¹, Y. Liao¹, O. Graham¹, K. Ping Sui¹, J. F. Pearson¹, T. Fakahau³, V. A. Cameron¹, S. D. S. Maggo¹; ¹Univ. of Otago, Christchurch, New Zealand, ²Univ. of Otago, Dunedin, New Zealand, ³Pacific Trust Canterbury, Christchurch, New Zealand

Abstract Body:

The enzyme cytochrome P450 2D6 (*CYP2D6*) metabolises approximately 25% of commonly prescribed drugs, including analgesics, anti-hypertensives, and anti-depressants, among many others. Genetic variation in drug metabolising genes can alter how an individual responds to prescribed drugs, including predisposing to adverse drug reactions. The majority of research on the *CYP2D6* gene has been carried out in European and East Asian populations, with indigenous and minority populations greatly underrepresented. However, genetic variation is often population specific and analysis of diverse ethnic groups can reveal differences in alleles that may be of clinical significance. For this reason, we set out to examine the range and frequency of *CYP2D6* variants in a sample of 202 Māori and Pacific people living in Aotearoa (New Zealand). We carried out long PCR to isolate the *CYP2D6* region, and then carried out nanopore sequencing to identify all variants and alleles in these samples. We identified 11 novel variants, one of which was an exonic missense mutation. Six of these occurred in single samples (<0.5% each) and one was found in 11 samples (5.4%). In addition, five new suballeles of *CYP2D6* were identified. One striking finding was that *CYP2D6**71, an allele of unknown functional status which has been rarely observed in previous studies, occurs at a relatively high frequency (9.2%) within this cohort. These data will help to ensure that *CYP2D6* genetic analysis for pharmacogenetic purposes can be carried out accurately and effectively in this population group.

Pharmacogenomics Posters - Wednesday

PB2725. An NGS approach to detecting novel LOF variants in *DPYD* across >69,000 individuals

Authors:

G-h. Kim, A. T. Chande, H. E. Williams, J. Van Den Akker, C. L. Neben, A. Y. Zhou; Color Hlth., Burlingame, CA

Abstract Body:

Background: The metabolism of chemotherapeutic drugs such as 5-fluorouracil (5FU), capecitabine, and tegafur are known to be affected by variants within *DPYD*. With improper dosing, these medications can lead to decreased treatment efficacy and potentially life-threatening toxicity, such as bone marrow suppression. As such, pharmacogenomic (PGx) testing in these patients is a crucial step to determine the correct dosing regimen.

Genotyping arrays are a common approach to clinical PGx testing, but use of this technology limits analysis to common, known variants and leaves rare, novel variants undetected. Furthermore, variants in *DPYD* have been shown to vary across genetic ancestry groups. In contrast, NGS allows for the detection of both known and novel variants and is not subject to the inherent ascertainment bias of genotyping arrays developed from individuals of European descent. Here, we analyzed NGS data from 69,226 individuals who had clinical PGx and identified novel potential loss of function (pLOF) variants in *DPYD*.

Methods: All individuals consented to have their de-identified information and sample used in anonymized studies. Variant data was interrogated for variants within *DPYD* and unique variants were normalized and subsequently annotated using Variant Effect Predictor. pLOF variants were defined as variants predicted to be deleterious with consequences labeled missense, start loss, stop gain, frameshift, or splice donor/acceptor variants.

Results: Approximately 1.6 million variants (2,765 unique variants) were detected among 69,226 individuals within the *DPYD* expanded genomic region, the majority of individuals were white (n=55,456, 79.4%, self-reported race/ethnicity). Of the 2,765 unique variants detected, 385 (13.9%) variants were classified as pLOF in 9,075 (13.1%) individuals. The vast majority of individuals (n=8794, 97.0%) were found to have one pLOF variant while one individual showed 10 pLOF variants. Across the entire cohort, we observed start loss (n=9, 2.33%), stop gain (n=4, 1.04%), splice donor/acceptor (n=33, 8.57%) and missense (n=321, 83.4%) variants, as well as variants with multiple predicted consequences (n=28, 7.27%). Self-reported ethnicity showed the majority of pLOF variants were found in white individuals, and a small number (n=17, 4.4%) were enriched in Asian and African individuals.

Conclusion: These data demonstrate that NGS is a useful tool to detect novel variants across a diverse population of individuals. Further investigation is needed to determine the functional consequences of these *DPYD* pLOF variants to avoid potentially life-threatening adverse effects and improve efficacy of treatment.

Pharmacogenomics Posters - Thursday

PB2726. Clinical genetic testing and reporting of homoplasmy or low-level heteroplasmy for *MT-RNR1* m.1555A>G conferring an increased risk for aminoglycoside induced ototoxicity and irreversible hearing loss.

Authors:

N. Cody¹, R. Zimmerman¹, J. Shaw¹, R. Rigobello¹, D. Ilg¹, S. Schnakenberg¹, M. Ghoname¹, R. Kadakia¹, F. Chaabani², C. Neiss², B. Ramey¹, C. Buchovecky¹, R. Kornrieck¹, S. A. Scott³, L. Edelmann¹; ¹Sema4, Stamford, CT, ²Illumina, San Diego, CA, ³Stanford Univ., Stanford, CA

Abstract Body:

Aminoglycoside antibiotic exposure may result in ototoxicity and irreversible hearing loss for patients harboring pathogenic variants in the mitochondrial *MT-RNR1* gene. Clinical Pharmacogenetics Implementation Consortium guidelines provide evidence and therapeutic recommendations for avoidance of aminoglycosides for *MT-RNR1* m.1095T>C, m.1494C>T, and m.1555A>G. Sema4 offers molecular supplemental newborn screening panels that include a pharmacogenetics (PGx) test containing 13 PGx gene/variants including *MT-RNR1* m.1555A>G. These panels are available for asymptomatic/presymptomatic children under the age of 10. Testing was performed on DNA extracted from buccal swab specimens. The m.1555A>G region was sequenced with >15000X coverage using Agilent SureSelect capture-based sequencing on the Illumina NovaSeq 6000 platform. Heteroplasmy was estimated as a percentage of reads harboring the m.1555A>G alteration over the total. Targeted genotyping was performed using PCR and single-base extension with Agena SpectroCHIP on a MassArray Analyzer, which was verified for performance of detecting homoplasmy and heteroplasmy >50%. Two of 1440 patients tested from 2019 to 2022 from two independent families were positive for m.1555A>G. One patient was considered homoplasmic (100%, 15000X) and the other patient had an estimated heteroplasmy of ~12% (2238/20651X). Reports for NGS results contained recommendations to avoid aminoglycosides, obtain genetic testing to confirm maternal inheritance, and perform additional testing for at-risk maternal relatives. A supplemental report using third-party software provided statements on impacted medications and avoidance of aminoglycosides. Impacted medication statements were not available on the PGx report if heteroplasmy was below 50%. In this instance, comments were included to clarify heteroplasmy may be variable across different tissues, and recommended avoidance of aminoglycosides. Report distribution followed with critical alert notifications to the healthcare provider and recommendations for genetic counselling support. Heteroplasmy, tissue specificity, and temporal changes in mutation burden create challenges for clinical testing of *MT-RNR1* m.1555A>G. Parallel testing and reporting positive findings was crucial when encountering low levels of heteroplasmy. Future validation of an Illumina microarray platform enhanced for PGx will enable the integration of two additional *MT-RNR1* variants to improve the clinical utility of this test. Supplementing clinical reports with discrete data and clinical decision support tools to assist healthcare providers is warranted.

Pharmacogenomics Posters - Wednesday

PB2727. Convergence of bipolar disorder treatments and gene knockdown on the transcriptome

Authors:

V. Vuokila^{1,2}, C-E. Castonguay^{1,2}, N. Farhangdoost¹, A. Khayachi¹, C. Liao^{3,4}, P. Dion¹, G. Rouleau^{1,2}; ¹Montreal Neurological Inst., McGill Univ., Montreal, QC, Canada, ²Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, ³Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

Common medications used for the acute treatment of bipolar disorder (BD) include mood stabilizers and antipsychotics, though how the drug mechanisms relate to BD disease etiology is poorly understood. We have conducted an analysis into the convergent transcriptomic effects of seven drugs commonly used to treat BD (Lamotrigine, carbamazepine, sodium valproate, chlorpromazine, clozapine, risperidone, and ziprasidone) with publicly available data from the LINCS L1000 perturbation database. Using Stouffer's Z-score method for convergent transcriptomic drug targets ($p_{val} < 0.05$) and the gProfiler2 R package for pathway enrichment analysis, we found that pathways related to extracellular exosomes ($q_{val} = 4.19 \times 10^{-8}$), extracellular vesicles ($q_{val} = 6.12 \times 10^{-8}$) and extracellular organelles ($q_{val} = 6.23 \times 10^{-8}$) were among the most significantly upregulated pathways by these drugs in wild-type neurons. Exosomes are small extracellular vesicles that carry molecular signals and are able to cross the blood-brain barrier, making them relevant in the context of psychiatric disorders. Although this transcriptomic screen has shed light on potential mechanisms through which drugs might reduce symptoms, it does not take into account the disease contexts in which pharmacological agents act. Therefore, we sought to model the effects of these BD drugs in a system that approximates BD biology. Genes that have a downregulatory association with BD include FURIN and MMD (Mullins et al. 2021). We aim to characterize the convergent transcriptomic signature of FURIN and MMD downregulation through CRISPRi in iPSC-derived excitatory cortical neurons. We demonstrate the knockdown effect of FURIN and MMD in their corresponding iPSC lines through TaqMan RT-qPCR ($RQ = 0.526$ and $RQ = 0.266$, respectively). We will differentiate our knockdown cell lines into neurons and assay the transcriptome through bulk RNA-sequencing in untreated and BD drug-treated knockdown neurons. We hypothesize that the knockdown of the FURIN and MMD will have common, convergent downregulatory effects in extracellular exosome, vesicle, and organelle pathways. Further, we hypothesize that treatment with BD drugs will ameliorate this downregulation. These results would highlight the importance of extracellular pathways in BD and further elucidate the genetic etiology of BD.

Pharmacogenomics Posters - Thursday

PB2728. CRISPR enrichment of complex *CYP2D6-D7-D8* loci allows for accurate detection of *CYP2D6* structural variation and phased haplotyping.

Authors:

A. Turner¹, A. D. Dereziński¹, A. Gaedigk², M. E. Berres³, D. Gregornik⁴, U. Broeckel¹, G. Scharer¹; ¹RPRD Diagnostics, Milwaukee, WI, ²Children s Mercy Kansas City, Washington, MO, ³Univ. of Wisconsin, Madison, WI, ⁴Children s Minnesota, Minneapolis, MN

Abstract Body:

The highly polymorphic *CYP2D6* gene is one of the most clinically tested pharmacogenetic genes contributing to the metabolism of >20% of common clinical drugs. A large portion of enzyme variability can be explained by single nucleotide polymorphisms (SNPs), as well as structural variants (SVs) such as copy number variation (CNV) and hybrid rearrangements with the highly similar *CYP2D7* pseudogene. CNVs/SVs are known to occur in different configurations and frequencies across populations. Simple SVs can be reliably detected using current methods, however clinically relevant complex SVs, especially those harboring *CYP2D6* and *CYP2D7*-derived hybrid sequences and/or multiple gene copies, are notoriously difficult to detect and characterize accurately. This can lead to incorrect enzyme activity assignment and impact drug dosing recommendations, particularly in underrepresented populations. To overcome *CYP2D6* genotyping limitations we developed a method utilizing *CYP2D6-CYP2D7-CYP2D8* locus targeted guide RNAs (gRNAs) along with CRISPR-Cas9 technology and PCR-free single molecule long-read sequencing to accurately determine diplotypes (phased haplotypes) regardless of SV complexity. This novel methodology was evaluated using previously characterized samples with a variety of SVs including complex *CYP2D6* CNVs and *CYP2D6/CYP2D7* hybrids, and further tested with high-molecular weight DNA obtained from multiple sample types including lymphoblastoid cell lines and blood (n=10). PCR-free single molecule libraries were prepared using targeted gRNAs and Cas9 Sequencing Kit, then sequenced using Oxford Nanopore MinION flow cells. Using standard and custom-build data analysis programs, data was base-called, filtered, aligned, sorted, indexed and coverage assessed and visualized over a ~38Kb region covering *CYP2D6*, *CYP2D7*, and *CYP2D8*. Samples with *CYP2D6* or hybrid copy number gains were enriched for reads up to 52kb. Analysis of reads >25kb with PHRED-scaled Qscores averaging >14 showed significant sequence enrichment of the targeted region (average read depth >150x). Our novel methodology accurately determined SNP level genotypes and complex diplotypes with SVs such as *CYP2D6**2/*36+*10 and *1/*13+*2, unconstrained by amplicon size or sample structural complexity. For the first time we enriched and analyzed the entire *CYP2D6-2D7-2D8* locus using PCR-free single molecule long-read sequencing. This allows fully phased dissection of the locus haplotype structure, including breakpoints, which offers unique advantages for clinical genotyping and phenotype prediction, overcoming the limitations of current testing platforms.

Pharmacogenomics Posters - Wednesday

PB2729. DEEPCT(Deep-learning-based Clinical Trial) simulates a drug clinical trial by exploiting germline variants as random assignment of drug.

Authors:

J. Kim^{1,2}, S. Lee^{1,3}, H. Kim¹, Y. SANGHYCK^{1,4}, E. Lee¹, D. KIM¹, N. Kim^{1,4}, E. Kim¹, J-H. Lee¹, S. Ko¹, S. Lee¹, S. Lee¹, W. Chung²; ¹Basgenbio, Seoul, Korea, Republic of, ²Dept. of Statistics and Actuarial Sci. Soongsil Univ., Seoul, Korea, Republic of, ³Dept. of Hlth.Policy, Korea Univ. Graduate Sch., Seoul, Korea, Republic of, ⁴Dept. of Epidemiology and Hlth.Promotion, Inst. for Hlth.Promotion, Graduate Sch. of Publ. Hlth., Yonsei Univ., Seoul, Korea, Republic of

Abstract Body:

Clinical trials are essential for evaluating the efficacy and safety of novel drugs and medical devices. However, clinical trials are costly, time-consuming, inappropriate for interventions with ethical issues, and have potential to pose unknown risks to the participants. In silico simulations of clinical trials prior to actual clinical trials may address the aforementioned challenges and improve the success rate of the actual trials. Using the data of 488,221 participants from the UK Biobank, including 20,437 participants who had gout events, we simulated the drug effect of xanthine dehydrogenase (XDH) inhibitors (e.g., allupurinol, febuxostat), which lowers serum uric acid and treat gout. The simulation algorithm consisted of deep learning approaches to a genome-wise association study (GWAS) and a genetic score calculation, followed by a mendelian randomization analysis. We refer to the series of the in silico methods as DEEPCT (Deep learning Clinical Trial).

We identified urate-level-lowering 10 variants within the XDH gene using a deep reinforcement learning algorithm. To calculate of the genetic score, the number of selected alleles was weighted by the GWAS effect of each variant on urate levels.

The XDH genetic score for lowering serum urate level by 0.1 mg was associated with a decrease of 14% in the risk of gout events (OR [95% CI] = 0.86 [0.77, 0.91]; $p < 0.001$). In addition, the XDH score for lowering serum urate level by 1 mg was associated with a reduced incidence rate of cardiovascular disease by 31% (OR [95% CI] = 0.69 [0.49, 0.96]; $p = 0.029$).

The drug effect estimated using the DEEPCT were in parallel with the drug effect of commercial XDH inhibitor drugs, which have been confirmed by actual clinical trials. The DEEPCT may be utilized prior to actual clinical trials to predict the effects of novel drugs and therapies.

Pharmacogenomics Posters - Thursday

PB2730*. Dynamics in blood transcriptomic networks responding to TNF inhibitors in patients with rheumatoid arthritis

Authors:

C-Y. Yu¹, H-S. Lee², Y. Joo², S-K. Cho², C-B. Choi², Y-K. Sung², T-H. Kim², J-B. Jun², D. Yoo², S-C. Bae², K. Kim¹, S. BANG²; ¹Kyung Hee Univ., Dondaemungu, Korea, Republic of, ²Hanyang Univ. Hosp. for Rheumatic Diseases, Seoul, Korea, Republic of

Abstract Body:

Although TNF inhibitors have widely been used for the effective treatment for rheumatoid arthritis (RA), one-third of patients do not respond well, subsequently showing poor prognosis. Our study aimed to uncover the main transcriptomic dynamics under anti-TNF treatments and identify biomarkers to stratify potentially well-responding patients. RNA-seq data from peripheral blood cells and clinical disease activity indices of patients with RA in a Korean cohort (n=62) and a Caucasian cohort dataset (Corrona CERTAIN, GSE129705; n=20) were obtained before and after treatment of TNF inhibitors. Response-associated genes were defined by estimating the differential expression changes after treatment between good and null responders. We found 458 response-associated genes that were significantly enriched in innate immunity and interferon response-related signaling pathways in the gene-level gene set enrichment analysis (GSEA). All longitudinal expression data from the two cohorts were used to build coexpression networks of which response-associated modules were subjected to a Bayesian key driver analysis to identify driver genes responsible in regulation of network-level expression changes upon treatment. A coexpression network analysis followed by a network-level GSEA identified 16 coexpression modules in the blood transcriptome. Of them, two modules were associated with innate immunity and interferon response-related signaling pathways, including significantly more response-associated genes, consistent with the gene-level GSEA. In the two response-associated coexpression modules, we identified 37 module-specific response-associated key drivers whose down-regulations were correlated with the improvement of response upon treatment in both the cohorts. Furthermore, the treatment-induced expression changes of key drivers in the interferon signaling networks were significantly associated with the change of erythrocyte sedimentation rate in moderate responders. In summary, this study provides the landscape of response-associated transcriptomic changes in peripheral blood cells from RA patients emphasizing the importance of innate immunity-related signatures in response prediction of anti-TNF biologics.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2022R1A2C2006073).

Pharmacogenomics Posters - Wednesday

PB2731. Elucidating the multi-omic drivers of hepatic drug metabolism in African Americans

Authors:

C. Clark, T. De, G. Yang, C. Alarcon, J. M. Avitia, M. Perera; Northwestern Univ., Chicago, IL

Abstract Body:

Background: Prediction of drug response in African Americans (AAs) using genetic variants discovered in other populations has been poor. Drug metabolizing enzymes (DMEs), located primarily in hepatocytes, metabolize most currently prescribed drugs. Phase I (i.e., cytochrome P450s, CYP450s) and Phase II (i.e., glucuronosyltransferases, UGTs) DMEs catalyze reactions that make drugs easier to excrete. Variations within these DME genes, their regulatory regions, or within regulatory proteins can contribute to interindividual variability of drug response and adverse drug events. This study aims to identify genetic variation and genes that associate with differences in drug metabolism using an AA cohort. **Method:** Hepatocytes were extracted from 75 AA cadaveric livers and subjected to a panel of validated probe substrates for DMEs (Midazolam, Bupropion, Mephenytoin, Diclofenac, Phenacetin, Acetaminophen) to measure metabolite formation rate (MFR). Hepatocytes underwent genome-wide genotyping, mRNA sequencing, and DNA methylation profiling. MFR was collected via LC-MS/MS, thus linking the donor's genetic, transcriptomic, and epigenetic data. A genome-wide association and differential gene expression study were conducted for the MFR of each probe substrate with age, sex, and principal components as covariates. **Results:** Wide variability in all MFRs were seen between individuals, reflecting the known variation in AA drug metabolism. Genome-wide significant associations were identified at the ANGPT2 locus ($p=1.8 \times 10^{-8}$). The top ANGPT2 SNP (rs2515447) was associated with decreased MFR. Loci reaching suggested level were also identified, including CYP2C (Mephenytoin, $p=4.1 \times 10^{-7}$), CYP4X1 (Bupropion, $p=7.5 \times 10^{-6}$), SEPT9 (Midazolam, $p=9.7 \times 10^{-7}$). Transcriptomic, epigenetic and fine-mapping analyses are ongoing. **Conclusion:** This first-of-its-kind study of MFR found genes previously linked to liver function or DMEs specific to African ancestry populations. ANGPT2 is a growth factor implicated in liver cancer prognosis and inflammatory pathways linked to CYP regulation. Both CYP2C and CYP4X1 enzymes have known roles in the metabolism of probe substrates. Importantly, CYP2C genetic variation has repeatedly been associated with clopidogrel (antiplatelet drug) and warfarin (anticoagulant) response. SEPT9 methylation is a clinical biomarker of hepatocellular carcinoma and is associated with liver fibrogenesis. These results provide needed African-ancestry pharmacogenomic data that will be critical for individualizing drug dosing and drug-type selection in AAs and in understanding the regulation of these important enzymes.

Pharmacogenomics Posters - Thursday

PB2732*. Estimating UK Biobank population-specific PGx allele and phenotype frequencies using PharmCAT

Authors:

B. Li¹, K. Sangkuhl¹, R. Whaley¹, M. Woon¹, K. Keat², R. Altman³, M. Carrillo¹, M. Ritchie⁴, T. Klein¹; ¹Stanford Univ., Stanford, CA, ²Univ. of Pennsylvania, Philadelphia, PA, ³Stanford Univ., Menlo Park, CA, ⁴Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Pharmacogenomics (PGx) investigates the genetic influence on drug response. It is an integral part of precision medicine and contributes to maximization of drug efficacy and reduction of adverse drug event risk. From a research perspective, it is important to understand PGx allele and phenotype distributions in different biogeographic groups. This knowledge provides guidance for future PGx studies and clinical genetic test panel design to serve individuals of wider biogeographic or genetic backgrounds. Nonetheless, to date, for many genes with published PGx-informed clinical drug prescribing guidelines, frequency information for some alleles is still missing, rarely reported, or seldomly tested due to various reasons. To address the issue, we applied the Pharmacogenomics Clinical Annotation Tool (PharmCAT) on an integrated UK Biobank genetic data set (N = 200,044), which incorporated quality-controlled whole exome sequencing, imputed, and microarray-based genotypes data to achieve a high-quality, comprehensive coverage of PGx positions. We mapped the self-reported ethnicities to five of the nine PharmGKB biogeographic groups: European (N = 187,660, 93.81%), Central/South Asian (N = 3,460, 1.73%), East Asian (N = 637, 0.32%), African American/Afro-Caribbean (N = 637, 0.32%), and Sub-Saharan African (N = 1,235, 0.62%); and an “other” group (N = 5,126, 2.56%) with participants who could not be classified into a single biogeographic group. For over 200K UK Biobank participants, PharmCAT reported alleles and phenotypes for 16 pharmacogenes (*ABCG2*, *CACNA1S*, *CFTR*, *CYP2B6*, *CYP2C19*, *CYP2C9*, *CYP3A5*, *CYP4F2*, *DPYD*, *IFNL3*, *NUDT15*, *RYR1*, *SLCO1B1*, *TPMT*, *UGT1A1*, and *VKORC1*) in five biogeographic groups. We observed frequency differences among biogeographic groups. E.g., for *CYP219*, which has guidelines for clopidogrel, proton pump inhibitors, antidepressants, etc., the percentage of poor metabolizers with clinical actionable recommendation was five-fold higher in the Central/South Asians and East Asians (10.98% and 11.77%, respectively) than in the Europeans (2.47%). We reported frequencies for no-function alleles that were rare or seldom tested in certain groups by previous studies. E.g., 19 (0.27%) and 570 (0.15%) *CYP2C19**35 in the Central/South Asians and Europeans, respectively, and 146 (3.79%) *SLCO1B1**31 in the African American/Afro-Caribbeans. This systematic estimate of PGx frequencies, which will be shared on the Pharmacogenomics Knowledge Base (PharmGKB), aims to bring awareness and crucial information to improve future PGx studies and clinical genetic test panels, particularly for underrepresented populations in genomics.

Pharmacogenomics Posters - Wednesday

PB2733. Evaluating the frequency and the impact of pharmacogenetic variants in an ancestrally diverse Biobank population

Authors:

K. Keat¹, S. Verma², B. Li³, G. Hoffecker⁴, M. Risman⁵, Regeneron Genetics Center, K. Sankuhl³, M. Whirl-Carrillo³, S. Dudek⁵, A. Verma⁴, T. E. Klein⁶, M. D. Ritchie⁵, S. Tuteja⁴; ¹Genomics and Computational Biology PhD Program, Univ. of Pennsylvania, Philadelphia, PA, ²Dept. of Pathology & Lab. Med., Univ. of Pennsylvania, Philadelphia, PA, ³Dept. of BioMed. Data Sci., Stanford Univ., Stanford, CA, ⁴Dept. of Med., Univ. of Pennsylvania, Philadelphia, PA, ⁵Dept. of Genetics, Univ. of Pennsylvania, Philadelphia, PA, ⁶Dept. of BioMed. Data Sci. and Med. (BMIR), Stanford Univ., Stanford, CA

Abstract Body:

Pharmacogenetics (PGx) aims to utilize a patient's genetic data to enable safer and more effective prescribing of medication. The Clinical Pharmacogenetics Implementation Consortium (CPIC) provides guidelines for 24 genes that affect 72 medications. Despite strong evidence linking PGx variants to drug response, there is a large gap in the actual implementation and return of actionable pharmacogenomic findings to patients in standard clinical practice. In this study, **we evaluated opportunities for genetically guided medication prescribing in a diverse health system** and determined the frequencies of actionable PGx alleles in an ancestrally diverse biobank population. A retrospective data mining of Penn Medicine electronic health record (EHR) data, which includes ~3.3 million patients between 2012-2020 provides a snapshot of the distribution of highly ranked CPIC drugs in the Penn Medicine health system. Additionally, the Penn Medicine BioBank (PMBB) consists of a diverse group of 43,359 participants whose EHR data is linked to genome-wide genotype data and whole exome sequencing. We used the Pharmacogenomics Clinical Annotation Tool (PharmCAT), which annotates pharmacogenetic alleles in a variant call format (VCF) file to identify samples with actionable PGx phenotypes. We identified ~316,000 unique patients that were prescribed at least 2 CPIC Level A or B drugs. Genetic analysis of samples from PMBB was used to measure PGx phenotype frequencies and clinical burden. All genotyped participants had at least one non-reference allele in a CPIC gene. PharmCAT annotation also identified that **98% of participants are carriers of one or more PGx actionable phenotypes**, indicating a modification in drug therapy would be recommended. We linked genetic data with prescription data in the EHR and identified that **13.3% of participants (n=5785) were prescribed medications impacted by their PGx alleles**. We found 849 participants who received clopidogrel with CYP2C19 intermediate or poor metabolizer phenotypes who were at increased risk for major adverse cardiovascular events. When we stratified by genetic ancestry, we found disparities in PGx allele frequencies and clinical burden. Notably, **clopidogrel users of Asian ancestry in PMBB had significantly higher rates of CYP2C19 intermediate or poor metabolizer phenotypes than European ancestry users** ($p < 0.0001$, $OR = 3.72$). This analysis provides an overview of the distribution of PGx associated genetic alleles and the prescription rates of medications impacted by PGx variants, illustrating the potential utility of pre-emptive genotyping for PGx and implementation in clinical care.

Pharmacogenomics Posters - Thursday

PB2734. Exploring polygenic risk scores as a biomarker of treatment response in systemic lupus erythematosus

Authors:

J. Esparza Gordillo¹, E. Arciero², D. Fraser³, L. McCarthy², S. Gowrisankar³, M. Chiano², C. Cox²; ¹GlaxoSmithKline, Genomic Sci., Tres Cantos, Madrid, Spain, ²GlaxoSmithKline, Genomic Sci., Stevenage, United Kingdom, ³Parexel, Durham, NC

Abstract Body:

Polygenic risk scores (PRS) enable the identification of individuals who are at high risk of disease and may help to improve clinical decision-making. We tested the utility of a PRS biomarker for predicting treatment response in patients with systemic lupus erythematosus (SLE). First, we selected 117 instruments from a published genome-wide association study for SLE and constructed a “SLE risk PRS” on the UK Biobank, where we observed a 30-fold higher SLE prevalence in individuals within the top versus the bottom percentile of the PRS distribution. Next, we tested for correlation between SLE PRS and response to belimumab in Phase 3 SLE clinical trials (N=755 individuals). Given that belimumab provides clinical benefit by blocking the B-cell activating factor (BAFF; also known as B-lymphocyte stimulator, BLyS), we hypothesized that individuals with “B-cell driven disease” are more likely to show clinical benefit. Thus, we implemented an additional “B-cell specific PRS” that used bioinformatic approaches, such as eQTL, gene expression, and overlap of credible set single nucleotide polymorphisms with B-cell enhancer marks, to select SLE risk variants potentially affected through impact on B cells. Both the “SLE PRS” and “B-cell specific SLE PRS” correlated with SLE severity at baseline (based on SELENA-SLEDAI scores) but did not independently predict treatment response (positive correlation between PRS quartile and “high dose” prednisone use at baseline OR [95% confidence interval (CI)] 1.19 [1.05, 1.35], p=0.005 and no correlation with SRI4 medicine response OR [95% CI] 1.00 [0.86, 1.18], p=0.96). Based on these results and on previous experience in the field we will discuss potential challenges and opportunities for the use of PRS on clinical decision-making.

Acknowledgments: Funding: GlaxoSmithKline (GSK study BEL218481). Editorial assistance (GSK-funded): Gosia Carless, PhD, Fishawack Indicia Ltd. UK, part of Fishawack Health.

Pharmacogenomics Posters - Wednesday

PB2735. Genetic associations of iron absorption among individuals of East Asian and European descent.

Authors:

A. Barad¹, Y. Wang¹, Y. Xu¹, E. Bender¹, S. Bird², Z. Gu¹, A. G. Clark^{3,4}, K. O. O'Brien¹; ¹Div. of Nutritional Sci., Cornell Univ., Ithaca, NY, ²Univ. of Rochester Med. Ctr., Rochester, NY, ³Dept. of Molecular Biology and Genetics, Cornell Univ., Ithaca, NY, ⁴Dept. of Computational Biology, Cornell Univ., Ithaca, NY

Abstract Body:

Iron (Fe) balance in humans is maintained by tight regulation of dietary non-heme Fe absorption. Although several genetic variants have been shown to be associated with biomarkers of Fe status, the associations between these genetic variants and non-heme Fe absorption remain unknown. Previously we showed an Asian-prevalent haplotype had a suggestive association with increased Fe absorption, which may help explain the higher body Fe stores in East Asians compared to Europeans. Our aim is to assess the genetic contribution of a functional measure of Fe homeostasis (*i.e.*, Fe absorption) to observed variability in Fe status in East Asian and European adults. We are conducting a nutrigenomic study (FeGenes; ClinicalTrials.gov ID: NCT04198545) in adults (18-50 y) of genetically confirmed East Asian (N=252) or European (N=252) ancestry. We are collecting data on Fe biomarkers and regulatory hormones, quantifying Fe absorption using a stable Fe isotope, and measuring genome-wide genetic variants with Illumina's Global Diversity Array. We carried out a preliminary association study of all previously reported Fe-associated SNPs and polygenic risk scores of enhanced Fe status (PRS-Fe) in Europeans and related this to Fe absorption in FeGenes participants. Consistent with published data, serum ferritin levels were higher in East Asian (N=87) compared to European (N=97) adults ($P=0.003$), and East Asian adults (N=30) absorbed more Fe ($P=0.008$) compared to European adults (N=27) for a fixed amount of storage Fe (ferritin). We replicated published results showing an association between *TMPRSS6*-rs855791 and Fe absorption with consistent effects in both ancestry groups (meta: $\beta=0.37$, $P=0.04$, N= 57); the Asian-prevalent haplotype in *HFE* was only associated with Fe absorption in East Asian adults ($\beta=0.70$, $P=0.03$, N=30). With the current sample size (N=57), we did not observe significant associations between the low-frequency *HFE* variants (H63D and C282Y) and Fe absorption. Of the variants tested, we found three additional SNPs in *VANGL1* (meta: $\beta=-0.53$, $P=0.03$), *NAT2* (meta: $\beta=0.47$, $P=0.03$), and *ARNTL* (meta: $\beta=-0.50$, $P=0.04$), respectively, to be associated with Fe absorption. Moreover, we computed PRS-Fe based on results from GWAS in Europeans and found significant positive associations between PRS-Fe and Fe absorption (meta: $P=0.03$). In sum, we found that known Fe-associated SNPs are associated with non-heme Fe absorption. On-going genetic association studies of Fe absorption in FeGenes may reveal new genes involved in the regulation of Fe homeostasis in humans and may provide an explanation to the observed population differences in risk of Fe overload or deficiency.

Pharmacogenomics Posters - Wednesday

PB2736. Genomic analysis of hydroxychloroquine-associated retinal toxicity in a large cohort

Authors:

E. Ullah¹, M. Ramachandran², D. M. McGaughey¹, A. E. Turriff³, P. A. Sieving⁴, R. B. Hufnagel¹, C. A. Cukras¹; ¹Natl. Eye Inst., NIH, Bethesda, MD, ²Univ. of Missouri-Columbia, Columbia, Columbia, MO, ³Natl. Human Genome Res. Inst., NIH, Bethesda, MD, ⁴UC Davis Eye Ctr., Sacramento, CA

Abstract Body:

Hydroxychloroquine (HCQ) is used as a first-line agent to treat autoimmune diseases including rheumatoid arthritis, systemic lupus erythematosus, juvenile idiopathic arthritis and Sjogren's syndrome. Chronic long-term use of HCQ has been associated with a risk of irreversible damage to photosensitive retinal layers leading to vision loss. DNA sequence variants in *ABCA4* gene has been investigated in multiple studies with varying results finding an association with risk of toxicity in some studies and protective effects in others. To better understand the potential association of genetic elements in a cohort of cases with HCQ-associated retinal toxicity, 50 cases and 81 controls with and without retinopathy following 5 years of hydroxychloroquine use were enrolled following comprehensive battery of clinical tests. Participant DNA was genotyped by using Illumina Infinium OmniExpressExome SNP microarray (960,919 SNPs), followed by data processing on GenomeStudio v2.0 and genome-wide association analysis (GWAS) by PLINK 1.9. NGS-based panel (*ABCA4*, *BEST1*, *CDH3*, *DRAM2*, *EFEMP1*, *ELOVL4*, *IMPG1*, *IMPG2*, *PROM1*, *RDS*, *RP111*, *TIMP3*, *TLL5*), exome and genome sequencing were also performed on genomic DNA samples. Sequence data was processed by in-house developed customized pipeline for alignment, quality control, variant calling and variant annotation. Rare variant association testing was done by c-alpha test, Fisher's exact test, and GWAS was performed by PLINK 1.9. In this multimodal genomic analysis, *ABCA4* variants were not observed to increase or decrease risk for HCQ-associated retinal toxicity following correction for variants with different frequency by ancestry. NGS-based panel of maculopathy genes and exome sequence also did not reveal any significant association. Using ancestrally-neutral SNPs from genome data, genome-wide association analysis of common variants did reveal a locus on chromosome 1, nearby loci for systemic lupus erythematosus and age-related macular degeneration, which may be associated with conferring risk towards HCQ-associated retinal toxicity. Further analyses are underway. This is the largest cohort reported for HCQ-based retinal toxicity, in which the genomic analysis noted no associated effect of *ABCA4* variants on affected status as previously reported. Furthermore, Genome sequencing followed by association analysis identified a locus on chromosome 1 which is being further investigated for its association with HCQ-based retinal toxicity.

Pharmacogenomics Posters - Thursday

PB2737. Impact of global and local ancestry on pharmacogenetics in African American persons with multiple sclerosis.

Authors:

M. Davis¹, L. C. Silva¹, S. W. Brugger²; ¹Brigham Young Univ., Provo, UT, ²Brigham Young Univ., Orem, UT

Abstract Body:

Multiple sclerosis (MS) is a disease of the central nervous system treated with disease-modifying therapies, including the biologic, interferon- β (IFN- β). Up to 60% of IFN- β -exposed MS patients develop abnormal biochemical liver test results, and 1 in 50 experiences drug-induced liver injury. SNP rs2205986, linked to differential expression of *IRF6*, has demonstrated a statistical association with drug-induced liver injury in patients of European ancestry. Studies have revealed a differential risk for MS development among African American individuals compared to those of primarily European ancestry. Additionally, the penetrance of a given MS risk allele near the centromere on chromosome 1 can differ significantly depending on whether the allele is European or African in origin. We aim to evaluate the association of rs2205986 with IFN- β -induced liver injury in African American MS patients in the context of both global and local ancestry. We acquired de-identified electronic health records (EHRs) for 70 African American MS patients from Vanderbilt University Medical Center BioVU. Patient genomes were genotyped on the Illumina MEGAex platform. Utilizing the ADMIXTURE software with a k value of 2, we ran a supervised ADMIXTURE analysis using the CEPH and YRI datasets from the 1KGP as European and African controls, respectively. The majority of samples had less than 50% African ancestry, with a wide distribution. Using RFMix, we will evaluate local ancestry across the genomes of our African American samples. We will analyze which regions of the genome have a local ancestry that differs significantly from the genome average. Significant signals will be analyzed for risk according to ancestry. By studying the genetic admixture of these patients, we aim to identify if differential risk of IFN- β -induced liver injury can be attributed to local ancestry deviations.

Pharmacogenomics Posters - Wednesday

PB2738. Implementation of clinical pharmacogenomics: How NCBI's Medical Genetics Summaries and MedGen can help educate and support clinicians.

Authors:

A. Malheiro, M. S. Kane, M. A. Hoepfner, T. Venkatapathi, B. Kattman; NCBI | NLM | NIH, Bethesda, MD

Abstract Body:

Pharmacogenomics (PGx) implementation depends upon specialized knowledge spread across many clinical disciplines and stages of clinical care. There is a need for a common foundation for PGx practice that enables information sharing for all stakeholders and a centralized resource to guide clinicians for optimized clinical care. The National Center for Biotechnology Information (NCBI) at the National Library of Medicine supports these efforts with the NIH Genetic Testing Registry (GTR), Medical Genetics Summaries (MGS), and MedGen which are all freely available. Approximately 335 US Food and Drug Administration (FDA)-approved drugs contain PGx information on the drug label; about 110 of which the FDA has deemed have sufficient evidence to suggest that genotype can affect drug metabolism and therapeutic effects. As the field of PGx expands so does the availability of clinical testing for PGx variants. Currently, GTR has information on over 880 tests for more than 200 PGx phenotypes from 112 laboratories—66 US and 46 international labs from 23 countries. MGS is a resource that aims to provide support to clinicians in need of PGx information by summarizing the latest research on the genetic factors that influence an individual's response to a drug, discussing available testing strategies, and providing access to practice guidelines from authoritative groups. The MGS articles compile practical information about PGx from authoritative sources for use in clinical settings in an unbiased approach and are extensively reviewed by experts. Currently, MGS covers 54 drugs with more in development. MedGen provides access to relevant clinical information on one page that is available for clinicians, laboratories, and researchers. MedGen aggregates and standardizes drug names, genes and gene products involved in drug metabolism and response, and pharmacogenetic phenotype names. Using automation and curated approaches, MedGen data is collected from authoritative sources: professional societies like CPIC and PharmGKB, and submitters to GTR and ClinVar. Data in MedGen is available for automated processing and can be integrated into products such as electronic health records. These clinical products at NCBI are cross-referenced to aid findability, linked to external authoritative groups, and offer targeted searches of the medical literature in PubMed and NCBI Bookshelf. Standardized terminology for drug-response clinical phenotypes in MedGen allows a clinician to search a specific drug name and discover information through this interconnectivity. By consolidating the expertise of the community, NCBI provides a common foundation for clinical PGx implementation.

Pharmacogenomics Posters - Thursday

PB2739. Investigating pharmacogenetics of dupilumab in the treatment of atopic dermatitis in an Asian clinical cohort

Authors:

Y. Yew^{1,2}, V. Koh³, F. Tai², M. Loh^{4,5,1}; ¹Natl. Skin Ctr., Singapore, Singapore, ²Lee Kong Chian Sch. of Med., Nanyang Technological Univ., Singapore, Singapore, ³Sch. of Biological Sci., Nanyang Technological Univ., Singapore, Singapore, ⁴Nanyang Technological Univ., Singapore, Singapore, ⁵Imperial Coll. London, London, United Kingdom

Abstract Body:

Background: Dupilumab is an approved targeted biologic therapy for moderate to severe atopic dermatitis (AD). While it is effective in many AD patients, its efficacy and adverse effects are variable. It can also be costly. Pharmacogenetics may allow us to better predict responders and episodes of adverse effects.

Aim: To better predict responders and adverse effects among AD patients treated with dupilumab by studying the variability in drug response due to genetic factors.

Method: Patients with atopic dermatitis (AD), treated with dupilumab from July 2020 to January 2021 at a tertiary skin centre were recruited and genotyped with the Infinium® Global Screen Array (GSA). Standard genotyping QC (quality control) procedures were performed using PLINK version 1.09. SNPs were excluded should they have a call rate < 98% and any possible gender swap after sex check. Genetic variants associated with treatment response, responders vs non-responders as a binary outcome were examined. Patients who achieved an Investigator Global Assessment (IGA) score of 0 and/or 1 at week 16 were considered as treatment responders. Those who have failed these treatment targets were deemed as non-responders. The Fisher's Exact Test was performed to compare allelic frequencies between responders and non-responders. A P-value of less than 0.05 was considered statistically significant. Statistical analyses and functional annotation were performed, and pathway analyses was conducted to understand the biological implications.

Results: A total of 15 patients were recruited with a total of 11 responders and 4 non-responders. Majority (86.7%) of patients were Chinese with 26.7% female. 160 variants with a minor allele frequency (MAF) > 50% were present in responders. These variants were significantly associated ($p < 0.05$) with responders to dupilumab. SNPs with the highest frequency include rs8110935, rs12154459, rs133434495, rs183693, rs2117655, rs3753931 and rs680930 with a MAF of 68.2%. Minor alleles from these 160 variants were not present in non-responders, strongly suggesting that patients who are heterozygous or homozygous for these minor alleles may be better suited for dupilumab treatment. Genetic variants significantly associated with dupilumab response were mostly involved in the immune system directly or indirectly. Other significantly associated pathways included those involving neurogenic inflammation.

Conclusion: These genetic variants could potentially be used as predictors, and provide better understanding of the underlying mechanisms involved in dupilumab treatment response. Future, larger sample size studies would be required to validate these findings.

Pharmacogenomics Posters - Wednesday

PB2740. LOF *CYP2C19* allele evaluation identifies significant number of Indian CAD patients are low responders to Clopidogrel

Authors:

G. Rastogi, V. Sharma, A. Sharma, L. GUPTA; NMC GENETICS INDIA PRIVATE LIMITED, GURUGRAM, India

Abstract Body:

Clopidogrel is used as an antiplatelet drug and is frequently prescribed to individuals with acute coronary syndrome and ischemic stroke. Despite the drug being effective in majority of the patients, some still experience ischemic events which might be due to poor platelet aggregation inhibition. *CYP2C19* plays an important role by converting Clopidogrel into its active form. Individuals with LoF alleles combinations are intermediate or poor metabolizers. An alternative antiplatelet therapy is recommended by Clinical Pharmacogenetics Implementation Consortium CPIC for poor and intermediate metabolizers. Genotyping tests are available to check the alleles of *CYP2C19* gene and the phenotypes are then interpreted as ultra, extensive, intermediate or poor metabolizer. Clopidogrel prescription in coronary artery disease patients is frequently given in South Asians especially in Asian Indians where the risk of CAD is almost double than Europeans. The goal of this study is to check the prevalence of cytochrome P450 2C19 *CYP2C19* loss-of-function SNPs or haplotypes in Indian CAD patients. In order to have accurate estimate of prevalence, the genotyping of 416 established CAD patients belonging to different states of India was performed. Perturbing results were observed with 50 % approximately of CAD patients were low responder to clopidogrel genetically; signifying evaluation of these LoF variations in CAD patients prescribed with clopidogrel in order to reduce recurrent ischemic event.

Pharmacogenomics Posters - Thursday

PB2741. Meta-analyses of genome wide association studies of toxicity to oxaliplatin and fluoropyrimidine chemotherapy in 1800 patients with advanced colorectal cancer

Authors:

K. Watts¹, C. Wills¹, A. Madi², C. Palles³, T. Maughan⁴, R. Kaplan⁵, N. Al-Tassan⁶, R. Kerr⁷, D. Kerr⁷, R. Houlston⁸, V. Escott-Price⁹, J. Cheadle¹; ¹Cardiff Univ., Cardiff, United Kingdom, ²The Clatterbridge Cancer Ctr. NHS Fndn. Trust, Wirral, United Kingdom, ³Univ. of Birmingham, Birmingham, United Kingdom, ⁴CRUK/MRC Oxford Inst. for Radiation Oncology, Oxford Univ., Oxford, United Kingdom, ⁵Univ. Coll. of London, London, United Kingdom, ⁶King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, ⁷Oxford Univ., Oxford, United Kingdom, ⁸The Inst. of Cancer Res., Sutton, United Kingdom, ⁹Dementia Res. Inst., Cardiff, United Kingdom

Abstract Body:

Background

Chemotherapies administered at normal therapeutic dosages can cause significant side-effects. Inter-individual variation in toxicity highlights the need for biomarkers to personalise treatment. We sought to identify such biomarkers by conducting meta-analyses of genome-wide association studies for ten toxicities in patients with advanced colorectal cancer treated with oxaliplatin and fluoropyrimidine chemotherapy ± cetuximab from the MRC COIN and COIN-B trials.

Methods

Association analyses were performed on 1,055 patients treated with capecitabine and oxaliplatin (XELOX) ± cetuximab and 745 with folinic acid, fluorouracil and oxaliplatin (FOLFOX) ± cetuximab. We also analysed rs6783836 in *ST6GALI* with hand-foot syndrome (HFS) in CRC patients from QUASAR2. Since genetic variation in *ST6GALI* is a risk factor for type-2 diabetes (T2D), a disease also associated with HFS, we sought to confirm an association between *ST6GALI* and T2D (17,384 cases, 317,887 controls) and analysed rs6783836 against markers of diabetes and inflammation using UK Biobank data.

Results

rs6783836 at *ST6GALI* was associated with HFS in patients treated with XELOX (OR=3.1, 95%CI=2.1-4.6, P=4.3x10⁻⁸) and was borderline significant in patients receiving capecitabine from QUASAR2, but with an opposite allele effect (OR=0.66, 95% CI=0.42-1.03, P=0.05). *ST6GALI* was associated with T2D (lead SNP rs3887925, OR=0.94, 95%CI=0.92-0.96, P=1.2x10⁻⁸) and the rs6783836-T allele was associated with lowered HbA1c levels (P=5.9x10⁻³) and lymphocyte count (P=2.7x10⁻³) beyond thresholds for multiple testing. We identified eight other loci associated with toxicities at suggestive significance levels (in XELOX treated patients: rs12433034 [*LOC105370399*] and rs143685874 [*LINC02504*] with stomatitis, and rs11295266 with diarrhoea; in FOLFOX treated patients: rs12029003 [*TMIGD3*], rs141748690 [*LOC105379133*] and rs1233378 with stomatitis; rs825249 [*CDH13*] with neutropenia and rs79848933 with nausea; all P<1.0x10⁻⁶).

Conclusion

rs6783836 in *ST6GALI* is a potential biomarker for HFS with links to T2D and inflammation. This and eight other putative biomarkers warrant further investigation for their potential clinical utility.

Pharmacogenomics Posters - Wednesday

PB2742*. Novel genetic signals identified for angiotensin-converting enzyme inhibitor-induced cough

Authors:

K. Coley¹, D. J. Shepherd¹, C. John¹, R. Packer¹, R. C. Free^{2,3}, E. J. Hollox⁴, L. V. Wain^{1,2}, M. D. Tobin^{1,2}, C. Batini¹; ¹Dept. of Hlth.Sci., Univ. of Leicester, Leicester, United Kingdom, ²NIHR Leicester BioMed. Res. Ctr., Univ. of Leicester, Leicester, United Kingdom, ³Dept. of Respiratory Sci., Univ. of Leicester, Leicester, United Kingdom, ⁴Dept. of Genetics and Genome Biology, Univ. of Leicester, Leicester, United Kingdom

Abstract Body:

Background: Angiotensin-converting enzyme inhibitors (ACEIs) are a class of antihypertensive drugs which are largely well tolerated. However, the most common adverse drug reaction (ADR) to their use is a persistent dry cough which affects 5-35% of users. In response to this ADR, clinical guidelines indicate a switch to an angiotensin-II receptor blocker (ARB). We have utilised such drug switches recorded in electronic health records (EHRs) as a proxy to study genetic determinants of ACEI-induced cough.

Methods: We have harnessed primary care EHRs linked to UK Biobank and the EXCEED study to identify cases who switched from ACEIs to ARBs within 12 months of initiating ACEIs, and controls who were continuous users of ACEIs. Utilising imputed genomic data in a meta-analysis of both cohorts, we identified ten independent sentinel variants (p -value $< 5 \times 10^{-8}$) in Stage 1, which were followed up in genome-wide association study (GWAS) summary statistics of ACEI-induced cough from the eMERGE Network (Stage 2), and meta-analysed. Functionally-informed fine-mapping was used to identify putative causal variants for variant-to-gene mapping.

Results: A total of 5,435 cases and 34,436 controls from four ancestral groups (African, East Asian, European and South Asian) were included in the Stage 1 meta-analysis of ~ 17 M variants. Of the 10 independent sentinel variants, 5 were also genome-wide significant (p -value $< 5 \times 10^{-8}$) in the Stage 1 and Stage 2 joint meta-analysis and 3 did not have available Stage 2 data. Variant-to-gene mapping identified *KCNIP4* which has been previously associated with ACEI-induced cough, as well as novel genes including *NTSR1* and *PREP* which are both involved in the mediation of neuropeptide activity. Further, PheWAS of rs6062847 (*NTSR1*) and genetic correlation analysis showed genetic overlap with chronic dry cough phenotypes.

Conclusion: We have utilised a proxy phenotype to perform the largest GWAS of ACEI-induced cough to date, identified putative causal variants and genes, and provided evidence for shared genetic determinants with chronic cough. In line with existing hypotheses, these findings highlight the role of proinflammatory mediators in sensory airway sensitivity and cough reflex modulation.

Pharmacogenomics Posters - Thursday

PB2743. Pharmacogenetics of tenofovir clearance among Southern Africans living with HIV.

Authors:

Z. Cindi¹, A. N. Kawuma¹, G. Maartens¹, P. Denti¹, R. Wasmann¹, Y. Bradford², M. Ritchie², L. Wiesner¹, S. Sokhela³, N. Chandiwana³, F. Venter³, D. W. Haas⁴, P. Sinxadi¹; ¹Univ. of Cape Town, Cape Town, South Africa, ²Univ. of Pennsylvania, Philadelphia, PA, ³Univ. of the Witwatersrand, Johannesburg, South Africa, ⁴Vanderbilt Univ, Nashville, TN

Abstract Body:

Background: Tenofovir disoproxil fumarate (TDF) and tenofovir alafenamide (TAF), prodrugs of the nucleotide analogue tenofovir, are prescribed in antiretroviral regimens to people with HIV (PWH). Plasma tenofovir exposure is higher when given as TDF and causes renal and bone toxicity in some patients. TAF yields lower plasma but higher intracellular tenofovir concentrations, and a more favorable safety profile. We characterized the pharmacogenetics of tenofovir clearance in Southern African PWH. **Methods:** Adults randomised to initiate TAF or TDF in dolutegravir-containing arms of the ADVANCE trial were studied. Plasma tenofovir clearance was estimated with non-linear mixed-effects models. Associations between pharmacogenetic polymorphisms, genome-wide polymorphisms, and unexplained variability in tenofovir clearance were examined using linear regression models stratified by study arm, with adjustment for 2 principal components. **Results:** 340 participants consented for genetic testing and associations were evaluable in 138 and 130 in the TAF and TDF arm, respectively. We found no significant associations with tenofovir clearance for either TAF or TDF among 5 polymorphisms previously associated with tenofovir pharmacokinetics (lowest P-value > 0.3 for each drug). Among 11 polymorphisms selected based on both prior strong association with any drug phenotype in PharmGKB and any genome-wide association with any trait in the GWAS catalog, the lowest P-value in both arms was for *IFNL4* rs12979860 (TAF: $P = 2.9 \times 10^{-3}$; TDF: $P = 3.0 \times 10^{-3}$), with the C allele associated with increased tenofovir clearance. This withstood correction for multiple testing. In genome-wide analyses, the lowest P-values for tenofovir clearance in the TAF and TDF arms were with *LINC01684* rs9305223 ($P = 3.0 \times 10^{-8}$) and intergenic rs142693425 ($P = 1.4 \times 10^{-8}$), respectively. Several additional polymorphisms were genome-wide significant [TAF: *LINC01684* rs4816969 ($P = 3.7 \times 10^{-8}$) and intergenic rs2829163 in chromosome 21 ($P = 4.5 \times 10^{-8}$); TDF: intergenic rs112914324 in chromosome 11 ($P = 1.6 \times 10^{-8}$) and intergenic rs11995962 in chromosome 8 ($P = 2.3 \times 10^{-8}$)]. **Conclusion:** Among Southern African PWH randomized to TAF or TDF, increased unexplained variability in tenofovir clearance was associated with an *IFNL4* polymorphism previously associated with resolution of hepatitis C virus infection. It is unclear how this gene would affect tenofovir disposition.

Pharmacogenomics Posters - Wednesday

PB2744. Pharmacogenomic analysis of drug metabolizing enzymes associated with plasma clozapine/N-desmethylclozapine ratio.

Authors:

S-I. Park^{1,2}, S. Kim³, K. Park⁴, J. Lee⁴, E. Kim^{3,5}; ¹Kangwon Natl. Univ., Chuncheon, Korea, Republic of, ²Kangwon Natl. Univ. Hosp., Chuncheon, Korea, Republic of, ³Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of, ⁴Kyung Hee Univ., Seoul, Korea, Republic of, ⁵Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract Body:

Background: Clozapine is a drug of choice for the treatment of resistant schizophrenia; however, inter-individual variations in the pharmacokinetics and pharmacodynamics limit its clinical benefits. N-desmethylclozapine, an active metabolite of clozapine, partially contributes to the clinical effect of clozapine. Clozapine is metabolized into N-desmethylclozapine by hepatic cytochrome P450 (CYP450) enzymes, with a major role in CYP1A2. Evaluation of genetic polymorphisms associated with the pharmacokinetic variability of clozapine can facilitate personalized pharmacotherapy.

Methods: Forty-five patients with flexible-dose escalation according to the clinician's instructions were included in the study. Genomic DNA was extracted from blood samples, and plasma metabolic ratio (clozapine/N-desmethylclozapine) was measured after eight (visit 2) and 18 weeks (visit 4) of the first dose of clozapine. Previously reported genes and SNPs related to the pharmacokinetics of clozapine were selected for targeted next-generation sequencing analysis. The genes included only the coding DNA sequence (CDS) region. For SNPs outside the CDS region, additional target probes were designed. This was followed by hybridization capture-based next-generation sequencing and data collection using a genome analysis tool kit. Statistical analysis was performed by using SPSS v 26.0. All *P*-values were based on two-sided comparisons, and *P* < 0.05 was considered statistically significant.

Results: Among the 81 variations identified by targeted capture sequencing, only 7 SNPs (2 for CYP1A2, 1 for UGT1A4, 1 for UGT1A1, and 3 for CYP2C19) showed nominal significance (*P* < 0.05) in metabolic ratio according to the genotype or phenotype of each gene. Phenotypes of CYP1A2 were determined by a combination of two SNPs, rs2069514 (c.-3860G > A) and rs762551 (c.-9-154C > A), showed significant associations with metabolic ratio (*P* = 0.020, t-test) at visit 4. In addition, phenotypes of CYP2C19 determined by three linked SNPs, rs12769205 (c.332-23A > G), rs4244285 (c.681G > A), and rs3758580 (c.990C > T), showed significant associations with metabolic ratio (*P* = 0.033, Kruskal Wallis test) at visit 2. The homozygous mutant group (N = 6) showed significantly higher metabolic ratio than the wild type + heterozygous mutant group (N = 39) at visit 2 (*P* = 0.010, Mann-Whitney test).

Conclusions: We found that the phenotypes of CYP1A2 and CYP2C19 were significantly associated with mid-term and earlier metabolic ratios of clozapine, respectively. These findings may provide a better understanding of the inter-individual variations in pharmacokinetics and pharmacodynamics of clozapine.

Pharmacogenomics Posters - Thursday

PB2745. Pharmacogenomics for All of Us: Approach to testing and return of results in 1M participants

Authors:

P. Empey¹, A. Radhakrishnan², J. D. Smith², S. McGee², E. Venner³, A. Haddad⁴, C. L. Neben⁵, A. Y. Zhou⁵, G-h. Kim⁵, H. Hoban⁵, K. Walker³, D. Kalra³, D. Murdock³, K. Doheny⁶, N. J. Lennon⁷, B. Shirts⁸, A. Wise⁹, A. Musick⁹, B. Ozenberger⁹, H. L. Rehm⁷, G. Jarvik², R. A. Gibbs³, J. C. Denny⁹, S. E. Topper⁵; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²Dept. of Genome Sci., Univ. of Washington, Seattle, WA, ³Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, ⁴Dept. of Pharmaceutical Sci., Univ. of Pittsburgh Sch. of Pharmacy, Pittsburgh, PA, ⁵Color Hlth., Burlingame, CA, ⁶Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ⁷Broad Inst. of MIT and Harvard, Cambridge, MA, ⁸Univ of Washington, Seattle, WA, ⁹All of Us Res. Program, NIH, Bethesda, MD

Abstract Body:

Pharmacogenomic (PGx) information is valued by patients and healthcare practitioners because genetic variation impacts drug pharmacokinetics and response. The All of Us Research Program (AoURP) is committed to return PGx reports along with other medically-actionable information to a diverse cohort of 1M participants. We designed an AoURP PGx return of results strategy to deliver value while mitigating potential risks within US regulatory frameworks.

In the AoURP, DNA derived from blood/saliva undergoes whole genome sequencing and participants provide secondary consent to elect to receive PGx results. Pharmacogenes, variants, phenotype translations, and medication associations to be reported were selected based on scientific evidence of actionability [Clinical Pharmacogenetics Implementation Consortium Level A gene-drug pairs, FDA PGx data and drug product labeling, professional society recommendations, and variants having a known altered function], the existence of validation controls, and the ability to support complex haplotyping. A priori analytical validation of PGx alleles was conducted using 221 clinical samples and cell lines. A participant-directed PGx report underwent multiple rounds of qualitative (n=21) and quantitative (n=205) user comprehension testing to ensure understanding and to minimize potential risks. The entire process was reviewed by the FDA for an Investigational Device Exemption (IDE). Finally, the value of returned PGx results was estimated by evaluating the prevalence of medication exposures among participants who provided electronic health record (EHR) data as of Jan. 2021.

Seven genes (CYP2C19, DPYD, G6PD, NUDT15, TPMT, SLCO1B1, and UGT1A1) and 46 variants were chosen for initial return. Validation results yielded >99% concordance with the truth set and between labs. User testing of the PGx report showed high comprehension for genetics knowledge (98.0%) and self-efficacy (98.3%). An IDE was awarded after 2 years of FDA engagement and importantly, included a reporting rubric for inclusion of 57 medications based on predicted phenotype associations. Among 194,418 AoURP participants with EHR medication data, 5.02% (2,618,872/52,158,718) of medication exposures were for PGx drugs included in the IDE and 60.0% (116,687/194,418) of participants had an exposure to at least one drug that had a potential PGx interaction.

The AoURP return of results strategy represents a path forward for returning PGx results to research participants. Information provided in the “Medicine and Your DNA” report is likely to be of interest to the majority of AoURP participants based on their common exposures to impacted medications.

Pharmacogenomics Posters - Wednesday

PB2746. Predicted expression of enzymes in the thiopurine metabolic pathway and risk of side effects in patients with inflammatory conditions treated with azathioprine

Authors:

L. Daniel, A. Dickson, P. Nepal, J. Zanussi, T. Miller-Fleming, P. Staub, W-Q. Wei, A. Davis, A. Hung, N. Cox, Q. Feng, M. Stein, C. Chung; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Introduction: Thiopurines are immunosuppressants with numerous side effects. We recently described that a higher risk score generated from the predicted liver expression of *AOXI* and *NME1* (enzymes in the thiopurine metabolic pathway) was associated with azathioprine discontinuation attributed to hematopoietic toxicity. Little is known about the association between this score and other known azathioprine side effects. We examined the association between the score and known azathioprine side effects. **Methods:** This was a retrospective cohort study of new azathioprine users with inflammatory conditions (e.g., systemic lupus erythematosus, rheumatoid arthritis); we limited the cohort to White TPMT/NUDT15 normal metabolizers and used MEGAchip genotyping with quality control and imputation. The risk score used PrediXcan (GTEx version 8) estimates of the genetically regulated expression of *NME1* and *AOXI* for each patient. Phecodes group phenotypes into clinically meaningful categories using ICD codes. We prespecified nine groups of phecodes (infections, hematologic, dermatologic, skin malignancies, other malignancies, hepatotoxicity, other gastrointestinal, constitutional, and other) corresponding to known azathioprine adverse effects. We tested the association between the score and the candidate phenotypes (n=99) using PheWAS. For outcomes that were significant ($p < 0.05$) in the PheWAS, we reviewed electronic health records to confirm the presence or absence of the side effects and performed logistic regression to assess the associations between the score and confirmed cases. **Results:** The cohort included 1259 patients (mean age 44.6 ± 17.6 years, 66% female) followed over a median of 1.6 years. The gene expression score (range 0.51 to 9.87 units) was significantly associated with the phecode for leukopenia (288.1); record review confirmed all 23 potential cases. Higher risk scores were associated with leukopenia [OR=1.19, (95% C.I. 1.02-1.39), $p=0.026$]. Results remained significant after adjustment for azathioprine daily. The score was not significantly associated with rash [OR 1.01, p-value 0.94], nausea and vomiting [OR 0.98, p-value = 0.75], thrombocytopenia [OR 0.97, p-value 0.75], or other known azathioprine adverse reactions. **Conclusion:** There was an association between a score including the predicted expression of *AOXI* and *NME1* with leukopenia. We found no associations between the score and other side effects, suggesting that these side effects are independent of *AOXI* and *NME1*.

Pharmacogenomics Posters - Thursday

PB2747. *PTPN2* and Discontinuation of Azathioprine Attributed to Myelotoxicity.

Authors:

A. Dickson, J. T. Zanussi, P. Nepal, L. L. Daniel, P. Straub, T. W. Miller-Fleming, W-Q. Wei, A. M. Hung, N. J. Cox, C. Stein, J. D. Mosley, Q. FENG, C. P. Chung; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Background: Azathioprine is an affordable immunosuppressant used globally to treat wide-ranging conditions, including multiple rheumatic conditions. Despite its broad application, dose-dependent side effects (e.g., myelotoxicity) frequently limit azathioprine use. Variants in genes within the thiopurine metabolic pathway—*TPMT* and *NUDT15*—predict some cases of myelotoxicity; however, most people who discontinue azathioprine for myelotoxicity are normal *TPMT* and *NUDT15* metabolizers. We hypothesized that a variant in *PTPN2*—rs11664064—that reached significance in a GWAS of azathioprine-associated leukopenia (defined as white blood cell count < 3.0 K/ μ L during azathioprine exposure) among these patients may predict discontinuation for myelotoxicity in a real-world clinical setting.

Methods: We conducted a retrospective cohort study among new users of azathioprine without an indication of organ transplant at a tertiary care center. We restricted the cohort to patients with EHR-reported White race and normal *TPMT* and *NUDT15* metabolizer status. The outcome was azathioprine discontinuation attributed to myelotoxicity in the clinical record. We used Cox hazard regression with competing risk (i.e., discontinuation attributed to other side effects) to examine the association between this variant and the clinical decision to discontinue azathioprine attributed to myelotoxicity. Given the small number of patients with the homozygous variant (n=4), we grouped patients as carriers or non-carriers of this variant for analysis.

Results: In the cohort of 1,184 patients, there were 39 outcomes among 1,063 non-carriers followed for a median of 21 [4-59] months; there were 10 outcomes among 121 among carriers followed for a median of 26 [5-75] months. In unadjusted competing risk analysis, carriers were at higher risk to discontinue azathioprine for attributed myelotoxicity: HR=2.28 (95%CI: 1.13-4.61, p=0.02). Results were similar without competing risk or when adjusted for 1) age and sex, and 2) age, sex, indication, initial dose-weight ratio, and prior testing for *TPMT*.

Conclusion: The rs11664064 variant in *PTPN2* was associated with a higher risk of azathioprine discontinuation attributed to myelotoxicity among White patients with indications other than organ transplants.

Pharmacogenomics Posters - Thursday

PB2748. Testing for common pharmacogenomic predictors of vigabatrin-associated visual field loss, under univariate & polygenic models.

Authors:

I. Boothman^{1,2,3}, L. Clayton^{4,5}, M. McCormack¹, R. Stevelink⁶, P. Moloney¹, R. KRAUSE⁷, W. Kunz⁸, S. Diehl⁷, T. J. O'Brien⁹, G. Sills¹⁰, F. Zara^{11,12}, B. Koeleman⁶, C. Depondt¹³, A. G. Marson¹⁴, H. Stefánsson¹⁵, K. Stefánsson¹⁵, J. Craig¹⁶, M. R. Johnson¹⁷, P. Striano^{11,12}, H. Lerche¹⁸, N. Delanty¹, S. Sisodiya^{4,5}, G. Cavalleri^{1,2,3}; ¹Royal Coll. of Surgeons in Ireland, Dublin, Ireland, ²The SFI Futureneuro Res. Ctr., Royal Coll. of Surgeons in Ireland, Dublin, Ireland, ³The SFI Ctr. for Res. Training in Genomics Data Sci., Galway, Ireland, ⁴Dept. of Clinical and Experimental Epilepsy, UCL Queen Square Inst. of Neurology, London, United Kingdom, ⁵Chalfont Ctr. for Epilepsy, Bucks, United Kingdom, ⁶Dept. of Genetics, Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, ⁷Luxembourg Ctr. for Systems Biomedicine, Univ. of Luxembourg, Esch-sur-Alzette, Luxembourg, ⁸Div. of Neurochemistry, Dept. of Epileptology, Univ. Bonn Med. Ctr., Bonn, Germany, ⁹Dept.s of NeuroSci. and Neurology, Central Clinical Sch., Monash Univ., The Alfred Hosp., Melbourne, Australia, ¹⁰Sch. of Life Sci., Univ. of Glasgow, Glasgow, United Kingdom, ¹¹"G. Gaslini" Inst., Genova, Italy, ¹²Dept. of NeuroSci.s, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Hlth., Univ. of Genoa, Genoa, Italy, ¹³Dept. of Neurology, Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium, ¹⁴Dept. of Pharmacology and Therapeutics, Univ. of Liverpool, Liverpool, United Kingdom, ¹⁵deCODE Genetics, Reykjavik, Iceland, ¹⁶Dept. of Neurology, Royal Victoria Hosp., Belfast Hlth. and Social Care Trust, Belfast, United Kingdom, ¹⁷Div. of Brain Sci., Imperial Coll. Faculty of Med., London, United Kingdom, ¹⁸Dept. of Neurology and Epileptology, Hertie Inst. for Clinical Brain Res., Univ. of Tübingen, Tübingen, Germany

Abstract Body:

Background: The anti-seizure drug vigabatrin (VGB) is an effective drug for controlling seizures, especially infantile spasms. However, use is limited by VGB-associated visual field loss (VAVFL). Approximately 33% of VGB-exposed adult patients experience this adverse reaction (Maguire, 2010) although the mechanisms by which VGB causes VAVFL remains unknown. Average peripapillary retinal nerve fiber layer (ppRNFL) thickness is correlated with the degree of visual field loss (measured by mean radial degrees) (Clayton, Dévilé et al. 2011). Duration of VGB exposure, maximum daily VGB dose, & male sex are associated with peripapillary retinal nerve fiber thinning. We hypothesize that common genetic variation is a predictor of VAVFL. Identifying pharmacogenomic predictors of VAVFL could potentially enable safe prescribing of VGB and broader use of a highly effective drug. Methods: Optical coherence topography (OCT) data were produced on VGB-exposed individuals (n=99) from the EpiPGX consortium. We conducted quantitative GWAS analyses for the following OCT thickness measurements: 1) Average ppRNFL, 2) inferior quadrant, 3) nasal quadrant, 4) superior quadrant, 5) temporal quadrant, 6) inferior nasal sector, 7) nasal inferior sector, 8) superior nasal sector, 9) nasal superior sector. We included sex, cumulative dose, maximum daily dose, duration of prescription of VGB (years) and 4 principal components as covariates. Using the summary statistics from the GWAS analyses we conducted gene-based testing using VEGAS2. To determine if VGB exposed patients were predisposed to having a thinner RNFL, we calculated their polygenic burden for retinal thickness. We conducted nine different PRS analyses using the OCT measurements. PRS alleles for retinal thickness were calculated using the summary statistics from a large-scale GWAS of inner retinal morphology using the OCT images of 31,434 UK Biobank participants (Currant, Hysi et al. 2021). Results: The GWAS analyses did not identify a significant association after correction for multiple testing. Similarly, the gene-based and PRS analyses did not reveal a significant association that survived multiple testing. Conclusion: We set out to

ASHG 2022 Annual Meeting Poster Abstracts

identify common genetic predictors for VGB-induced ppRNFL thinning (as a marker of VAVFL), under univariate & polygenic models. Results suggest that large-effect common genetic predictors are unlikely to exist for VGB-induced VAVFL. Sample size was a limitation of this study; but recruitment is a challenge as VGB is rarely used today because of this adverse reaction. Rare variants may be predictors of this ADR and were not studied here.

Pharmacogenomics Posters - Wednesday

PB2749. The Alabama Genomic Health Initiative Integrating Genomics into Primary Care

Authors:

N. Limdi¹, D. Absher², I. Asif¹, G. Barsh², L. Bateman¹, K. Bowling³, G. Cooper², B. Davis¹, K. East², C. Finnila², B. Goff¹, M. Kelly², S. Hiatt², W. Kelley², B. Korf¹, D. Latner², J. Lawlor², T. May⁴, M. Might¹, I. Moss¹, M. Nakano-Okuno¹, T. Osborne¹, S. Sodeke⁵, A. Stout², M. Thompson²; ¹Univ. of Alabama at Birmingham, Birmingham, AL, ²Hudson Alpha Inst. for Biotechnology, Huntsville, AL, ³Washington Univ. Sch. of Med., Saint Louis, MO, ⁴Washington State Univ., Spokane, WA, ⁵Tuskegee Univ., Tuskegee, AL

Abstract Body:

Introduction The Alabama Genomic Health Initiative (AGHI) is a state-supported, IRB-approved research study launched in 2017 to evaluate the impact of population-based genomic screening. In 2021, AGHI adopted a new clinic-based model, partnering with UAB Family Medicine providers with a primary goal of integrating genomic medicine into primary care. Participants and providers receive results about variation in medically actionable genes and pharmacogenetics (PGx). **Methods** Participants are enrolled in UAB clinics (Birmingham, Hoover, and Selma, AL), and a blood sample, personal/family history, and medications are collected. DNA is analyzed using the Illumina Global Diversity Array and Taqman assays (*CYP2D6* copy number changes). Sanger sequencing is used to confirm variation in medically actionable genes. Sanger sequencing and PGx analyses are performed in CLIA-certified labs. Genetic counselors and pharmacists review results in the context of the participant's personal/family history and medications, respectively, and highlight potential risks in return-of-results (ROR) reports. Provider reports are returned via the electronic medical record (EMR), and patient reports are mailed approximately two weeks later. Providers are notified when a high impact result is identified. Genetic counselors and pharmacists are available to provide education and guidance to providers and patients. Provider feedback is used to identify and refine processes.

Results From July 2021-June 2022, 485 individuals (62% non-white; 75% female) have enrolled in AGHI, and 286 have completed analysis and ROR. Of these, 31.1% (n=89) had a result affecting a current medication, and 1% (n=4) had a medically actionable result (*KCNQ1*, *TTN*, *BTD*, and *CACNA1S*). To enable the incorporation of genomic information into the EMR, we worked with UAB Health System Information Services, to develop 1) automated transfer of reports into the EMR, 2) Genetics and Pharmacogenetics Reports folders and 3) a Genomic Medicine MPage in Cerner, which allows providers to view genetic reports in one location.

Conclusion There has been substantial interest on the part of family medicine providers in being involved in AGHI, and we have obtained important feedback. Many of their patients are from disadvantaged socioeconomic backgrounds, and much of the feedback has entailed refinement of messaging about participation and return of results. Additionally, embedding genetic results in the EMR allows this information to be available and inform future care and prescribing.

Pharmacogenomics Posters - Wednesday

PB2750. Vaccinations against pneumonia and the flu may protect carriers of particular genotypes against Alzheimer's disease

Authors:

S. Ukraintseva¹, K. Arbeev¹, H. Duan¹, A. Yashkin¹, A. Tropsha², I. Akushevich¹, A. Yashin¹; ¹Duke Univ., Durham, NC, ²UNC, Chapel Hill, NC

Abstract Body:

Background: Vaccines with beneficial off-target effects could be promising candidates for repurposing for Alzheimer's disease (AD) prevention. Earlier we found that the association of pneumococcal vaccine with AD can be modified by rs2075650 polymorphism in *TOMM40*, a known genetic risk factor for AD. Here we further explore the modulating effects of genetic risk factors for AD on the associations between vaccinations and AD risk. In this analysis we focus on the SNP rs6859 in *NECTIN2* gene that was linked to *both* AD and vulnerability to infections in prior research, including our own. **Method:** We evaluated associations between vaccinations against pneumonia and the flu received between ages 65 and 75 and the risk of AD later in life (a binary variable defined as 0 = "no AD, and lived longer than 75 years", and 1 = "AD onset after age 75") in the sample of 5,146 participants of the Cardiovascular Health Study, taking into account the rs6859 polymorphism. Logistic model with covariates including sex, race, birth cohort, education, smoking status, and number of A alleles of rs6859 (AD risk factor) was used. We also performed stratified analysis by genotype (carriers vs. non-carriers of A alleles of the rs6859). **Results:** The total count of vaccinations against pneumonia and the flu received between ages 65-75 was associated with the lower risk of AD later in life in the model with all covariates (OR=0.87; p=0.01). In the stratified analysis, the total count of vaccinations was associated with the lower risk of AD only in carriers of A allele of the rs6859 (OR = 0.84; p=0.04), but not in the non-carriers. Number of flu shots, separately, didn't show a significant association with AD. Pneumococcal polysaccharide vaccine received between ages 65 and 75 (with or without an accompanying flu shot) was associated with 33% lower risk of AD onset later in life in carriers of A allele (OR=0.67; p=0.05), but not in the non-carriers. All odds ratios were close to 1 in the non-carriers of A allele group. **Conclusion:** In this study, pneumococcal polysaccharide vaccine, as well as the total count of vaccinations against pneumonia and the flu received between ages 65-75, had genotype-specific associations with the risk of AD, with protective effects of vaccinations seen only in carriers of A allele of the rs6859, but not in the non-carriers. These results support the repurposing of pneumococcal and flu vaccines for personalized, genotype-tailored, AD prevention.

Pharmacogenomics Posters - Thursday

PB2751. Variant Calling and Orthogonal Verification of 20 Pharmacogenomics (PGx) Clinical Pharmacogenetics Implementation Consortium (CPIC) Level A Genes including 4 Challenging Genes Using Illumina DRAGEN Bio-IT Platform.

Authors:

S. Catreux¹, K. Mukherjee¹, E. Roller¹, V. Onuchic¹, I. Filip¹, S. Tzur¹, J. Han¹, M. Axford², M. Curtis², L. Lau²; ¹Illumina, San Diego, CA, ²The Hosp. for Sick Children, Toronto, ON, Canada

Abstract Body:

Pharmacogenomics (PGx) is a field of research that studies how a person's genes affect how he or she responds to medications. Information from PGx testing is used by healthcare providers to tailor drug dose, predict side effects, and measure effectiveness of drugs for individuals with defined genotypes.

The Clinical Pharmacogenetics Implementation Consortium (CPIC) curates and posts freely available, peer-reviewed gene/drug clinical practice guidelines. Out of the list of PGx-related genes, there is a set of 20 genes classified as "Level A" by CPIC (as of March 2022), with moderate to high levels of evidence in favor of changing drug prescribing.

Variants from 16 of 20 genes can be called with high accuracy using standard whole genome sequencing (WGS) or panel-based workflows. However, 4 genes have been difficult to interrogate with next-gen sequencing (NGS): HLA-A/B genes due to high levels of polymorphism and CYP2B6/CYP2D6 due to paralogous regions in the genome. CYP2B6 has one pseudogene (CYP2B7), with ~ 93% shared sequence homology. CYP2D6 has two pseudogenes CYP2D7 (~97% homology) and CYP2D8P (~91% homology). Here, we introduce the application of 3 gene specific targeted callers (HLA-A/B, CYP2D6, CYP2B6) as well as describe the integrated star allele caller which genotypes the following remaining 16 genes: CACNA1S, CFTR, CYP2C19, CYP2C9, CYP3A5, CYP4F2, IFNL3, RYR1, NUDT15, SLCO1B1, TPMT, UGT1A1, VKORC1, DPYD, G6PD and MT-RNR1. The PGX calling function in DRAGEN provides an easy-to-use workflow that can genotype these CPIC Level A genes in a single streamlined process starting from sequence read data (FASTQ or BAM) to final star allele calls.

The performance of the DRAGEN PGx calling function was assessed through two independent studies. The first study showed that DRAGEN's PGx star allele caller achieved 100% concordance with external PGx callers such as PharmCAT and Aldy when tested on around 3300 samples from Coriell and the 1k Genome project. The CYP2D6 and CYP2B6 callers were tested against GeT-RM PGx calls and found to have 99.3% and 92% concordance respectively. The HLA-A/B caller has an accuracy of 99.6% when compared against a reference truth set generated via Sanger sequencing.

The second study was executed by the Hospital for Sick Children (SickKids). They examined 20 samples using orthogonal PGx testing via MassArray, the external PGx caller PharmCAT (using DRAGEN 3.10.8 variant calls), and the DRAGEN integrated star allele caller in v4.0 and found close to 100% concordance between the calls.

Variant calling for all 20 genes including 4 gene specific targeted callers will be available in DRAGEN v4.0.

Pharmacogenomics Posters - Wednesday

PB2752. Variation in *CYP2A6*, a nicotine metabolism gene, associates with lung diseases: A Phenome-wide Association and Mendelian Randomization study.

Authors:

H. Giratallah^{1,2}, M. J. Chenoweth^{1,2}, J. Pouget^{1,2,3}, A. El-Boraie^{1,2}, C. Lerman⁴, J. Knight⁵, R. F. Tyndale^{1,2,3}; ¹Dept. of Pharmacology & Toxicology, Univ. of Toronto, Toronto, ON, Canada, ²Campbell Family Mental Hlth.Res. Inst., CAMH, Toronto, ON, Canada, ³Dept. of Psychiatry, Univ. of Toronto, Toronto, ON, Canada, ⁴Norris Comprehensive Cancer Ctr., Univ. of Southern California, Los Angeles, CA, ⁵Data Sci. Inst., Lancaster Univ. Med. Sch., Bailrigg, United Kingdom

Abstract Body:

Background: *CYP2A6* is a genetically variable enzyme that inactivates nicotine, activates carcinogens (e.g. nitrosamines), and metabolizes many pharmaceuticals. Genetic variation in *CYP2A6* changes smoking behaviours and the risk for tobacco-related diseases. This Phenome-Wide Association Study (PheWAS) examined associations between our previously developed *CYP2A6* weighted genetic risk score (wGRS), a robust genotypic marker of *CYP2A6* activity, with diseases in the UK Biobank (N~400,000). The top associations were then evaluated using Mendelian Randomization (MR) to estimate the potential causal effects of *CYP2A6* activity on these diseases. **Methods:** The wGRS was computed for participants in the UK Biobank after extracting raw or imputed genotypes for each of the *CYP2A6* wGRS variants or their proxy variants ($r^2 > 0.8$). The reconstructed wGRS using proxy variants, was validated in an independent sample (N=922). The phenome consisted of mapped UK Biobank diagnosis codes (ICD-9 and ICD-10) to PheCodes. PheCodes-wGRS associations were tested in linear models adjusting for and/or stratifying on relevant covariates. Two-sample MR used the independent sample as *CYP2A6* exposure and the phenome-wide significant (PWS) signals in the UK biobank as outcomes in current smokers. **Results:** In the total sample and current smokers, six PWS signals were identified: two lung cancers and four obstructive respiratory diseases, where higher wGRS (i.e. faster *CYP2A6* activity) was associated with higher disease risk ($P < 1 \times 10^{-6}$). No PWS signals were detected in former or never smokers. In current smokers, the top lung cancer signal had a higher estimate in females versus males. However, when restricted to lighter smokers (≤ 20 cigarettes/day), females and males estimates were similar. MR effects of *CYP2A6* activity on the six PWS signals in current smokers were significant (OR=1.168-1.442; $P < 0.001$), providing evidence supporting causation. MR sensitivity analyses by Egger and leave-one-out methods indicated no evidence of horizontal pleiotropy. Whether the *CYP2A6* effect is via nitrosamine activation or mediated by increased smoking is now being investigated using mediation and mediation MR analyses. **Conclusions:** This is the first study showing, in a hypothesis-free approach in a large biobank, an association of *CYP2A6* wGRS with lung cancers and obstructive respiratory diseases and providing evidence supporting causation. Our findings support a role for *CYP2A6* variation as a risk factor for lung cancer and obstructive respiratory diseases, with potential utility in early lung cancer screening programs that currently only rely on smoking history and age.

Pharmacogenomics Posters - Thursday

PB2753*. Whole Exome Sequencing adds power to prioritize drug targets when combined with other genetic data types

Authors:

S. Ranganathan¹, J. Flannick²; ¹Broad Inst. of Harvard and MIT, Boston, MA, ²Boston Children's Hosp., Boston, MA

Abstract Body:

90% of drugs that enter clinical trials do not get approved, increasing the cost of drug development. This high attrition rate is partly due to the inability of preclinical models to accurately predict efficacy. Human genetic evidence can address these limitations by providing 'experiments of nature' that inform on the effects of drug target perturbation in vivo. Existing studies have presented the odds ratio (OR) of drugs supported by genetic evidence being approved over drugs without genetic evidence getting approved, focusing on common variant associations from GWAS (OR=1.5) and Mendelian disease associations from OMIM (OR=2.6). Whole exome sequencing (WES) studies present a third-class of associations that balance the benefits of GWAS associations and Mendelian disease associations. We sought to evaluate (a) the ability of WES associations to predict drug success rate; (b) the extent to which various ways of analyzing WES data impact this ability; and (c) the extent to which WES data complements the other classes of genetic associations in providing genetic support for drug targets. For our analysis, we used WES gene-trait associations in the UK Biobank and previously published drug target-indication data. We compared drug approval ORs and the number of drugs with genetic support across GWAS, OMIM, and WES data, and evaluated the effects of various parameters (p-value thresholds to consider an association significant, variant mask and statistical used to compute a gene-trait association) on these values for WES data. Comparing the gene-trait associations identified by the different databases, though WES prioritized fewer associations than OMIM and GWAS, it identified new gene-trait associations (7) even with the strictest significance thresholds of $p < 1e-05$. WES gene-trait associations under different classes of diseases revealed a significant difference in the approval ORs for the different classes of diseases (OR = 1.5 for cardiovascular diseases, OR = 2.5 for neoplasms). For all parameter choices, WES data showed significant drug approval ORs, with the highest OR produced by a significance threshold of $p < 1e-04$ for a burden test of pLoF variants and SKAT test of missense variants ((OR = 2.61). Integrating this with GWAS and OMIM by taking the union of the gene-trait associations in the three databases, OR = 1.71 with 131862 gene-trait associations fulfilling the different parameters. Overall, these results demonstrate that incorporating WES data into a pipeline for drug target validation can add substantial information. Investigating the optimal way to integrate WES, OMIM and GWAS into a genetic support pipeline could be a valuable research direction.

Evolutionary and Population Genetics Posters - Wednesday

PB2754. A Bayesian framework to capture the probability of polymorphism with sequence context across the human genome.

Authors:

C. Adams, M. Conery, B. Auerbach, S. T. Jensen, I. Mathieson, B. F. Voight; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Germline mutations are the primary vehicle by which genetic variation is created and maintained in species, with inferences derived from mutation rate models fundamental to many population genetics inference methods. Previous models have demonstrated that nucleotides flanking polymorphic sites - local sequence context - explain variation in the probability that a site is polymorphic in the genome. However, limits to these models exist as the size of the local sequence context window expands, which includes the lack of robustness to data sparsity for practical applications, lack of regularization to generate parsimonious models, and finally a lack of quantified uncertainty in estimated rates to facilitate comparison of rates between models. To address these limitations, we present Baymer, a regularized Bayesian hierarchical tree model to capture the heterogeneous effect of sequence contexts on polymorphism probabilities. Baymer implements an adaptive Metropolis-within-Gibbs Markov Chain Monte Carlo sampling scheme to estimate the posterior distributions of sequence-context based probabilities that a site is polymorphic. We show that Baymer accurately infers polymorphism probabilities and emits well-calibrated uncertainty estimates based on the posterior distributions, is robust to circumstances of data sparsity, appropriately regularizes to return parsimonious models, and is computationally scalable at least up to 9-mer sized context windows. We demonstrate application of Baymer in two ways -- first, in identifying differences in polymorphism probabilities between continental populations in the 1000 Genomes Phase III dataset, and secondly in a sparse-data setting to demonstrate the portability of polymorphism models as a proxy for de novo mutation probabilities as a function of variant age, sequence context window size, and demographic history. Altogether, we demonstrate Baymer is a dynamic polymorphism probability estimation algorithm that is well suited to handling the sparsity of existing datasets, varying information content in sequence context window sizes, and accounts for uncertainty in these parameters, thus providing a novel tool for population genetics inference problems.

Evolutionary and Population Genetics Posters - Thursday

PB2755. A likelihood-based framework for demographic inference from genealogical trees

Authors:

C. Fan¹, N. Mancuso², C. Chiang²; ¹Univ. of Southern California, Los Angeles, CA, ²Univ. OF SOUTHERN CALIFORNIA, Los Angeles, CA

Abstract Body:

Demographic inference of a population not only sheds light on its population size and migration histories, but also serves as a null model for patterns of genetic variation, from which one can further construct non-neutral evolutionary models. However, traditional methods to reconstruct demographies from genetic data rely heavily on simplifying assumptions or summary statistics of the underlying genealogy, thereby undermining accuracy and/or flexibility of the results. For instance, populations are often assumed to be homogeneous with little or no migration, or that the genealogical tree is trimmed, losing subtle information contributed by ancient population histories. Here we propose a genealogical likelihood framework, gLike, to derive the likelihood of an ancestral recombination graph (ARG) given a proposed demography. Intuitively, we use a continuous time Markov process to describe the migration of lineages among populations, and the coalescence of lineages within the same population. A given genealogical tree may result from different traveling histories of the lineages (called paths), and the total probability of the tree is the summation over all possible paths. Implemented with maximum likelihood methods, gLike gives the most plausible demography to explain the observed ARG. The fact that our method is directly based on the ARG — rather than genetic data as in previous methods — lessens the assumptions and dependence about the mutation process, and naturally connects the flexibility of the ARG structure with the variety of demographic models, especially those with complex admixtures. Through extensive simulations on hundreds of samples, we show that demographic parameters estimated by gLike are unbiased and more accurate compared with existing non-ARG-based methods. Our work sheds light on the stochastic process of the formation of a genealogy under an arbitrary population history, which may facilitate investigations into genetic questions such as the areas including population structure analysis and history of genetic disease alleles. To our knowledge, gLike is the first full likelihood demographic model inference method based on the ARG, and because our framework requires only few light assumptions, we expect it to be a firm foundation on which other applicable theories or methods about the ARG could be built.

Evolutionary and Population Genetics Posters - Wednesday

PB2756. A robust kinship inference approach based on machine learning method for missing person identification

Authors:

M. Huang; Univ. of North Texas Hlth.Sci. Ctr., Fort worth, TX

Abstract Body:

Estimating the relationships between individuals is the fundamental component of missing person identification. Recent advances in whole genome sequencing (WGS) and genotyping microarray technology have enabled rapid profiling of millions of single nucleotide polymorphisms (SNPs), which provides much higher cumulative discrimination power than that of the traditional Short Tandem Repeats (STRs) analysis. However, for missing persons cases, the samples (e.g., bones) are usually highly degraded, and ancestry information of these samples are usually unknown. Thus, the genotyping error of these samples could be high (e.g., ~5% or higher depending on the quality filtering of the data). The Identity-by-Descent (IBD) segment measures can be easily interrupted by genotyping errors, but the genome-wide relatedness measures (e.g., King-robust) that are based on individual markers may be more robust to errors. In this study, we describe a machine learning approach that combines 17 genome-wide relatedness measures to train classification models aiming to reduce the effect of genotyping error and improve the accuracy of relationship estimation. A hierarchical classification strategy was implemented, in which a top-level classifier was to determine the relationship degrees (e.g., 1st, 2nd, 3rd, and unrelated) and three second-level classifiers were for determining the specific relationships within each degree. (e.g., parent-child or full-sibling for 1st degree). Both simulated and real datasets with various error rates (from 1% to 10%) were used to train and test the classifiers. The results showed that this approach outperformed the individual measures (e.g., KING-robust). In addition, higher accuracies could be obtained using the training sets with the correct error rates (i.e., training and test datasets share the same error rate). This new approach also had a robust performance with lower density SNPs (e.g., as low as 50K).

Evolutionary and Population Genetics Posters - Thursday

PB2757. Addressing population stratification in GWAS with variational autoencoders.

Authors:

D. Yang, J. Bartlett, W. Bush; Case Western Reserve Univ., Cleveland, OH

Abstract Body:

Variational autoencoders (VAEs) are generative deep learning models in which a mirrored pair of neural networks seek to first compress a set of input variables into a low-dimensional latent space and then recreate the input data to provide a measure of information loss due to the compression. Unlike the more standard principal components analysis (PCA), they can incorporate non-linear relationships that allow a low dimensionality representation within a latent space with better preservation of global geometry. Recently, VAEs have been adopted for single-cell sequencing and transcriptome and methylome data analysis because of their high capacity to manage large-scale datasets. Population stratification acts as an omnipresent threat to the validity of genome-wide association studies (GWAS) that could cause spurious associations of risk factors to diseases. Traditionally, PCA has been used for visualization and inference of population structure in genotype data but often results in multiple dimensions capturing linear patterns only. Here, we explored the utility of VAEs for addressing population stratification in GWAS using popvae (Batty *et al.* 2020) that took unphased genotypes from unrelated individuals from the Alzheimer's Disease Sequencing Project dataset (n = 14,604) to train the popvae algorithm and obtained sampled coordinates in a two-dimensional latent space that captures subtle aspects of individuals' ancestral information. We compared the results from popvae, PCA, and K-means methods to the self-reported ancestry via visualization. Then we conducted GWAS of common variants using the latent coordinates as ancestry estimation and contrasted performance to adjustments for PCs. We used the genomic inflation factor (λ) and identified risk factors for quality assessment. Popvae classified all samples into five distinct clusters with only two dimensions compared to ten PCs with PCA. Between our classifications and self-reported ancestry, we found higher rates of agreement for non-Hispanic Black and Hispanic groups than the non-Hispanic White, non-Hispanic Asian, and other non-Hispanic populations. We identified a similar λ value (VAEs: 1.28; PCA: 1.14) and significant risk factors near the *APOE* gene and the *APP* gene. Overall, VAEs are a promising new approach that may allow an equivalent or better identification of genetic ancestry components and subgroups for genetic association analyses that expend fewer degrees of freedom than PCA, especially in the setting of large-scale data with multi-ancestry samples. They could also be integrated into many other multi-omics analyses to improve the performance of dimensionality reduction tasks.

Evolutionary and Population Genetics Posters - Wednesday

PB2758. Analysis of evolutionary pressure on gene expression identifies patterns of selection in tissues and disease

Authors:

A. Sartori¹, T. Dupuis², E. T. Dermitzakis¹, A. A. Brown²; ¹Université de Genève, Dept. of Genetic Med. and Dev., Genève, Switzerland, ²Univ. of Dundee Sch. of Med., Dundee, United Kingdom

Abstract Body:

As our collections of populations and species with DNA sequence expands, across multiple species, geographical regions and time-frames with ancient DNA, so does our ability to identify regions of the genome either conserved due to negative selection or under more recent selection specific to the environment. In parallel, molecular studies and disease genetics studies provide information on how the genome functions: taken together, these can help us understand the mechanisms by which natural selection has shaped human populations. Using expression data from 48 GTEx tissues, we defined sets of specifically expressed genes (2560 genes per tissue, Wilcoxon p value $< 4E-22$) and looked for evidence that molecular processes in those tissues were more likely to be under selection. We found significant evidence that genes specific to esophagus (p value = $4.9E-05$), stomach (p value = 0.001) and colon sigmoid (p value = 0.03), part of the digestive system, were more likely to be under recent selection than genes specific to our chosen baseline tissue whole blood, consistent with different dietary pressures, as well as genes specific to the cerebellum (p value = 0.003) and cerebellum hemisphere (p value = 0.04) regions of the brain. These regions are central to physical movement, underpinning our development into *Homo erectus*. In contrast, genes specific to the amygdala (p value = $7.5E-24$) and hippocampus (p value = $8.4E-33$) are significantly more likely to be conserved across primates; the amygdala is central to vigilance and threat detection, crucial to survival across species. Both regions are part of the limbic system, hippocampus controls sexual behavior and the ovary's reproductive functions. Our results also allow us to re-evaluate known findings of positive selection: investigating the *LCT* gene which allows Northern Europeans to digest milk and the *EPAS1* gene which aids survival at high altitude, we see the former is specific to the small intestine and the latter to whole blood, artery tissue and lungs. Finally, we used the PhenomeXcan database to identify genes associated to 21 diseases. We found that genes associated with three autoimmune diseases, T1D, Lupus and IBD, were more likely to be under recent positive selection in Europeans than East Asians. Such diseases have also higher incidence in Europeans, explaining their larger effect on survival and reproduction there. This result also suggests that risk alleles for these diseases may be being purged from European populations, highlighting the need for genetic studies to use a wide and diverse range of populations.

Evolutionary and Population Genetics Posters - Thursday

PB2759. Analysis of human disease variants from ancestrally diverse Asian genomes

Authors:

S. Chan¹, Y. Bylstra², P. Tan^{2,3,4}, S. S. Jamuar⁵, J. NGEOW^{1,6}, W. Lim², SG10K_Health Consortium; ¹Natl. Cancer Ctr. Singapore, Singapore, Singapore, ²SingHlth.Duke-NUS Inst. of Precision Med., Singapore, Singapore, ³Genome Inst. of Singapore, Agency for Sci., Technology and Res., Singapore, Singapore, ⁴Precision Hlth.Res. Singapore (PRECISE), Singapore, Singapore, ⁵KK Women's and Children's Hosp., Singapore, Singapore, ⁶Lee Kong Chian Sch. of Med., Nanyang Technological Univ., Singapore, Singapore

Abstract Body:

Despite comprising most of the global population, Asian populations remain poorly characterized in terms of genetic epidemiology of disease given their under-representation in human genomics research. In this study, we characterized clinically significant genetic variation for monogenic disorders and pharmacogenomics from 9,051 human genomes representing populations of East Asian, South Asian and severely under-represented Austronesian-speaking Southeast Asian ancestries. Of the sequenced individuals, 3.4% (309/9,051) harbored a medically actionable pathogenic variant in at least one of the 73 genes in the American College of Medical Genetics and Genomics (ACMG) list for reporting secondary findings (SF v3.0), and an additional 1.6% (148/9,051) of individuals carrying variants of uncertain significance with potential for pathogenic classification upon further clinical or experimental evidence. Twenty-seven percent (10/37) of genes linked to severe recessive disorders with appreciable carrier frequencies in Asians are currently missed by carrier screening panels, and we conservatively estimated 0.5% Asian couples at risk of having an affected child. At the population level, disparate burden of genetic risks attributable to ancestry-specific recurrent variants were observed, however, we also detected individuals harboring pathogenic variants highly specific to ancestries discordant to their self-reported ethnicity, mostly due to cryptic admixture. Beyond genetic disease risk, examination of the pharmacogenomic landscape revealed that 99.7% of sequenced individuals carried at least one actionable pharmacogenetic finding in 23 pharmacogenes with high confidence gene-drug associations; predominantly driven by the high frequency of the Asian-prevalent *VKORC1* rs9923231 allele affecting sensitivity to warfarin. Combining pharmacogenomics with genetic risk assessment revealed that 22.5% (32/142) of Asians predisposed to Centers for Disease Control and Prevention (CDC) Tier 1 genetic conditions would benefit from pre-emptive pharmacogenomics for therapy optimization. Our findings illustrate the genetic variability related to human disease in Asians, providing insights towards understanding genetic epidemiology of disease and improving delivery of precision medicine in a large, diverse and substantially under-represented population.

Evolutionary and Population Genetics Posters - Wednesday

PB2760. Analysis of Latvian genome diversity combining the whole-genome sequencing with genome-wide genotyping data.

Authors:

J. Klovins, R. Rescenko, M. Briviba; Latvian BioMed. Res. and Study Ctr., Riga, Latvia

Abstract Body:

Establishing a population-specific genome variation reference is an obvious step needed to implement genome-based applicability into personalized medicine. As a pilot, to achieve this goal, we have performed whole-genome sequencing of 600 individuals from a population-based cohort included in the Genome Database of Latvian Population (LGDB). We provide the first information on the quality and composition of variants obtained from this experiment. The resulting data were cross-validated with genotypes obtained by an Illumina GWAS chip from the same individuals with high concordance between both platforms. We also included 4500 genotyped samples from the LGDB to evaluate the Latvian population's genetic structure and ancestry admixture. We observed a distinct population stratification of the Latvian sample based on ethnic and regional origin. Self-reported ethnicity correlated to a high degree with the composition of ancestral populations in admixture analysis compared with data from neighboring countries. The group of people with Latvian ethnicity displayed an unusually high proportion of Western Hunter-gatherer ancestry compared to other populations. We also demonstrate significantly improved imputation of low-frequency variants using a haplotype reference panel specific to the Latvian population. This study provides the most extensive to-date analysis of genetic variation in Latvia, providing a background for a public reference resource. Our results show that the genetic composition of the Latvian population is uniquely formed and should be considered an essential factor in future genetic and biomedical studies.

Evolutionary and Population Genetics Posters - Thursday

PB2761. Ancestral genetic tracing defies socio-cultural structure in the north western human populations of India

Authors:

A. Singh^{1,2}, M. Thakur², V. Sahajpal³, S. Kumar¹; ¹Natl. forensic Sci.s Univ., Gandhinagar, India, ²Zoological Survey of India, Kolkata, India, ³State Forensic Sci. Lab., Shimla, India

Abstract Body:

The Himalayas, a mountainous home of varied ethnic populations is one of the most undisturbed regions considering the tough terrain and climatic conditions. Several populations of the north-western Himalayas followed strict social barriers for thousands of years culminating in the present caste endogamy. With this preserved genetic pool based on the caste system, genetic differentiation among the Himalayan populations must follow social division. However, prior migrations and demographic changes may have impacted the genetic makeup of the populations, which must be analyzed on a bigger scale. Based on 20 STRs, we obtained the genetic data of 1005 unrelated individuals from 12 groups from the high elevational locations of India's north-western Himalayas. Most of these populations adhere to stringent marriage traditions within their caste, which reduces the likelihood of shared genetic components leading to population sub-structure. Both Bayesian and non-Bayesian studies, however, indicated a paucity of population substructure, indicating extensive ancestral mixing and panmixia. Regardless of caste or socioeconomic structure, majority of the populations demonstrated close genetic connections. However, when compared to the other groups of the central and eastern Himalayas, the inhabitants of the western Himalayas grouped indicating geographical boundaries of the Himalayan regions. These findings show that the genetic structure of western Himalayan populations is panmictic, defying the ancestral gradient associated with the caste system.

Evolutionary and Population Genetics Posters - Wednesday

PB2762. Ancestral haplotype prediction using neural networks

Authors:

S. Belsare, A. D. Kern; Univ. of Oregon, Eugene, OR

Abstract Body:

Reconstructing the genealogy of a sample of DNA sequences in the form of an Ancestral Recombination Graph (ARG) can give access to the history of that sample, and help gain insights into the biological processes that gives rise to that set of sequences or individuals. Indeed recently, approaches aimed at ARG reconstruction have matured to the point of application in studying the genetics of related individuals and understanding evolutionary events.

A key step in inferring genealogies is generating the ancestral haplotypes at internal nodes in the tree which gives rise to the sample of interest. Existing methods for inferring ancestral haplotypes use heuristic approaches, e.g. generating consensus haplotypes from extant sequences. We present here a deep learning based method to improve ancestral haplotype inference. This method trains neural networks on simulated data of extant and ancestral sequences to predict ancestral haplotype segments. We show that our predicted ancestral haplotype segments from the neural network are more accurate than existing heuristic approaches. Full length ancestral haplotypes are then generated by algorithmic concatenation of the predicted segments. We incorporate our neural network haplotype predictions into existing genealogy inference methods and demonstrate better performance of our predictions by comparing reconstructed trees resulting from these methods with tree comparison metrics.

Evolutionary and Population Genetics Posters - Thursday

PB2763. Ancient trans-species polymorphism at the Major Histocompatibility Complex in primates

Authors:

A. Fortier, J. Pritchard; Stanford Univ., Stanford, CA

Abstract Body:

Genes within the Major Histocompatibility Complex (MHC) are responsible for peptide presentation to T cells, thus playing a central role in immune defense against pathogens. These genes are subject to strong selective pressures including both balancing and directional selection, resulting in exceptional genetic diversity---thousands of alleles per gene! Moreover, some alleles appear to be shared between primate species, a phenomenon known as Trans-Species Polymorphism (TSP) or Incomplete Lineage Sorting, which is rare in the genome overall. However, despite the clinical and evolutionary importance of *MHC* diversity, we currently lack a full picture of primate *MHC* evolution. To start addressing this gap, we used Bayesian phylogenetic methods to determine the extent of TSP at six classical *MHC* genes. We find strong support for TSP in all six genes, including between humans and old-world monkeys in *HLA-DRB1* and even, remarkably, between humans and new-world monkeys in *HLA-DQB1*. Despite the long-term persistence of ancient lineages, we observe rapid evolution at amino acids within the peptide-binding domain. The most rapidly evolving positions are also strongly enriched for autoimmune and infection disease associations. Together, these results suggest complex selective forces arising from differential binding of pathogen peptides, which drive short-term allelic turnover within lineages while also maintaining deeply divergent lineages for at least 40 million years.

Evolutionary and Population Genetics Posters - Wednesday

PB2764. Ancient virome analyses of ancient individuals who lived in the Japanese archipelago

Authors:

L. Nishimura^{1,2}, R. Sugimoto³, H. Kanzawa-Kiriyama⁴, K. Shinoda⁴, I. Inoue^{3,2}; ¹Natl. Inst. of Genetics, Mishima, Japan, ²The Graduate Univ. for Advanced Studies (SOKENDAI), Mishima, Japan, ³Natl. Inst. of Genetics, Mishima, Japan, ⁴Natl. Museum of Nature and Sci., Tsukuba, Japan

Abstract Body:

Ancient viral sequences have been discovered in historical samples such as bones, teeth, and mummified tissues. Various ancient human pathogenic viruses have been discovered since the RNA of ancient influenza viral was analyzed in 1997. Those ancient viruses help elucidate past pandemic events and long-term viral evolution. However, the number of discovered ancient viruses has been limited. Here, we analyzed whole genomic sequencing (WGS) data derived from ancient individuals who lived in the Japanese archipelago for more than thousands of years to discover ancient viruses. We conducted several bioinformatic analyses to detect the ancient viruses from WGS data: de novo assembly to obtain longer sequences, contigs and reads alignment to known viral reference sequences, metagenomic profiling, and non-homologous viral detection using CRISPR immunological memories of host bacteria. Firstly, we recovered an almost complete sequence of Siphovirus contig89 (CT89) by assembling and characterized its genomic features. Based on its genomic characteristic, we compared the ancient CT89 genomes and modern ones to comprehend the long-term evolution of the CT89 virus. The result indicated that the most recent common ancestor of CT89 was around 7,900 years ago. We also detected five ancient viral sequences that did not show homology with modern viral sequences. It might reflect highly diverged or extinct ancient viral genomes. Then, we characterized ancient viromes of each ancient individual based on the reads alignment results and metagenomic profiling. It revealed the differences between ancient and modern Japanese viromes, and it might reflect the different dietary behavior of ancient people. Overall, our results suggested that the ancient viral analyses might be useful for understanding the existence of viruses in the past and long-term viral evolution.

Evolutionary and Population Genetics Posters - Wednesday

PB2765. Between-population genetic differences for human complex traits

Authors:

V. Hivert¹, J. A. Revez¹, M. E. Goddard^{2,3}, N. R. Wray^{1,4}, L. Yengo¹, P. M. Visscher¹; ¹Inst. for Molecular BioSci., The Univ. of Queensland, Brisbane, QLD 4072, Australia, ²Agriculture Victoria Res., Agribio, 5 Ring Road, Bundoora, VIC 3083, Australia, ³Faculty of Vet. & Agricultural Sci., Univ. of Melbourne, Parkville, VIC 3010, Australia, ⁴Queensland Brain Inst., The Univ. of Queensland, Brisbane, QLD 4072, Australia, Australia

Abstract Body:

It is notoriously difficult to draw inference about between-population phenotypic differences using genetic data, because there are many reasons why genetic differences between populations do not lead to phenotypic differences and, conversely, why phenotypic differences are not caused by genetic factors, especially in human populations. Environmental or cultural effects, gene-by-environment and gene-by-gene interactions, and natural selection can all cause observed phenotypic mean differences to deviate from genetic differences. Here, we ask what between-population differences are expected for human complex traits, assuming a simple neutral and additive model of population divergence, in the absence of genotype-by-environment interaction or covariance. Under such a model, the expected mean genetic difference is zero but the variance of the difference is not. When estimating genetic means from Genome-Wide Association Study (GWAS) data, the expected variance of the mean difference in polygenic scores between any pair of populations equal twice the among-population additive genetic variance explained by the markers used in the score. We perform a simulation study to validate theoretical expectations of between-population differences that we derive as a function of parameters we can observe from genetic data, such as the fixation index (F_{ST}) and the within-population heritability of a trait (h^2_w). Under this demo-genetic model, meaningful between-population genetic and, given our assumptions, phenotypic differences are possible. For example, for European and African populations with $F_{ST}=0.14$, and the traits height and schizophrenia (both $h^2_w=0.7$), we would expect the standard deviation of the between-population phenotypic difference for height to be 2.8cm, and the expected range in lifetime prevalence of schizophrenia to differ by more than 100-fold. Finally, we show that our results still hold when simulating a more realistic model of human evolution including different demographic events.

Evolutionary and Population Genetics Posters - Thursday

PB2766. Biological functions and spatial distribution of rapidly evolving brain-expressed genes in the human lineage

Authors:

T. Ajumobi¹, A. DeCasien¹, K. Wagstyl², A. Raznahan¹; ¹NIMH Intramural Res. Program, NIH, Bethesda, MD, ²Wellcome Ctr. for Human Neuroimaging, Univ. Coll. London, London, UK, London, United Kingdom

Abstract Body:

Distinct features of the human brain are likely to reflect evolutionary modifications of human genome structure (e.g. DNA sequence) and function (e.g. gene expression) in the human lineage. However, it remains unclear how these two modes of evolution are related at the gene-level and whether genes that differ in these signatures also differ in their neurobiology. Here, we estimated sequence divergence (SD) and expression divergence (ED) for N=7108 genes with 1:1 homologs between humans and chimpanzees and human and rhesus macaques (Aim 1); compared the functions of genes with high SD or high ED (Aim 2); identified genes that may have been particularly critical to human brain evolution as those exhibiting simultaneously high SD and ED (Aim 3); and provided functional and spatial annotations for these genes (Aim 4). We estimated: i) sequence divergence as the rate of nonsynonymous to synonymous changes (dN/dS) in the human lineage (since our last common ancestor with chimpanzees); and ii) expression divergence as the average difference in expression level between human vs. chimpanzee and human vs. macaque brain tissues. We found that genes tend to exhibit either high SD or high ED ($\rho < 0.2$, Aim 1), genes exhibiting brain-specific expression patterns exhibit lower SD but higher ED than non-brain-specific genes, high ED genes are associated with brain development while high SD genes are associated with general biological processes (Aim 2). We then identified N=99 genes with simultaneously high SD and ED (Aim 3) and found that these genes are collectively most highly expressed in cortical areas that have undergone most marked expansion during human evolution ($\rho = 0.432$; $p_{\text{spin}} = 0.027$) (Aim 4). Furthermore, these high ED and high SD genes are enriched within spatially derived co-expression networks that are associated with radial glial cells and autism. These findings indicate that changes in gene sequence and gene expression represent two largely dissociable evolutionary signals in the human lineage and intersect with different neurobiological processes. The few genes that show both these evolutionary signatures are highly expressed in cortical regions that have undergone marked evolutionary expansion and are associated with early brain development and genetic risks for neurodevelopmental disorders.

Evolutionary and Population Genetics Posters - Wednesday

PB2767. CAFE: Ancestry and genotype calling uncertainty-adjusted ancestry-specific allele frequency estimation from admixed samples.

Authors:

J. Wang¹, S. Zöllner^{1,2}; ¹Dept. of Biostatistics, Univ. of Michigan, Ann Arbor, MI, ²Dept. of Psychiatry, Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Estimates of allele frequency distributions across different populations provide important insights into the transferability of GWAS results, population-specific drivers of disease etiologies, and interpopulation demographic histories. Allele frequencies from recently admixed groups are often more difficult to interpret in this context due to heterogeneous ancestry background of each allele. Existing methods proposed to deconstruct the allele frequencies of the underlying ancestral groups from admixed samples are typically based on “best-guess” local ancestries and genotypes. These methods ignore uncertainty in local ancestry and genotype calling, which can bias the allele frequency estimates especially for rare variants and can overestimate the differentiation among ancestral groups.

Here, we present CAFE, a novel local ancestry based method for estimating ancestry-specific allele frequency, which adjusts for ancestry and genotype calling uncertainty. CAFE is an Expectation-Maximization (EM) algorithm and it informs allele frequency estimates by averaging individual genotype dosages characterized by probabilistic local ancestry calls, genotype calling uncertainties and estimated allele frequency distributions. CAFE is capable of interpreting large-scale ($N > 10,000$) phased or unphased data generated from array-based or sequence-based genotyping techniques and of multi-way admixture patterns. In simulation studies, we evaluated the impact of allele frequency distribution, population structure, genotyping error rate, sequencing error rate and depth, and sample size on the estimates obtained from CAFE and from methods not adjusting for the uncertainty. The results demonstrate high estimation accuracy of CAFE in both array-based and sequence-based genotyping data (Pearson’s R with true values > 0.999). The simulation studies also demonstrate that ignoring the uncertainty inflates bias in estimating allele frequencies 16-fold and 9-fold on average from array-based genotyping or low-depth sequencing data respectively, even with corresponding low error rates ($\sim 0.1\%$) and large sample size ($\sim 10K$). We further applied CAFE to 1000 Genomes sequencing data to estimate African allele frequencies and European allele frequencies from admixed African samples. The resulting allele frequency estimates are highly correlated with allele frequencies for their corresponding superpopulations (Pearson’s R, AFR = 0.995; EUR = 0.965). Thus, we provide a new, computationally efficient method that provides precise estimates of ancestral allele frequencies and can thus support the analysis of samples from recently admixed populations.

Evolutionary and Population Genetics Posters - Thursday

PB2768. Cardiovascular associations and signals of selection at the *Adrenergic Receptor Alpha-1A* gene (*ADRA1A*) in high-altitude Andeans.

Authors:

E. Moya, J. Reeves, W. Gu, E. Lawrence, J. Hall, T. Simonson; Univ. of California San Diego, La Jolla, CA

Abstract Body:

People living at high-altitude exhibit respiratory and cardiovascular adaptations to the stressful environment produced by hypoxic conditions that may be associated with genetic adaptations. Mechanisms involving adrenergic pathways are especially relevant, since adrenergic receptors have been involved in regulation of cardiovascular consequences observed in animal models exposed to hypoxia. We hypothesize that variants in the *Adrenergic Receptor Alpha-1A* gene (*ADRA1A*) underlie cardiovascular phenotypes in a high-altitude Andean population. We localized adaptive signals based on the composite of multiple selection scan, which demonstrated selection at the *ADRA1A* locus in both Andean and Tibetan populations. We identified a missense, non-synonymous variant (rs1048101, A>G, p.[Arg347Cys]) within the second exon of *ADRA1A* and examined its associations with cardiovascular phenotypes in an Andean population (N = 194). We found that the putatively adaptive A allele of rs1048101 is enriched in Peruvians compared to the rest of the world (both Andeans and Peruvian A allele frequency = 55%) and is associated with a lower chronic mountain sickness score ($p < 0.045$) and a higher heart rate in male participants when breathing room air ($p < 0.021$). We conclude that additional copies of the putatively adaptive allele A at rs1048101 are linked to beneficial physiological traits, such as lower prevalence of chronic mountain disease and increased heart rate, suggesting a potential adaptive role involving the *Adrenergic Receptor Alpha-1A* gene. Future studies in controlled models are necessary to establish the direct cellular and physiological consequences of this variant under hypoxic conditions.

Evolutionary and Population Genetics Posters - Wednesday

PB2769. Challenging continental groupings of humans using a data-driven approach.

Authors:

M. Palma Martínez¹, Y. Posadas García², C. Quiroz López³, T. Lasisi⁴, K. A. Bird⁵, A. Zaidi⁶, M. Sohail¹; ¹UNAM, Cuernavaca, Mexico, ²UASLP, San Luis Potosí, Mexico, ³ENAH, Ciudad de México, Mexico, ⁴Univ. of Southern California, Los Angeles, CA, ⁵Michigan State Univ., East Lansing, MI, ⁶Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Over the last decades, population geneticists have explicitly pushed back against the claim that humans can be grouped into commonly defined racial categories on the basis of genetic variation. Current approaches have used genetic ancestry as a grouping alternative. However, reliance on genetic ancestry as defined by continental groupings can lead to approaches that recapitulate racial thinking (Lewis et al., 2022). Here we use a data-driven approach to challenge the validity of continental groupings.

We reanalyze 1000G project and HGDP data using previously reported genetic metrics by the 1000G consortium. We show that many genomic signatures of interest are not structured by continental groupings. For example, the number of singletons found in 1000 Genomes populations showed patterns that superseded continental groupings (e.g. Finnish vs. British, Luhya vs. Yoruba, etc.).

The idea of continental groups is sometimes justified with ADMIXTURE-type analyses, where the inferred ancestry components are thought to represent ‘pure’ ancestral groups. We push against this framing by first showing that ADMIXTURE components do not represent ‘true’ underlying groups but are rather a function of the sample being analyzed. For example, we show that in the 1000G, ADMIXTURE detects within-continent substructure even before the emergence of all major continental labels. We also show that some components are shared across continents as exemplified by populations from Europe, South, and Central Asia (K=7). To further illustrate how sampling scheme affects ADMIXTURE inference, we show if only one or a few individuals from the JPT subpopulation are included in the analysis, the ADMIXTURE component corresponding to the JPT populations is no longer detected. Furthermore, inclusion of samples from underrepresented regions such as Africa and Asia leads to the emergence of new clusters, representing substructure inside continental groups. This also changes the order at which other pre-existing components are detected. This highlights the fact that ADMIXTURE results, although useful to understand population structure, represent statistical properties of the data rather than some ground truth about continental ancestry.

Our results emphasize the fact that population genetic variation tracks differences in demographic history rather than arbitrarily defined continental groups and that such groups are often not useful for applications in human genetics. We also caution against the use of typological thinking whereby reference populations within a continent are treated as “type specimens” that accurately represent all other populations within that continent.

Evolutionary and Population Genetics Posters - Thursday

PB2770. Characterizing the origin of African ancestral tracts in admixed individuals of the UK Biobank

Authors:

N. N. Shah, T. Tan, G. E. Tietz, E. G. Atkinson; Baylor Coll. of Med., Houston, TX

Abstract Body:

High-throughput genome sequencing has allowed for fine characterization of genetic ancestry leading to better understanding of the complex demographic and migratory history of diverse human populations. Recent genetic studies have elucidated and confirmed the historical records of major demographic events and the timeline of when admixture occurred for multiple admixed American populations. However, limited work has thus far investigated the ancestral origins of African ancestry in admixed European populations.

Here, we characterize the geographic origin and demographic history of the African components of modern admixed individuals of the UK Biobank (UKB) to sub-continental resolution. Using a subset of ~4,500 two-way admixed African-European UKB individuals, we explore the population structure of these individuals using both allele frequency and local ancestry informed methods, and construct a demographic model for this population. We performed PCA and ADMIXTURE analyses, followed by ancestry-specific PCA (ASPCA) to identify the fine-scale geographic areas within the African continent from which African local ancestry tracts in our admixed sample derive. We conducted ASPCA in two steps - first, PCAdmix was used for local ancestry estimations, and then each ancestry was independently masked to be projected against a reference population of continental African and European populations from the 1000 Genomes Project and African Genome Variation Project. We observed highly variable proportions of African ancestry, with an average of ~56% of West-Central and West African ancestry. To characterize the demographic history of these admixed African descent UKB participants, we analyzed the length of African ancestry tract segments to understand how recently in the past pulses of admixture occurred. We did this by testing different demographic models fitting a variety of potential migration histories using the Tracts software program, which utilizes local ancestry inference calls.

In sum, we quantified and investigated that African haplotypes in a subset of admixed UKB individuals, tracing the bulk of tracts in this cohort back to the West-Central African ancestral origins, and identified the optimal demographic model for two-way admixed African-European individuals in the UKB. This knowledge sheds light on the fine-scale ancestral history of modern European individuals of the African diaspora. Further it will help inform the anticipated patterns of genetic variation in diverse admixed cohorts which is vital for optimal benchmarking and calibration of statistical methods.

Evolutionary and Population Genetics Posters - Wednesday

PB2771. *Cis*-regulatory genetic variants influence distinct expression profiles of adipose and muscle tissue transcripts between African and European ancestry individuals.

Authors:

C. Langefeld, M. Comeau, S. Das; Wake Forest Univ. Sch. of Med., Winston-Salem, NC

Abstract Body:

Genetic regulation of transcription plays a key role in the etiology of many diseases. Genetic regulatory variants may contribute to the ethnic disparities in disease predisposition by generating variability in transcript expression among different populations. Here we utilized adipose and muscle tissue expression data to identify differences in gene expression between individuals of African and European ancestry. European Americans (EA, N=99) and African Americans (AA, N=36) from the Arkansas metabolic study cohort were simultaneously metabolically phenotyped, molecularly profiled (transcript expression by Illumina HT12-V4 chip and RNA-seq), and genotyped (Illumina Infinium Omni5Exome-4 Chip) using the same methods. Self-identified race and genotype-defined admixture were used as predictors in two different linear regression models adjusting for age, sex, and BMI. We identified 505 and 297 genes in adipose and muscle tissue, respectively; exhibiting differential expression between AAs and EAs or associated with African admixture (FDR<0.05). Increased expression of nine genes (e.g., GSTM3, GSTT2B, CRYBB2) and lower expression of 20 genes (e.g., NARS2, DHRS4, SLC25A26) was observed in both adipose and muscle tissue of AAs compared to EAs. Pathway enrichment analysis suggested decreased expression of mitochondrial genes and higher xenobiotics metabolism genes in AAs. Expression quantitative trait analysis identified *cis*-regulatory SNPs (*cis*-eSNPs) for a subset of these transcripts (FDR<0.05) in the EA cohort or AAs from AAGMEx (N=256). To test the effect of genetic variants on differential expression of transcripts between African and European ancestry individuals, we recomputed linear regression by including the top *cis*-eSNP genotypes for each transcript as covariates in the model. Adjustment for *cis*-eSNPs removed statistical significance of differential expression >2 log-scale for 144 (64.9%) and 65 (46.4%) transcripts in adipose and muscle, respectively. Current work is exploring the impact of both allele frequency and linkage disequilibrium on the observed patterns. In summary, this study suggests a significant contribution of *cis*-eSNPs in determining the differential expression of transcripts between population groups.

Evolutionary and Population Genetics Posters - Thursday

PB2772. Coding variants associated with healthy aging: Findings from the Long Life Family Study (LLFS)

Authors:

T. Gunasekaran¹, R. Cheng^{1,2}, B. Thyagarajan³, A. I. Yashin⁴, T. T. Perls⁵, J. M. Zmuda⁶, M. A. Province⁷, S. Cosentino^{1,2}, N. Schupf^{1,2}, J. H. Lee^{1,2}, B. Vardarajan^{1,2,8}; ¹The Gertrude H. Sergievsky Ctr., Columbia Univ., New York, NY, ²Taub Inst. for Res. on Alzheimer's Disease and the Aging Brain, Columbia Univ., New York, NY, ³Dept. of Lab. Med. and Pathology, Univ. of Minnesota, Minneapolis, MN, ⁴Duke Univ., Durham, NC, ⁵Boston Univ. Sch. of Med., Boston, MA, ⁶Dept. of Epidemiology, Univ. of Pittsburgh, Pittsburgh, PA, ⁷Washington Univ. Sch. of Med., St. Louis, MO, ⁸Dept. of Neurology, Columbia Univ., New York, NY

Abstract Body:

Background: Exceptional longevity (EL) is a complex trait that is likely the result of an interaction between environmental, behavioral, genetic and stochastic factors. Over 20 genetic loci have been associated with healthy aging and several were implicated in aging-related diseases. Identifying the genetic determinants of EL could also lead to the discovery of genes that affect the rate of aging and predisposition for or against age-related diseases. Such discoveries would be helpful in the prediction, prevention and treatment of age-related diseases as well as predictors of and strategies for healthy aging. **Methods:** We selected 420 unrelated individuals of European ancestry from the Long Life Family Study (LLFS) families (one per family). Individuals were selected with the least summary disease score (SDS) defined as the sum of the presence (1) or absence (0) of phenotypes in eight health domains - heart disease, stroke, lung disease, hypertension, diabetes, peripheral artery disease, cancer and dementia. Elderly individuals of the same genetic background were selected from Washington Heights and Inwood Community Ageing Project (WHICAP) as controls. Whole exome sequencing (WES) was conducted followed by joint variant calling, QC and annotation. We conducted single marker and gene-burden analysis to identify variants enriched for EL in LLFS cohort and compared minor allele frequencies (MAF) with variants in gnomAD. **Results:** Single marker analysis identified rs28357075 in *C4A* ($p=2.95E-06$, MAF=0.13), rs201667800 in *FAM185A* ($p=5.35E-06$, MAF=0.19), rs1809279 in *NPEPPS* ($p=3.69E-04$, MAF=0.07) and rs140030732 ($p=3.75E-04$, MAF=0.009) in *TET1* associated with EL. In gene-based rare variant analyses, *HLA-DRB1* was the strongest association (nvariants=11, $p=3.89E-13$) detected with EL phenotype. In addition, associations in *PABPC1* (nvariants=4, $p=6.47E-13$) and *PRKRA* (nvariants=8, $p=6.25E-12$) genes were detected at genome-wide significance threshold. Genes with variants enriched in the LLFS samples were involved in exonuclease activity and galactose metabolism pathways. Genes depleted in EL were involved in disease risk pathways including auto-immune diseases, diabetes, peptidase activity and response to stress. **Conclusion:** WES in individuals with EL identified multiple genetic loci that have been previously associated with longevity, cancer and aging related phenotypes such as Alzheimer's Disease. This suggests that the genes identified in EL are likely to confer protection against the common diseases and eventually promoting longer life span.

Evolutionary and Population Genetics Posters - Wednesday

PB2773. Comparing human centromeres with TandemAligner

Authors:

A. Bzikadze, P. A. Pevzner; UC San Diego, La Jolla, CA

Abstract Body:

Recent advances in long-read sequencing technologies led to rapid progress in centromere assembly. We previously presented *centroFlye* — the first tool for centromere assembly from long error-prone reads. In 2019, we joined efforts with the Telomere-to-Telomere Consortium that, in Apr 2022, published a landmark achievement — the complete sequence of a human genome that includes sequences of all centromeres. This progress opens a possibility to address the long-standing questions about the centromere architecture. Centromeres are formed by *extra-long tandem repeats* (ETRs) representing some of the most rapidly evolving regions of the human genome. Since variations in centromeres are linked to cancer and infertility, comparison of centromeres is an important task for understanding their evolution and genomic diversity. A common approach to variant calling is based on mapping reads that originated from the query genome to the target sequence. Provided assembled sequences of centromeres, the difficult problem of calling variants in centromeres is substituted by a seemingly simpler problem of alignment of their sequences. Surprisingly, there are no algorithms that approach that problem. At first glance, it appears that similarities between centromeres can be revealed by the standard sequence alignment based on dynamic programming. However, we show that it is an inadequate computational model for centromeres analysis since the highest-scoring pairwise alignment does not reveal the evolutionary events that made two centromeres different.

We propose the TandemAligner algorithm to address the ETR comparison problem and illustrate its performance using the centromere on chromosome X from a haploid cell line CHM13 (CHM13X) and from a male individual HG002 (HG002X). We define the *count* of a substring P in a string S as the number of occurrences of P in S and say that P is *rare* if its count in S does not exceed a defined threshold, and *frequent*, otherwise. Since ETRs are rich with frequent substrings, selecting a correct match between these substrings in two ETRs is challenging. Indeed, if the count of a substring P in string S (in string T) is $N(M)$, the total number of matches is $N*M$. Instead, TandemAligner introduces a novel alignment scoring that prioritizes matches of rare substrings since they are more likely to be relevant to the evolutionary relationship between ETRs. It further finds the optimal alignment with respect to this scoring and attempts to reconstruct the evolutionary scenario that led to generation of CHM13X and HG002X from their common ancestor. We show that the mutation rate across centromeres substantially exceeds the average rate across the human genome.

Evolutionary and Population Genetics Posters - Thursday

PB2774. ConsensuSV - HPC-ready, ML-enhanced automated pipeline for Illumina-based variant detection.

Authors:

M. Chilinski^{1,2}, D. Plewczynski^{2,1}; ¹Warsaw Univ. of Technology, Warsaw, Poland, ²Univ. of Warsaw, Warsaw, Poland

Abstract Body:

We have developed two bioinformatics pipelines: ConsensuSV-core, and ConsensuSV-pipeline for the identification of several types of mutations in human DNA sequences of the whole genomes. The first algorithm serves the purpose of merging the calls from multiple, independent Structural Variant (SV) callers using consensus approach. For that purpose, we are using deep neural network, that we pretrained on New York Genome Center (NYGC) high-quality Illumina data. The algorithm itself is flexible, and can be easily retrained for any other sequencing method (for example Oxford Nanopore Sequencing, or PacBio), and other Structural Variant callers can be easily added. The second algorithm serves as a wrapper for the first one - it is high performance computing (HPC) ready pipeline for variant detection developed using luigi framework. It is completely automated, and incorporates using raw Illumina fastq data and aligning it accordingly to the Human Genome Structural Variation Consortium (HGVS) alignment recommendations. After the alignment step, 8 gold-standard SV callers are utilised, along with SNP and Indel caller. Finally, it merges the all sets of SVs using our ConsensuSV-core, that was pretrained on two trios, and tested on third one from 1000 Genomes project (CHS, PUR, YRI). ConsensuSV-pipeline is open source software, and comes with Docker image (as well as Dockerfile), and complete documentation allowing super easy setup on personal computers and HPC centres. Availability: <https://github.com/SFGLab/ConsensuSV-pipeline> <https://github.com/SFGLab/ConsensuSV-core>

Evolutionary and Population Genetics Posters - Wednesday

PB2775. Conservation and function of the *NXF2* palindrome in male fertility.

Authors:

A. Lawson, J. Mueller; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Large (>10kb), nearly identical (>99% nucleotide identity) palindromes harbor genes with testis-biased expression and are enriched on the mammalian X and Y chromosomes. Previous studies have identified the independent acquisition of large X- and Y-linked palindromes; however, whether some X-palindromes are deeply conserved across mammals remains unclear. We hypothesize some mammalian X-palindromes are as old as the origin of the sex chromosomes and persisted to benefit male fertility. To determine the extent of X-palindrome conservation across mammals, we are sequencing and assembling putative palindromic regions across seven mammalian species based upon their known conservation across primates. To study the function of ancient palindrome-associated genes, we selected nuclear export factor 2 gene *Nxf2*, a gene family harbored in a X-palindrome conserved across mammals. Mouse *Nxf2* is necessary for male fertility, demonstrated by a previous study of *Nxf2*^{-Y} mice exhibiting age- and strain-dependent male germ cell depletion due to an unknown mechanism. We hypothesize mouse *Nxf2*-related sub/infertility is caused by failure to silence transposable elements (TE). This hypothesis is based on the observation that fly *nxf2* is necessary for TE silencing and mouse *Nxf1* suppresses transcripts harboring TEs. To study the role of mouse *Nxf2* in TE regulation we re-created the previously characterized *Nxf2*^{-Y} mice to assess TE activity and generated NXF2-3X-FLAG mice to assess the molecular mechanism by which *Nxf2* disruption results in spermatogenic defects. Understanding the extent of X-palindrome conservation throughout mammals will inform us to how large palindromes have shaped sex chromosome evolution and acquired roles in spermatogenesis.

Evolutionary and Population Genetics Posters - Thursday

PB2776. Continuous measures of genetic diversity in biomedical research.

Authors:

A. Dixit, A. Mañas, L. Seninge; Coral Genomics, San Francisco, CA

Abstract Body:

The use of discrete race or ethnicity categories in biomedicine and, more specifically, genetics research has a long and complex history. Their use risks reinforcing the false notion that there are a set of discrete biological types that map directly onto historical race categories. A possible alternative, would be to ignore such categories altogether. Such an approach risks exacerbating health disparities by leading to biomedical advances with poor applicability of predictions or interventions in diverse patient populations. The preceding points lead to a dilemma for researchers who want to advance biomedical research in an equitable, socially aware manner. A metric is presented for characterizing the representation of human diversity in clinical research without relying on categorial variables. A related metric is proposed for evaluating the performance of biological advances. Both metrics apply in the specific context that genetic data is available on the study population.

Our derivation of continuous metrics is based on a reference population that is as representative of global human diversity as possible (ex: panels incorporating 1000 Genomes, the All of Us initiative, UK Biobank, and H3Africa). The genetic diversity of the reference panel is encapsulated in a minimal set of orthogonal axes. These axes are constructed to be robust to genotyping platform. For each axis, the representation of diversity is compared to the study population and a continuous measure of the difference is derived based on information theory.

We apply this algorithm to both synthetic datasets and hundreds of previous genetic studies to quantify representation and diversity on a continuous basis for the first time at scale. Finally, we apply these diversity axes towards evaluating the performance of polygenic risk scores without resorting to categorical race or ethnicity labels.

Evolutionary and Population Genetics Posters - Wednesday

PB2777*. Contrasting patterns of positive selection at *Plasmodium falciparum* and *Plasmodium vivax* associated loci in Oman.

Authors:

P. Haffener¹, A. Z. Al-Riyami², S. Zadjali³, G. B. J. Busby⁴, M. Al-Rawahi⁵, S. Al Hosni², A. Al Marhoobi⁵, A. Al Sheriyani⁶, E. M. Leffler¹; ¹Univ. of Utah, Salt Lake City, UT, ²Sultan Qaboos Univ. Hosp., Muscat, Oman, ³Sultan Qaboos Comprehensive Cancer Ctr., Muscat, Oman, ⁴Allelica, New York, NY, ⁵Sultan Qaboos Univ., Muscat, Oman, ⁶Royal Oman Police Hosp., Muscat, Oman

Abstract Body:

Plasmodium falciparum and *P. vivax*, the two most common causes of malaria in humans, utilize distinct molecular strategies and have driven different human adaptations. *P. falciparum* tends to cause higher parasitemia resulting in more severe disease and is suggested to be a stronger selective pressure than *P. vivax*. Nevertheless, the Duffy null allele that removes an essential receptor for *P. vivax* invasion from red blood cells has strong evidence for natural selection in African and admixed African populations. Other variants have been associated with susceptibility to severe *P. falciparum* malaria. Like the Duffy null allele, these are more common in African populations and in populations with African admixture. Oman has had a long history of contact with African populations, in part due to East African settlements and trade of the Omani Empire. In addition, Oman has recently struggled with endemic *P. falciparum* and imported *P. vivax* malaria. To learn about the relative impact of the two parasite species in human populations, we aimed to contrast evidence for selection on loci associated with either *P. falciparum* or *P. vivax* malaria in Oman after their introduction through genetic admixture with African populations. We conducted whole genome sequencing of DNA collected from 100 healthy Omani blood donors and merged the resulting variant calls with the 1000 Genomes Phase 3 variant call set. We first assessed global ancestry proportions and found that, combined across all Omani genomes, about 1/5 of genetic ancestry was derived from African populations. To determine whether malaria-associated loci were introduced via admixture, we inferred local genetic ancestry and searched for significant deviations from genome-wide expectations. Although we did observe African ancestry on some haplotypes carrying alleles associated with protection from *P. falciparum* malaria, none displayed a significant excess of African ancestry relative to the genome-wide average. In contrast, we found that the Duffy null allele was at high frequency in Omanis (86%) and that this locus harbored the greatest excess of African ancestry across the entire genome. Interestingly, other tests for selection also did not reveal any signal at *P. falciparum*-associated loci, indicating that *P. falciparum* was likely not a strong selective pressure following historical admixture. We conclude that the excess of African ancestry at the Duffy null allele is the result of positive selection in response to pressure from *P. vivax* malaria and that the historical prevalence of *P. falciparum* may have been too low to provide sufficient protection to overcome the fitness costs of *P. falciparum*-associated alleles.

Evolutionary and Population Genetics Posters - Thursday

PB2778. *De novo* assembled and phased human genomes from Persian Arab trios show divergent and novel sequence versus CHM13 and GRCh38, providing valuable population specific reference genomes for middle eastern region.

Authors:

M. Ghorbani¹, I. Alazwani¹, S. Moosa¹, R. Razali², Z. Siddig¹, R. Mathew¹, M. Al Hashmi¹, A. El Khouly¹, S. Poolat¹, S. Tomei¹, K. Fakhro^{1,3,4}, E. Eichler⁵, Y. Mokrab^{1,3}; ¹Human Genetics Dept., Sidra Med., Doha, Qatar, ²Dept. of BioMed. Sci., Coll. of Hlth.Sci., Qatar Univ., Doha, Qatar, ³Dept. of Genetic Med., Weill Cornell Med., New York, NY, ⁴Coll. of Hlth.and Life Sci., Hamad Bin Khalifa Univ., Doha, Qatar, ⁵Washington Univ., Dept. of Genomic Sci., Seattle, WA

Abstract Body:

The genetic diversity of the Middle Eastern region is poorly captured in GRCh38 and the various reference genomes published to date. Genomic analysis of SNV and SV calls for both disease association studies and rare gene discovery were shown to be improved significantly using population-specific genomic references and annotations. Based on three family trios with Persian Arab ancestry from Qatar, we built *de novo* reference genomes using a combination of PacBio long-read sequencing (50-70X), Bionano optical mapping (70X), and Illumina short-read sequencing (30-60X). For four individuals from these trios, we generated high quality draft assemblies for both collapsed (N50 = 59.5-79.2 Mb) and haplotype-resolved using trio-based phasing (N50 = 10.3-22.7 Mb). Haplotypes were resolved accurately using independent assessment of assemblies by parental inherited short reads k-mers, showing 95~99% of k-mers being correctly placed in each haplotype. Also, high QV (quality value) values were achieved for the assemblies' haplotypes (range 34-38) which is equivalent to >99.9% accuracy. In terms of essential gene completeness, these assemblies contain >98% of essential genes from BUSCO analysis. In addition, our collapsed genome results scaffolds contiguity is comparable to that of GRCh38. Relative to the latest CHM13 T2T and GRCh38 references, our assemblies show novel sequence up to 15.2 Mb. Up to 24K SV calls were made versus GRCh38 and 22K SVs versus CHM13 T2T highlighting divergence relative to these references. This highlights the need for such references as population-specific tools to advance the understanding of genetic variants implicated in disease in the Middle Eastern region and beyond.

Evolutionary and Population Genetics Posters - Wednesday

PB2779. Detection of Genomic Regions undergoing Selection in post-Neolithic Transition

Authors:

C. Bhattacharyya¹, P. Ghosh¹, S. K. Paine¹, A. Basu²; ¹Natl. Inst. of BioMed. Genomics, Kalyani, India, ²Natl. Inst. of BioMed. Genomics, Kalyani, India

Abstract Body:

From early domestication of plants to a full-fledged settled agriculturist society, human populations have undergone several environmental and demographic changes. Altogether those changes might have put evolutionary pressure on the people who have adapted to the complex and sedentary agricultural lifestyle. Here, we hypothesize that people who adopted the practice of farming and formed large complex societies; manifest a footprint of recent positive selection in their genomes that has been advantageous in the changing environment. From the data generated in GenomeAsia100k Consortium, we selected 171 individuals from 15 large populations which are sedentary, complex in social relationships, dependent on an agrarian economy. We consider these populations as ones who were early to transform to an agrarian lifestyle. In contrast to these populations, we selected 166 individuals from 17 extant tribal populations who do not exhibit any of the above features of agrarian lifestyle, even today. To explore the signatures of selection in the trait(s) pertinent in transition to farming cultures, we have employed two types of test: 1) Population Branch Statistic (PBS), that identifies variants whose frequency has significantly altered in 'agriculturist' compared to 'non-agriculturist' when compared against an outgroup population 2) cross-population Extended Haplotype Homozygosity (XP-EHH), which identifies recently selected long-haplotypes among 'agriculturists'. After stringent filtering, we identified 347 variants with PBS scores above 99th percentile, clustering into 29 unique 1MB regions of the genome and 1603 variants with significantly high post-adjustment XP-EHH score (more than 4) clustering into 7 genomic regions. Significant genes from the regions identified through PBS are COL12A1, CALCR, KITLG, ANO5, NELL1; which are associated with mineralization of bones and protection from osteoporosis. Our top hit region identified through XP-EHH contains 894 variants (mean adjusted score 5.9, SD 1.5), is chr15:48416360-48891965, spanning the genes SLC12A1, DUT, FBN1 which strongly impacts kidney function. SLC12A1 is a major sodium potassium chloride cotransporter, mostly expressed in the thick ascending limb of the loop of Henle. We propose that changes in genes involved in mineralization of bones and reabsorption of salt can be in response to sedentary agrarian life and salt-rich diet.

Evolutionary and Population Genetics Posters - Thursday

PB2780. Did Neanderthals have different 3D genome folding from modern humans?

Authors:

E. McArthur¹, D. Rinker¹, E. Gilbertson², G. Fudenberg³, M. Pittman², K. Keough⁴, K. Pollard⁵, J. Capra²; ¹Vanderbilt Univ., Nashville, TN, ²Univ. of California San Francisco, San Francisco, CA, ³Univ. of Southern California, Los Angeles, CA, ⁴Gladstone Inst.s, San Francisco, CA, ⁵Gladstone Inst.s / UCSF, San Francisco, CA

Abstract Body:

Introduction: Changes in gene regulation were a major driver in the divergence of archaic hominins--Neanderthals and Denisovans--and modern humans. The three-dimensional (3D) folding of the genome is critical for regulating gene expression; however, its role in recent human evolution has not been explored because the degradation of ancient samples does not permit experimental determination of archaic 3D genome folding.

Hypothesis: We hypothesize that the 3D genome differed between modern and archaic hominins and that these differences can be revealed by reconstructing the genome folding via machine learning methods and experimental Hi-C. Furthermore, we hypothesize that these 3D differences contributed to phenotypic divergence in the evolution of modern humans.

Results: We apply novel deep learning methods to infer 3D genome organization from DNA sequence to Neanderthal, Denisovan, and 20 diverse modern human genomes. Comparing the resulting 3D contact maps across the genome, we identify 167 distinct regions with diverged 3D genome organization between archaic and modern humans. We show that these 3D-diverged loci are enriched for genes related to immune response, cognition, and morphology of the eye, supra-orbital ridges, hair, and lungs. Despite these specific diverged loci, the 3D genome of archaic and modern humans are more similar than expected based on sequence divergence. This suggests that the pressure to maintain 3D genome organization constrained hominin sequence evolution. Finally, we identify loci where modern Eurasians inherited novel 3D genome folding patterns via introgression from archaic ancestors that associate with traits including height, fat distribution, and blood pressure. Our results provide a putative molecular mechanism for the phenotypes associated with these introgressed haplotypes. To validate modern human loci with hypothesized archaic 3D genome folding patterns, we conduct experimental HiChIP on diverse lymphoblastoid cell lines with and without Neanderthal ancestry.

Conclusion: Using deep learning to predict archaic 3D genome organization, we illustrate the potential of inferring molecular phenotypes from ancient DNA to reveal previously unobservable biological differences contributing to human evolution and disease.

Evolutionary and Population Genetics Posters - Wednesday

PB2781. Discovering determinants of short tandem repeat stability in local sequence context using gradient-based attribution scores.

Authors:

R. DeVito, M. Gymrek; Univ. of California San Diego, La Jolla, CA

Abstract Body:

Short tandem repeats (STRs), tracts in DNA where a short motif is repeated, are among the most polymorphic and mutation prone loci in the human genome. Only around 19% of STRs are stable, meaning they have a fixed copy number, while the remaining unstable STRs vary in copy number or have base substitutions in the population. This variation has been shown to affect a variety of traits, including gene expression, monogenic disorders such as Huntington's Disease, and complex traits. Multiple studies have identified total repeat copy number as the primary determinant of STR stability. While other determinants of STR stability have been identified, including GC content and recombination rates, these explain relatively little variation in mutation rates across loci and have shown inconsistent associations across studies.

In this work we explored whether local sequence around an STR may help explain differences in STR stability across loci. Using sequence data from the reference genome and stability labels from 2,504 unrelated 1000 Genomes samples, we trained a 1-D convolutional neural network to predict STR stability from local sequence context with a balanced set of 37,198 dinucleotide STRs of lengths 12 and 13 bp. Our best model achieved auROC 0.82 on an orthogonal test set of STRs. We then used integrated gradients, a gradient-based attribution score, to identify sequence features contributing to our model's performance.

Through analysis of these gradient-based scores, we identified three repetitive patterns in the sequences directly flanking an STR that are significantly correlated with stability (rank-biserial correlations of 0.24, 0.20, and 0.11). All of these determinant patterns involve the base at the boundary of the STR and maintaining a pattern with a periodicity of 2 or 1. For example, an AC repeat followed by a sequence that continues the "C" at every other base pair is more likely to be unstable. Overall, our study highlights a strong and previously unreported impact of the local sequence immediately adjacent to STRs as an important contributor to STR stability. These findings will help develop improved models of STR mutation properties and can give new insights into STR mutation mechanisms.

Evolutionary and Population Genetics Posters - Thursday

PB2782. Disease-associated STR loci in diverse populations

Authors:

C. Steely¹, N. Russell², W. Watkins², L. B. Jorde²; ¹Univ. of Utah, Sch. of Med., Salt Lake City, UT, ²Univ. of Utah, Salt Lake City, UT

Abstract Body:

Short tandem repeats (STRs) are tandemly repeated sequences of 1-6 bp motifs. STRs compose only 3% of the genome, but they have been frequently used in studies of population genetics and evolution, as well as in forensic analyses due to their high heterozygosity. Normal population variation at some of these loci has also been associated with variation in gene expression. Expansion of STRs in certain loci (both intronic and exonic) has been linked to over 50 human diseases, including Friedreich ataxia, Huntington disease, amyotrophic lateral sclerosis (ALS), and fragile X syndrome. Using publicly available data from gnomAD and the genomes of 300 individuals included in the Simons Genome Diversity Project, we analyzed 57 STRs that have been associated with disease. The genomes included in gnomAD and the Simons Genome Diversity Project include individuals from understudied and underrepresented groups. Similar to the analysis performed on the gnomAD data, we used ExpansionHunter to genotype 57 STRs in disease-associated genes in the Simons Genome Diversity Project. Of the 57 loci analyzed in this study, 39 (~68%) had significantly different average STR lengths among the populations analyzed. Because longer STRs mutate more readily, we also analyzed the relationship between the average length of these loci in each population and the prevalence of the associated disease. Huntington disease is more prevalent in non-Finnish European individuals, and the STR in *HTT* had the highest mean allele length in the non-Finnish European individuals. This study highlights the importance of including diverse populations to better understand the genetics of disease.

Evolutionary and Population Genetics Posters - Wednesday

PB2783. Evaluation of imputation performance of different reference panels in a Pakistani population.

Authors:

J. Xu^{1,2}, A. Arsalan³, G. Genovese⁴, D. Liu¹, B. Fennessy¹, E. Cheng¹, A. W. Charney¹, J. A. Knowles⁵, M. Ayub⁶, R. E. Peterson⁷, T. B. Bigdeli⁸, L. M. Huckins^{1,2}; ¹Pamela Sklar Div. of Psychiatric Genomics, Icahn Sch. of Med. at Mount Sinai, New York, NY, ²Dept. of Psychiatry, Yale Sch. of Med., New Haven, CT, ³Univ. of Peshawar, Peshawar, Pakistan, ⁴Stanley Ctr. for Psychiatric Res., Broad Inst., Cambridge, MA, ⁵The Human Genetics Inst. of New Jersey, Rutgers Univ., Piscataway, NJ, ⁶Univ. Coll. London, London, United Kingdom, ⁷The Virginia Inst. for Psychiatric and Behavioral Genetics, Virginia Commonwealth Univ., Richmond, VA, ⁸SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY

Abstract Body:

Despite calls for increasing diversity, GWAS studies are almost exclusively performed in European individuals, impeding the discovery of biologically important loci. Powerful imputation reference panels allow us to include rare variants in GWAS; however, performance of these panels is uneven across ancestries. Here, we compare for the first time the performance of three imputation panels in a large Pakistani population. We genotyped 2,219 Pakistani individuals using the Illumina Infinium™ Global Screening Array (v3.0). Genotypes were called following the MoChA pipeline using Google Cloud. Imputation was carried out using three reference panels: TOPMed, 1000 Genomes [1000G], and Genome Asia Pilot [GAsP]. First or second-degree relatives were removed. Among genotyped SNPs, the average mask R^2 (correlation between true genotyped values and imputed dosages) for common SNPs ($MAF > 0.05$) was 0.87, 0.87, and 0.86, respectively, for TOPMed, 1000G and GAsP; the average R^2 for rare SNPs ($MAF < 0.01$) was 0.57, 0.56, and 0.53. Among imputed SNPs, the average R^2 for common SNPs was 0.88, 0.84, and 0.84, whereas average R^2 for rare SNPs was 0.06, 0.18, and 0.20. TOPMed imputed 7.7-fold more rare SNPs than 1000G, and 20.6-fold more than GAsP. If we restrict to only SNPs shared across all 3 panels, the average R^2 for common SNPs was 0.89 for TOPMed, 0.87 for 1000G, and 0.85 for GAsP, and the average R^2 for rare SNPs was 0.37, 0.28, and 0.23. In terms of SNP counts, TOPMed, 1000G and GAsP imputed 5,281,437, 5,266,802, and 3,822,999 common SNPs with $R^2 \geq 0.8$. For rare SNPs, these 3 panels imputed 4,349,982, 1,415,749 and 646,301 SNPs with $R^2 \geq 0.8$. In addition, 648,080 multi-allelic SNPs and 812,111 indels with $R^2 \geq 0.8$ were imputed using TOPMed, compared to 189,677 multi-allelic SNPs and 820,967 indels using 1000G. No multi-allelic or indels was imputed using GAsP. We found 4 major categories of imputed SNPs: downstream, intergenic, intron and upstream. Among these, TOPMed has more SNPs with $R^2 \geq 0.8$ in each category. Our findings suggest that TOPMed with its much greater sample size ($n=97,256$) was able to impute more high-quality SNPs, especially the rare ones, with the caveat that it also imputes more low-quality rare SNPs that need to be filtered out prior to GWAS. For next step, we will examine R^2 between imputed SNPs and SNPs from targeted sequencing in the same set of individuals for evaluation of imputation accuracy. Overall, this work will contribute to more inclusive genetic studies of global populations.

Evolutionary and Population Genetics Posters - Wednesday

PB2784. Evidence of potential natural selection in African American individuals post admixture.

Authors:

J. Jaworski¹, J. Bartlett², R. Durodoye², H. Diaz-Zabala¹, S. Williams³, M. Aldrich⁴; ¹VUMC, Nashville, TN, ²Case Western Reserve Univ., Cleveland, OH, ³Case Western Reserve Univ, Cleveland, OH, ⁴Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Introduction. Admixed populations are formed by the mixing of genetically differentiated ancestral populations. Alleles that are highly differentiated, after an admixture event, can be present in frequencies locally that deviate from the global genome-wide admixture average. This can be due to natural selection post admixture. The extent of such recent selection in African American individuals is an unresolved question with prior work giving conflicting evidence with respect to directional selection since admixture in African American individuals. We looked for evidence of non-random patterns of local ancestry, perhaps due to natural selection, in a subset of the Southern Community Cohort Study. **Methods.** The sample consisted of 1,375 unrelated African American participants that were previously genotyped on a set of Illumina arrays (1M to 2.5). After merging and applying standard quality control metrics to the dataset, a total of 604,224 markers were available for analysis. The software, Efficient Local Ancestry Inference (ELAI) was used to measure African (AFR) and European (EUR) ancestry on an overlapping set of ~583K markers. We performed pairwise comparisons to search for regions for evidence of genome-wide selection. Tests were done using multiple reference AFR and EUR populations from the 1000 Genome reference dataset. We subsequently ran the same comparisons using the program RFMix (v1.5.4) to verify the results. **Results.** The average genome-wide AFR ancestry among was 84.4% (standard deviation = 0.835). A region on chromosome 2 (74.37 Mb to 75.10 Mb) showed an excess EUR ancestry over 5 standard deviations above the genome-wide mean ancestry and was observed in all reference comparisons. In addition, regions on chromosome 11 (57.41 Mb to 57.97 Mb) and chromosome 14 (45.26 Mb to 46.74 Mb) had excesses EUR ancestry in most of the AFR/EUR comparisons. Findings were replicated in the RFmix analysis. None of these regions have been previously reported. Within the chromosome 2 region, a nearby gene of interest is *HK2*, which catalyzes the phosphorylation of hexose and is associated with Type 2 Diabetes Mellitus, Thyroid Dysmorphogenesis 1, and Chondroblastoma. **Conclusion.** Our analysis shows three potential chromosomal regions of selection since admixture in an African American population. Future analyses will estimate the selection coefficient to explain the observed excess in EUR ancestry.

Evolutionary and Population Genetics Posters - Thursday

PB2785. Examining polygenic adaptation in time-stratified genome samples

Authors:

X. Cheng, M. Steinruecken; Univ. of Chicago, Chicago, IL

Abstract Body:

With the rapid accumulation of ancient DNA (aDNA) genomes and evolve-and-resequence (E&R) data, more time-stratified population genomic datasets are emerging. Such time-series data allow us to examine the temporal dynamics of natural selection and can lend power to detecting its footprints. Few selection-detecting methods, however, are tailored to jointly consider multiple samples of the same population at different times, especially when more than one locus is involved. Meanwhile, increasing evidence is supporting the polygenic nature of most traits under selection, underscoring the need for approaches that account for multiple loci. Here, we constructed a hidden-Markov model (HMM) framework based on the Wright-Fisher diffusion model explicitly for directional or stabilizing selection on polygenic traits. To reduce the computational load, we use a normal approximation for the step-wise transition between generations in the underlying diffusion model. We implemented this framework as a Python package with its command-line-interacting script, and benchmarked its performance with both forward and backward simulations. Further, for each major complex trait in the UKBioBank, we extract variation records of their significant loci in historical British populations during the past ~4500 years from the Allen aDNA Resources dataset, and assign these loci their respective effect size estimates in UKBioBank. With the composite likelihood incorporating independent loci across the genome, we were able to obtain estimates of the predominant mode of selection, i.e., directional or stabilizing, and its intensity, for each trait, gaining insights into recent human adaptation.

Evolutionary and Population Genetics Posters - Wednesday

PB2786. Exploring population health disparities through identity by descent genetic communities in The Biobank at The Colorado Center for Personalized Medicine

Authors:

J. Shortt¹, M. Lin¹, N. Rafaels², C. Caggiano³, R. Shemirani⁴, G. Belbin⁵, N. Zaitlen⁶, E. Kenny⁷, K. Crooks⁸, C. Gignoux¹; ¹Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ²Univ. of Colorado Sch. of Med., AURORA, CO, ³UCLA, Marina del Rey, CA, ⁴Icahn Sch. of Med. at Mount Sinai, Palo Alto, CA, ⁵Gencove Inc., New York, NY, ⁶UCLA, Los Angeles, CA, ⁷Icahn Sch. of Med. at Mt Sinai, New York, NY, ⁸Univ Colorado, Aurora, CO

Abstract Body:

Human demographic history and selection have shaped the genetic variation found within human populations for tens of thousands of years, and thus play a profound role in the risk of disease during an individual's life. However, understanding an individual's risk of disease is complicated by interacting environmental exposures that vary by population. Here, we use a data-driven approach to derive population substructure information and use it to characterize the risk of complex disease within high-resolution genetic communities. We identified identity by descent (IBD) segments indicative of recent shared ancestry among nearly 75,000 unrelated individuals genotyped in The Biobank at the Colorado Center for Personalized Medicine (CCPM). Individuals were clustered into 40 different genetic communities using pairwise cumulative IBD as input to iterative Louvain clustering. Communities are enriched in different regions of Colorado and have both overlapping and distinct global ancestry profiles indicative of population substructures within CCPM. We estimated the prevalence of over 1,200 phecodes derived from linked electronic health record data within each genetic community and identified community-specific enrichment or depletion across the phenome using logistic regression models. At FDR=5%, we identify 65 associations between 14 unique communities and 48 unique phecodes reflecting health disparities between communities, including two European ancestry communities with contrasting risk for skin cancer (OR[95% CI]=0.73[0.62-0.86] and 1.6[1.37-1.88]). Genetics is used to identify communities, but differences in risk between communities may largely reflect sociocultural, socioeconomic, and other environmental differences between communities. Communities may therefore be useful in accounting for these factors.

Evolutionary and Population Genetics Posters - Thursday

PB2787. Extensive sex-biased gene flow in Native Americans across the Andes

Authors:

V. Borda¹, L. Fehren-Schmitz², C. Sanchez³, O. Caceres³, C. Padilla³, E. Tarazona-Santos⁴, T. O'Connor¹; ¹Univ. of Maryland Sch. of Med., Baltimore, MD, ²Univ. of California Santa Cruz, Santa Cruz, CA, ³Centro Natl. de Salud - Inst. Natl. de Salud, Lima, Peru, ⁴Univ. Federal de Minas Gerais, Belo Horizonte, Brazil

Abstract Body:

The Central Andes region, including coast, highlands and adjacent Amazon rainforest, was the main scenario for the development of complex societies during the last 5000 years. Archaeological, linguistic and genetic evidence support the idea of two interaction spheres emerging in the North and the South regions of the Central Andes. Recent studies highlighted gene flow events in-between some North Coast/Highland and North Amazon populations. Here we explored how extensive was the gene flow across the Andes into the adjacent Amazon populations and how different were the sex-specific contributions. We used autosomal and x-chromosome genotyping data from previously published Native American samples resulting in 565 individuals. A subset of this data includes 11 populations from the adjacent Amazon forest. By performing genetic clustering analysis on both, autosomal and x-chromosome, we detected two main clusters explaining the genetic composition of the Amazon groups, Amazon related and an Andean related clusters. By comparing autosomal and x-chromosome ancestry proportions, we detect less Amazon related ancestries in x-chromosome with respect to the autosome in all 12 Amazon groups. To explore the nature of this observation, we performed admixture modeling using qpAdm to identify plausible models and admixture proportions that explain the genetic composition of each Amazon group. Nine out of 11 Amazon groups fit as the result of two-way admixture involving North Highlands (Ancient sample from La Galgada 4100BP) and Amazon related (Present-day Karitiana) sources. Interestingly, the ancestry proportions inferred were contrasting between autosomal and x-chromosome data. Furthermore, we computed Z scores to estimate how different were the ancestry proportions between autosomes and the x-chromosome. Six Amazon groups resulted in Z score values higher than 2.5 (2.5-5.19) suggesting a predominant male-based contribution from the Amazon related source into these populations living beside the Andes. We are providing substantial evidence for sex-biased gene flow involving the cultural interactions occurred across the Andes.

Evolutionary and Population Genetics Posters - Wednesday

PB2788. Fine-scale ancestry mapping revealed population-specific associations with complex disease

Authors:

M. Isshiki^{1,2}, S. Raj¹; ¹Albert Einstein Coll. of Med., New York, NY, ²The Univ. of Tokyo, Tokyo, Japan

Abstract Body:

In this project we aimed to identify the relationship between fine-scale genetic ancestry and prevalence of complex diseases in populations of African descent. Individuals of African ancestry in the United States often have African ancestry from multiple ancestral population groups, which reflect their complex history. We hypothesized that these different ancestries can be identified in the genome, and that each of these ancestries show different correlations to complex disease outcomes. We chose to focus on chronic diseases as nearly half of the US population has at least one chronic disease, and two-thirds of deaths globally are caused by the presence of one or more of the following chronic diseases: heart disease, cancer, stroke, chronic obstructive pulmonary disease, and diabetes. The prevalence and manifestation of these diseases varies among individuals and populations with different ancestral backgrounds. We used the methods PCA and ADMIXTURE to assess population structure of individuals of African descent from the TOPMED and All of Us datasets against a reference panel enriched for African ancestral populations. We then used fineSTRUCTURE to assess fine-scale population structure of these individuals. The methods RFMix, ELAI, and Gnomix were used to infer local ancestry. The accuracy of our local ancestry inference was validated by comparing the global mean values with the results of PCA and ADMIXTURE. We correlated local ancestry proportions with different complex diseases prevalence. We then used admixture mapping to perform association tests with the most correlated traits, adjusting for local ancestry. We found population-specific associations that underscore the importance of understanding fine-scale population structure in under-represented populations with complex ancestries, such as African-Americans. Our results suggest that local ancestry inference can be a powerful tool in understanding the contribution of ancestry to disease risk. Our findings support previous studies showing that African ancestry in the United States is not a monolith. We take this one step further and find associations with common disease phenotypes.

Evolutionary and Population Genetics Posters - Thursday

PB2789. Generation of synthetic genotypes with neural networks.

Authors:

M. Geleta^{1,2}, D. Mas Montserrat³, M. Perera², X. Giro-i-Nieto², A. G. Ioannidis^{3,4}; ¹Computer Sci. Dept., Univ. of California, Berkeley, Berkeley, CA, ²Univ.t Politècnica de Catalunya, Barcelona, Spain, ³Dept. of BioMed. Data Sci., Stanford Univ., Stanford, CA, ⁴Inst. for Computational and Mathematical Engineering, Stanford Univ., Palo Alto, CA

Abstract Body:

Although the number of sequenced genomes has grown substantially, there remains a clear imbalance between the ancestries that are represented with the vast majority of sequenced samples still coming from populations of European descent. This results in an underrepresentation of training data for non-European populations. Here we investigate genomic data synthesis tools to overcome the problem of imbalanced ancestry groups. We present some of the novel methodologies for the generation of realistic synthetic genotypes using different connectionist approaches: Variational Autoencoders (VAEs) and Generative Moment Matching Networks (GMMN); and introduce an evaluation scheme to qualitatively and quantitatively assess the quality of the synthetic single nucleotide polymorphism (SNP) sequences. Additionally, we show how such synthetic genotypes can be used to address privacy and data sharing concerns.

Evolutionary and Population Genetics Posters - Wednesday

PB2790*. Genetic and phenotypic high-altitude adaptation at *EPAS1/HIF2A* in Andean highlanders

Authors:

W. Gu¹, E. S. Lawrence¹, R. J. Bohlender², C. Anza-Ramirez³, A. M. Cole⁴, J. J. Yu¹, H. Hu², E. C. Heinrich⁵, K. A. O'Brien^{1,6}, C. A. Vasquez⁷, Q. T. Cowan⁷, P. T. Bruck⁷, K. Mercader⁸, M. Alotaibi⁸, T. Long⁸, J. E. Hall¹, E. A. Moya¹, M. A. Bauk¹, J. J. Reeves¹, M. C. Kong⁹, R. M. Salem¹⁰, G. Vizcardo-Galindo¹¹, J-L. Macarlupu¹¹, R. Figueroa-Mujica¹¹, D. Bermudez³, N. Corante³, E. Gaio¹², K. P. Fox⁸, V. Salomaa¹³, A. S. Havulinna¹³, A. J. Murray⁶, A. Malhotra¹, F. L. Powell¹, M. Jain⁸, A. C. Komor⁷, G. L. Cavalleri⁴, C. D. Huff², F. C. Villafuerte³, T. S. Simonson¹; ¹Dept. of Med., Div. of Pulmonary, Critical Care, and Sleep Med., Univ. of California San Diego, La Jolla, CA, ²Dept. of Epidemiology, Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, ³Laboratorio de Fisiología Comparada/Fisiología del Transporte de Oxígeno-LID, Departamento de Ciencias Biológicas y Fisiológicas, Univ. Peruana Cayetano Heredia, Lima, Peru, ⁴Sch. of Pharmacy and Biomolecular Sci., Royal Coll. of Surgeons in Ireland, Dublin, Ireland, ⁵Div. of BioMed. Sci., Sch. of Med., Univ. of California, Riverside, Riverside, CA, ⁶Dept. of Physiology, Dev. and NeuroSci., Univ. of Cambridge, Cambridge, United Kingdom, ⁷Dept. of Chemistry and Biochemistry, Univ. of California, San Diego, La Jolla, CA, ⁸Dept. of Med. and Pharmacology, Univ. of California, San Diego, La Jolla, CA, ⁹Dept. of Bioengineering, Univ. of California, San Diego, San Diego, La Jolla, CA, ¹⁰Dept. of Anthropology and Global Hlth., Univ. of California, San Diego, La Jolla, CA, ¹¹Laboratorio de Fisiología Comparada/Fisiología del Transporte de Oxígeno-LID, Departamento de Ciencias Biológicas y Fisiológicas, Univ. Peruana Cayetano Heredia, Lima, Perú, Lima, Peru, ¹²Laboratório de Fisiologia Respiratória, Faculdade de Med., Univ. e de Brasília, Brasília, Brazil, ¹³Dept. of Publ. Hlth. and Welfare, Finnish Inst. for Hlth. and Welfare, Helsinki, Finland

Abstract Body:

High-altitude populations have been exposed to environmental stress for thousands of years and provide a unique opportunity to study human evolutionary processes and genetic variants selected for hypoxia tolerance. Prior studies have identified Hypoxia-Inducible Factor (HIF) pathway genes as crucial components of adaptation in human and non-human high-altitude species. However, links between precise variants within adaptive gene regions and physiological traits remain largely unexplained. To elucidate the genetic basis of high-altitude adaptation in Andean highlanders, we conducted a composite of multiple signals selection scan on 40 high-coverage whole genomes. We intersected a list of prioritized candidate genes with strong evidence for positive selection with a list of *a priori* genes and looked for deleterious, non-synonymous mutations within the selected regions. We highlighted a missense variant (rs570553380, A>G, p.[His194Arg]) within the sixth exon of *Endothelial PAS Domain Protein 1 (EPAS1)*. The wildtype histidine at this residue is highly conserved among vertebrates except for the coelacanth, *Latimeria chalumnae*, a lobe-finned fish. In humans, this derived variant is largely unique to Andean populations (Andean: 10.0% vs. 0.20% global allele frequencies) and is likely under early-stage selection (estimated age = 9,845 ± 1969 years, which approximately aligns with the first human habitation of the Andes). As hemoglobin concentration has been correlated with the *EPAS1* locus in the Tibetans, we tested whether rs570553380 was associated with hematocrit in an expanded cohort of Andean highlanders (N = 224) and found males with more copies of the putatively adaptive G allele exhibited lower hematocrit (P < 0.003). We then identified serum-metabolite correlations between rs570553380 and 1 known and 11 unknown metabolites, which are associated with hematocrit in Andeans and also a large genome cohort (FINRISK). Furthermore, we introduced rs570553380 into HEK293T cells using CRISPR ABE base-editing and found that cells with one copy of the adaptive variant demonstrated decreased expression of canonical HIF targets relative to wildtype cells under hypoxia. Associations between *EPAS1* and hematocrit and hemoglobin concentration across Andean and Tibetan suggest

ASHG 2022 Annual Meeting Poster Abstracts

phenotypic convergence, likely via distinct molecular and physiological changes. These findings provide the first example of convergent phenotypes associated with genetic adaptation across high-altitude human populations.

Evolutionary and Population Genetics Posters - Thursday

PB2791. Genetic landscape and demographic history of Britain and Ireland

Authors:

A. Shanmugam^{1,2}, A. M. Molloy³, L. Brody⁴, M. Merrigan⁵, S. O'Reilly⁵, G. L. Cavalleri^{1,2,6}, R. McLaughlin^{3,2,6}, R. P. Byrne³, **E. Gilbert**^{1,6}; ¹Royal Coll. of Surgeons in Ireland, Dublin, Ireland, ²The SFI Ctr. for Res. Training in Genomics Data Sci., Galway, Ireland, ³Trinity Coll. Dublin, Dublin, Ireland, ⁴NHGRI (NIH), Bethesda, MD, ⁵Genealogical Society of Ireland, Dún Laoghaire, Ireland, ⁶The SFI FutureNeuro Res. Ctr., Dublin, Ireland

Abstract Body:

Background: Recent studies have identified subtle but discrete genetic groups across Britain and Ireland. Leveraging patterns of identity-by-descent (IBD) segment sharing, we built upon our understanding of fine-scale population structure in Britain and Ireland and derived insights into recent demographic change in these populations.

Methods: We assembled genotype data from 6,724 individuals across four studies with associated regional ancestry. The Leiden community detection algorithm was applied recursively to identify genetic clusters in the dataset. Additionally, we inferred regional effective population sizes across time using IBDNe, and estimated recent migration rates using MAPS to reveal insights into recent demographic history.

Results: We identified fine-scale genetic population structure which stratified by geography and detected new subgroups in Ireland (for ex. Donegal-Cavan) and South-East England (for ex. Norfolk-Suffolk). Using IBDNe, we provide novel insights into changes in regional effective population sizes over time, with all population clusters showing patterns of recent exponential growth. We have also identified evidence of gene flow barriers through estimated migration surfaces.

Conclusions: Our findings extend our knowledge of fine-scale population structure across Britain and Ireland, adding novel findings with previously undetected genetic clusters. The analysis of changes in recent effective population size and migration surfaces deepens our understanding of the history of these population genetic groups.

Evolutionary and Population Genetics Posters - Wednesday

PB2792. Genetic variation related to the regulation and expression of hemoglobin isoforms among global populations

Authors:

J. Hall¹, M. C. Kong¹, W. Gu¹, J. J. Yu¹, C. Anza-Ramirez², K. Fox¹, F. Villafuerte², M. Gotesman³, T. S. Simonson¹; ¹Univ. of California, San Diego, La Jolla, CA, ²Univ. Peruana Cayetano Heredia, San Martín de Porres, Peru, ³The Lundquist Inst., Torrance, CA

Abstract Body:

Despite limited oxygen availability in each breath of air, human populations living at high altitude within the Andes, Tibetan plateau, and Ethiopian highlands exhibit both unique physiological responses to this environmental stress and extremely strong natural selection for genes involved in oxygen transport. In previous studies, we identified several gene regions under natural selection within in an Andean cohort from Cerro de Pasco, Peru, which contained fetal hemoglobin qualitative trait loci (QTLs) that influence variance in fetal hemoglobin levels in patients with sickle cell disease (SCD). Fetal hemoglobin has been suggested to play a role in oxygen transport in high-altitude populations and is upregulated under conditions of stress erythropoiesis and in patients with SCD. We hypothesize that genes involved in fetal hemoglobin persistence and symptom severity in SCD patients also play a role in adaptation to adverse physiological responses to hypoxic stress at high altitudes. Hemoglobin expression relies on the interplay of a multitude of transcription factors and can be influenced by alterations in transcription factor binding sites for the hemoglobin genes or upstream transcription factors. To identify population-specific variations in transcription factor sites that may be involved with variance in hemoglobin regulation and expression, we compiled a comprehensive list of all known single nucleotide variants (SNVs) within genomic regions identified as under selection in our Andean cohort coding for the hemoglobin isoforms and upstream transcription factors. These genomic regions encompass the hemoglobin gene cluster, hemoglobin locus control region, and the genes *BCL11A* and *SOX6*. Variants were filtered to include SNVs with allele frequencies greater than 5% within African (AFR) and East Asian (EAS) populations and the Andean cohort. SNP2TFBS software was used generate SNV lists specific to each population to identify shared and unique transcription factor site SNVs. From this, we identified hundreds of transcription factor binding site variants shared between two or more of the populations, along with 38 unique SNVs in Africans, 18 in Andeans, and 3 in East Asian populations that we hypothesize may play a role in population-specific variation in hemoglobin expression within these populations.

Evolutionary and Population Genetics Posters - Thursday

PB2793. Genome-wide patterns of tandem repeat abundance across species are primarily driven by retrotransposons

Authors:

Y. Jin, M. Maksimov, **M. Gymrek**; Univ. of California San Diego, La Jolla, CA

Abstract Body:

Short Tandem Repeats (STRs) are sequential repeat elements in DNA sequences that are abundant in most eukaryotic species. STRs are known to contribute to variation in gene function and gene expression across species. It has been observed that the abundance and characteristics of STRs vary widely across genomes of different species. However, the evolutionary origins of this variation and its implications remain largely unexplored. Previous work suggested that species- or clade-specific retrotransposon elements may represent a major source of STRs in eukaryotic genomes. Here, we explored the composition of STRs across the tree of life with species ranging from yeast to primates and investigated the critical role retrotransposons play in driving species-specific differences in genomic STR content. Our results confirm previous reports that homopolymers are strongly enriched in primate genomes, and that this trend is driven by SINE retrotransposon elements. Further, we identify DNAREP1_DM and TART_DV elements as drivers of dinucleotide abundance in multiple fruit fly species and CR1-Y1_Ave and CR1-Y2_Aves as drivers of pentanucleotide abundance in birds. Finally, we find that dinucleotide repeats are enriched in abundance and tend to be longer in rodent species, but that this enrichment is likely not driven by rodent-specific retrotransposon elements and is therefore likely driven by alternative mechanisms. Overall, our work highlights the major role retrotransposon elements play in shaping the composition of STRs across the genomes of diverse species.

Evolutionary and Population Genetics Posters - Wednesday

PB2794*. Genomic deserts: A global survey of underrepresented populations in genomics research

Authors:

S. Mangul¹, W. Wolfsberger^{2,3}, K. Chhugani¹, Y-N. Huang¹, D. Yu⁴, S. Knyazev⁵, S. Nataneli¹, J. Hu⁶, V. Munteanu⁷, S. Groppa⁸, T. Oleksyk^{2,3}; ¹Dept. of Clinical Pharmacy, Sch. of Pharmacy, Univ. of Southern California, Los Angeles, CA, ²Dept. of Biological Sci., Oakland Univ., Rochester, MI, ³Uzhhorod Natl. Univ., Uzhhorod, Ukraine, ⁴Dept. of Quantitative and Computational Biology, Dornsife Coll., Univ. of Southern California, Los Angeles, CA, ⁵Dept. of Pathology and Lab. Med., David Geffen Sch. of Med., Univ. of California Los Angeles, Los Angeles, CA, ⁶Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ⁷The Faculty of Computers, Informatics and Microelectronics, Technical Univ. of Moldova, Chisinau, Moldova, Republic of, ⁸Dept. of Neurology, Nicolae Testemitanu State Univ. of Med. and Pharmacy, Chisinau, Moldova, Republic of

Abstract Body:

Genomic sequencing provides clues to understand the physiological and pathological mechanisms of humans. Studying population genomics over diverse populations can promote genomics research across all the populations around the world at an unprecedented scale and increase the health equity across diverse populations. Despite the development and contribution of national genome sequencing projects, such projects currently represent only a small fraction of the world populations, primarily in high income and developed countries, and thus contribute to the growing global inequality in genomic research. Typically, high income and developed countries have more resources to develop national genome projects compared to the low income or developing countries, resulting in many countries and regions in the world becoming blind spots lacking their own genomic representation. An effective way to fill the existing blind spots in the global genomic map is to initiate national level genome projects to enable genomic-based medicine for the population of this country. To address the emerging issues of global genomic diversity and to further emphasize the urgent need to engage diverse populations in genomic research, we have completed a systematic survey of national genomic projects across 195 countries and regions of the world in order to show inconsistencies representing worldwide populations in genomics research and data sharing practices. Our results show locations of numerous “genome deserts” in the world across all parts of the world (Europe, Africa, Asia, Australia, Oceania and the Americas). The regions containing the most deserts were South America and the regions with the highest density of genomic projects were Asia. We also illustrated how many genomic projects failed to represent minority populations limiting the capacities of personal medicine for these groups. We observed a large variability of availability of genomic data ranging from sharing the genomic data freely (14.39%) and upon request (85.61%). Among the genomic data, 72.99 % (47,984 samples) of the genomic data is generated from uni-national genome projects and 27.01% (11,761 samples) of the genomic data is generated from international collaborative genome projects. We discussed the advantages and limitations of various sharing models and their role in empowering local scientific communities. We also investigated the data sharing practices adopted by national genomic projects and concluded with discussion with scientists from countries lacking the national genomic projects on recommendations, opportunity and pitfalls on establishing and implementing national genomic projects.

Evolutionary and Population Genetics Posters - Thursday

PB2795. Genomic variation and chromatin structure across the human-specific *NOTCH2NL* duplications

Authors:

T. Real¹, M. Vollger¹, D. Dubocanin¹, P. Dishuck¹, A. Sedeno Cortes¹, X. Guitart¹, K. Munson¹, J. Ranchalis¹, E. Eichler^{1,2}, A. Stergachis¹; ¹Univ. of Washington, Seattle, WA, ²Howard Hughes Med. Inst., Seattle, WA

Abstract Body:

Segmental duplications (SDs) are nearly identical stretches of DNA that exist in multiple copies within the genome. SDs are among the fastest-evolving genomic regions and can underlie human diseases and novel genes with human-specific functions. For example, the *NOTCH2NL* loci, which expanded to 5 gene paralogs selectively in humans, underlies four neurodevelopmental diseases and encodes a novel gene associated with expansion of the human frontal cortex. However, the sequence and function of SDs, including *NOTCH2NL*, has remained largely opaque, as standard sequencing technologies cannot overcome their high sequence identity. Advances in long-read genomic and epigenomic sequencing are beginning to open previously impenetrable portions of the genome for studying. Here, I leverage these emerging technologies to study genetic diversity and gene regulatory function at *NOTCH2NL* loci. Specifically, using 1 human reference and 4 non-human primate (NHP) accurate long-read assemblies of the *NOTCH2NL* loci, I have reconstructed the evolutionary history of *NOTCH2NL* within the Great Ape lineage, demonstrating that *NOTCH2NL* has likely independently evolved in humans. Additionally, using 94 parental haplotype-resolved assemblies from the Human Pangenome Reference Consortium (HPRC) project, I have characterized the scope of genetic variation at these loci across a diverse human population. This demonstrated pervasive large-scale structural and genetic changes affecting all *NOTCH2NL* paralogs. Finally, using single-molecule chromatin fiber sequencing of human brain tissue I have defined the chromatin landscape of each *NOTCH2NL* paralog, showing paralog-specific promoter chromatin architectures. Overall, this study begins to delineate the genetic and epigenetic architecture and evolution of *NOTCH2NL* paralogs in humans, and provides a framework for studying these unexplored segmentally duplicated regions of the genome that likely play an essential role in what makes us human.

Evolutionary and Population Genetics Posters - Wednesday

PB2796*. Germline genetic determinants of T-cell fraction in ~110,000 diverse individuals

Authors:

H. Poisner¹, A. Faucon¹, N. Cox², A. Bick³; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ Med Ctr., Nashville, TN, ³Vanderbilt, Nashville, TN

Abstract Body:

T-cell abundance is subject to competing selection in humans. Germline genetic determinants of human T-cell abundance have not been studied at scale, because complete blood counts only assay lymphocyte abundance which is the sum of T-cells, B-cells, and NK-cells. To address this gap, we extended the T-cell ExTRECT method for use with whole-genome sequence data. It estimates T-cell fractions from read depth along the V(D)J recombination region of the T-Cell Receptor Alpha locus. We leveraged the NHLBI BioData Catalyst cloud platform to estimate blood T-cell fractions in the multi-ethnic TOPMed study at scale (N~110k, ~40K non-European) and performed a genome-wide association study. The SNP heritability of T-cell fraction was 0.10 (SD 0.02) in European ancestry and 0.42 (SD 0.04) in African ancestry individuals (SumHer BLD-LDAK model). We identified 12 genome-wide significant loci that affect one or more hematopoietic lineages. Three affect both lymphoid and myeloid lineages (*ACKR1*, *SH2B3*, *IRF8*). Three are lymphocyte-specific (*CD69*, *KLF2*, *IL7*). Three are T-cell-specific (*BCL2L11*, *TRDV2*, *SHH*). Interestingly three are myeloid lineage-specific (*HBB*, *CSF3*, *CSF3R*), implying that changing the relative proportion of myeloid cells changes the fraction of whole blood DNA derived from T-cells. We developed a polygenic model with a Julia port of PRS-cs for estimating T-cell fractions in TOPMed Europeans. Projecting this model into Vanderbilt BioVU (N~72k, EUR) revealed that neutrophil and lymphocyte measurements were inversely correlated with predicted T-cell fractions. This inverse relationship was also replicated in the UK Biobank. We used fine-mapping and manual curation to prioritize causal variants in T-cells. At three of our loci we prioritized four variants found in enhancer regions for *BCL2L11* (T-cell apoptosis), *IL7* (T-cell development), and *CD69* (T-cell proliferation). All three have differential allele frequencies in European and African ancestry populations. Phenome-wide analyses identified associations of these loci with immune-related diseases including Multiple Sclerosis (*IL7*) and hypothyroidism (*CD69*). In summary, by exploiting a new method to estimate T-cell abundance directly from whole blood sequence data, we obtained new insights into the genetic regulation of T-cells in diverse populations, with implications for autoimmune disease susceptibility.

Evolutionary and Population Genetics Posters - Thursday

PB2797. Guaranteeing unbiasedness in selection tests based on polygenic scores.

Authors:

J. Blanc, J. Berg; Univ. of Chicago, Chicago, IL

Abstract Body:

Population stratification is a well-studied problem in genome-wide association studies, leading to biases in the estimated strength of phenotypic association for individual genetic variants. In short, if environmental effects on the phenotype are correlated with ancestry gradients within a GWAS panel, any variant that is stratified along this ancestry gradient will receive a biased effect size estimate. While state of the art methods to correct for stratification are generally effective in reducing the number of significant false positive associations, even subtle biases in effect size estimates can accumulate across loci, leading to systematic biases in polygenic scores. In turn, these biases in the distribution of polygenic scores can lead to false positives in downstream analyses, such as tests for polygenic adaptation or other analyses of among group genetic differences. One approach is to attempt to be overly aggressive in controlling for stratification. However, there is currently no way to tell conclusively if confounding effects have been removed. A second approach is to conduct the GWAS in an evolutionarily diverged sample that is less likely to share population genetic structure with the test panel. This renders potential biases in the effect sizes irrelevant to the test, but comes at the cost of significantly reduced statistical power due to the issue of poor portability of polygenic scores across samples of divergent ancestry. Here using theory from population and statistical genetics, together with simulations, we show how this second approach can be generalized to panels that do share genetic structure, and that it is possible to guarantee the unbiasedness of selection tests without needing to guarantee that the effect sizes are fully unbiased. Specifically, if the researchers performing the GWAS also have access to the panel of test individuals and have identified the specific test to be performed ahead of time, then it is possible to compute a covariate to include in the GWAS, which will guarantee that the test is unbiased. More generally, our results have implications beyond tests for selection as any analysis that attempts to quantify the correlation between polygenic scores and demographic or environmental variables is subject to the same type of stratification biases.

Evolutionary and Population Genetics Posters - Wednesday

PB2798. Haplotype-based analyses of phylogeny and regional genome diversity among laboratory rats using the latest rat reference genome

Authors:

Y. Pan¹, A. Ozel², A. Palmer³, H. Akil¹, H. Chen⁴, J. Li⁵; ¹University of Michigan, Ann Arbor, MI, ²Univ. of Michigan, Ann Arbor, MI, ³UCSD, LA JOLLA, CA, ⁴Univ. of Tennessee, Knoxville, TN, ⁵Univ Michigan, Ann Arbor, MI

Abstract Body:

The common rat (*Rattus norvegicus*) is a key species in biomedical research. We previously reported Illumina whole-genome sequencing of eight inbred rat lines that are founders of the Heterogeneous Stock. Variant calls based on the previous rat reference genome, rn6, revealed regions with unusually high levels of heterozygosity, which span >9% of the genome and harbor >1,700 genes. These regions likely reflect mis-assemblies in rn6. We and colleagues in the International Rat Omics Consortium (IROC) have collaborated with the Vertebrate Genome Project to complete a new reference genome, termed mRatBN7. Re-analyses of the same data with mRatBN7 led to a significant reduction of high-heterozygosity regions (<0.5% of the genome). The two references differ in several large regions (>10 MB), including reversed orientation and placement to a different chromosome. mRatBN7 showed a higher human-rat concordance than rn6, including a 20 MB region on ChrX. Compared to rn6, mRatBN7 led to fewer zero-coverage regions (0.8% vs. 4.5% for males, and 0.3% vs. 2.4% for females), higher rate of mapping (97.3% vs 96.3%), and fewer variants being called (21,497,403 sites vs 23,893,814 sites). IROC investigators have performed linked-read WGS and joint variant calling for 168 samples representing most of the commonly used laboratory rat strains, including inbred lines, several RI panels, and technical replicates. Using the resulting variants data we constructed global phylogenetic trees at the sample, sub-strain and strain levels to investigate their evolutionary relationships, and compared with prior data for a subset of the samples using genotyping platforms. We generated region-by-region phylogenetic trees and defined shared ancestral haplotypes among the strains. In doing so, we developed a clustering-based method to infer local haplotypes and their likely boundaries. Shared haplotypes arise from co-ancestry relationships among various strains and, as a result, an eight-ancestry panel may have fewer than 8 unique haplotype in a given region, sometimes as few as 1. The catalog of the size, distribution, and composition of these ancestral haplotypes along the entire genome provides insights into the evolution of rat strains and informs regional variation of genetic diversity that supports genetic mapping studies.

Evolutionary and Population Genetics Posters - Thursday

PB2799*. High coverage whole genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios

Authors:

M. Byrska-Bishop¹, U. S. Evani¹, X. Zhao², A. O. Basile¹, H. J. Abel³, A. A. Regier³, A. Corvelo¹, W. E. Clarke¹, R. Musunuri¹, K. Nagulapalli¹, S. Fairley⁴, A. Runnels¹, L. Winterkorn¹, E. Lowy⁴, Human Genome Structural Variation Consortium, P. Flicek⁴, S. Germer¹, H. Brand², I. M. Hall⁵, M. E. Talkowski², G. Narzisi¹, M. C. Zody¹; ¹New York Genome Ctr., New York, NY, ²Massachusetts Gen. Hosp., Boston, MA, ³Washington Univ., St. Louis, MO, ⁴European Molecular Biology Lab., Cambridge, United Kingdom, ⁵Yale Univ., New Haven, CT

Abstract Body:

The 1000 Genomes Project (1kGP) is the largest fully open resource of whole genome sequencing (WGS) data consented for public distribution of raw sequence data without access or use restrictions. The final, phase 3 release of the 1kGP included 2,504 unrelated samples from 26 populations and was based primarily on low coverage WGS. Here, we present a new, high coverage WGS 1kGP resource that has been extended to 3,202 samples including 602 complete trios and sequenced to a targeted depth of 30X using the Illumina NovaSeq 6000 system. We performed single nucleotide variant (SNV) and short insertion and deletion (INDEL) calling against the GRCh38 reference using GATK HaplotypeCaller which resulted in discovery of over 111 million SNVs and over 14 million INDELS with false discovery rates of 0.3% and 1.15%, respectively, across the entire cohort of 3,202 samples. We also discovered and genotyped over 173 thousand structural variants (SVs), including insertions, deletions, duplications, inversions, and multiallelic copy number variants, by integrating multiple algorithms and analytic pipelines, including GATK-SV, svtools, and Absinthe. When compared to the low coverage phase 3 1kGP dataset, the variant counts in the high coverage callset reflect an estimated average increase of 190,885 SNVs (1.05-fold), 268,182 INDELS (1.47-fold), and 5,835 (2.81-fold) SVs per genome, and a cohort-level increase of over 18.6 million SNVs (1.24-fold), 9.8 million INDELS (4.05-fold), and ~100 thousand SVs (2.47-fold), across the original 2,504 unrelated samples. Although phase 3 did achieve the goal of capturing almost all common SNVs (MAF > 1%), the higher coverage plus advances in sequencing and analytic methods greatly expanded the discovery of all rare variants and of INDELS and SVs across the frequency spectrum. As part of this work, we also released an improved reference imputation panel based on the high coverage WGS consisting of SNV, INDEL, and SV calls across the 3,202 1kGP samples, including full trios. Our phased panel leveraging pedigree correction provides improvements in power across the board, but particularly in the imputation of many more common INDELS and SVs, making these accessible through imputation for association studies. Given the large effect size of non-SNV variation, this increase in power is likely to enable discovery of new genetic associations that help pinpoint the causative variant. We make all the data generated as part of this project publicly available and we envision this updated version of the 1kGP callset to become the new de facto public resource for the worldwide scientific community working on genomics and genetics.

Evolutionary and Population Genetics Posters - Wednesday

PB2800. Human populations and demographics in Qatar from the Neolithic to the Late Iron Age.

Authors:

A. D'Aurelio¹, M. Baldoni^{2,3}, F. R. Vempalli¹, F. De Angelis², F. Sakal⁴, F. Al-Naimi⁴, M. Al-Hashmi¹, K. Wang¹, L. Wang¹, G. Wang¹, O. Soloviov¹, F. Castorina⁵, C. Martínez-Labarga², S. Tomei¹; ¹Integrated Genomics Services, Res. Dept., Sidra Med., Doha, Qatar, ²Dept. of Biology, Univ. of Rome Tor Vergata, Rome, Italy, ³PhD Program in Evolutionary Biology and Ecology, Dept. of Biology, Univ. of Rome Tor Vergata, Rome, Italy, ⁴Dept. of Archaeology, Qatar Museums, Doha, Qatar, ⁵Dept. of Earth Sci., Univ. of Rome Sapienza, Rome, Qatar

Abstract Body:

Motivation: Despite the production of genetic data related to the Middle East and Qatar specifically is increasing dramatically, a critical unmet research and public need relies on the study of the prehistorical societies living in this region of the World. Recent archaeological studies have discovered thousands of prehistoric burials that represent a great opportunity to investigate the population dynamics and sociocultural changes in prehistorical Qatari societies. The archaeological record has been the only way by which the Qatari prehistoric populations have been researched. Here we have applied the study of ancient DNA (aDNA) to understand the origin, migration patterns, genetic relationships, admixtures, kinship, and changes in prehistorical Qatari societies.

Material and methods: We performed aDNA analyses on 20 samples selected by taking into account preservation status and availability of either tooth or petrous bone. The same samples were also submitted for radiocarbon dating and isotopic and morphological analysis.

DNA was extracted in a dedicated clean lab facility starting from 50-135 mg of bone powder, following a silica-based protocol tailored to aDNA, modified by adding a further digestion step. Illumina double-stranded libraries were prepared and treated with partial uracil-DNA-glycosylase (UDG) to prevent nucleotide misincorporation.

Final NGS libraries were sequenced through NovaSeq 6000 System on an S4 Flow Cell.

Using a combination of bioinformatics tools, we performed quality control analysis on the sequenced reads, selecting endogenous reads with map and base PRHEd score quality above 30.

Results: According to the radiocarbon dating, our samples cover a transect of time going from the Neolithic to the Late Pre-Islamic period (7.4-1.4 kya). We report quality control scores on the sequencing data and their correlation with sample type, burial location and burial time frame. Even though the DNA preservation was mined by detrimental environmental factors, we tried estimating ancestries from both multi-locus genotype data and model-based approaches.

Conclusions: To the best of our knowledge, our study represents the first attempt to analyze aDNA in the Arabian Peninsula. We successfully retrieved aDNA sequences from human samples older than 1500 years excavated in Qatar. We believe that our results have the potential to pave the way for further paleogenetic studies in the region.

This work was supported by a grant from the Qatar National Research Fund (NPRP10-0208-170411). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the Qatar National Research Fund.

Evolutionary and Population Genetics Posters - Thursday

PB2801. Improved detection of evolutionary selection highlights potential bias from different sequencing strategies in complex genomic-regions

Authors:

T. Hayeck^{1,2}, Y. Li¹, T. Mosbrugger¹, J. Bradfield³, A. Gleason¹, G. Damianos¹, G-W. Shaw¹, J. Duke¹, L. Conlin⁴, T. Turner⁵, M. Fernandez Vina⁶, M. Sarmady⁷, D. Monos¹; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ³Quantinuum Res. LLC, San Diego, CA, ⁴Children s Hosp. of Philadelphia, Philadelphia, PA, ⁵Washington Univ. Sch. of Med., St. Louis, MO, ⁶Stanford Univ., Palo Alto, CA, ⁷Children S Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Balancing selection occurs when multiple alleles are kept at elevated frequencies in equilibrium due to opposing evolutionary pressures. A new statistical method, LD-ABF, was developed to test for selection using efficient Bayesian techniques, accounting for both the density of polymorphisms and strength of linkage disequilibrium. It demonstrated the most robust detection of selection in a variety of simulation scenarios. Established targets as well as 45 novel genes were identified to be under balancing selection after a genome wide scan of hundreds of clinical trios. Selection peaks were observed across multiple gene families whose biological functions favor diversification through both allele polymorphism and gene duplications: olfactory-receptor, immunoglobulin, zinc-finger, etc. Surprisingly, *SIRPA*, demonstrated dramatic selection signal second only to the MHC in most clinical populations and was replicated in separate whole-genome long-read samples from the Pangenome. Despite being known to likely be under strong selection, the MHC is a notoriously complex region that is difficult to properly sequence and characterize. To improve the detection of selection signal in the MHC, we took advantage of high-resolution HLA typing from thousands of diverse individuals and whole-genome long-read sequencing data. This analysis revealed strong balancing selection in expected peptide-binding domains as well as previously understudied intronic and intergenic regions of the HLA genes while demonstrating consistent patterns in diverse populations when taking advantages of higher quality sequencing technologies. The improved statistical method combined with higher quality sequencing results in more consistent findings of selection and improved localization of variants under selection, especially in complicated regions.

Evolutionary and Population Genetics Posters - Wednesday

PB2802. Imputation around the world: Assessing imputation quality across diverse global populations.

Authors:

J. Cahoon^{1,2}, X. Rui³, E. Tang², C. Simons², J. Langie³, M. Chen³, Y-C. Lo³, C. W. K. Chiang^{3,2}; ¹Dept. of Computer Sci., Univ. of Southern California, Los Angeles, CA, ²Dept. of Quantitative and Computational Biology, Univ. of Southern California, Los Angeles, CA, ³Ctr. for Genetic Epidemiology, Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA

Abstract Body:

Genotype imputation is now an essential component in human genetic studies. By predicting unobserved genotypes based on sequenced individuals, imputation increases marker density and enables large-scale genome-wide association studies. The state-of-the-art imputation reference panel released by the Trans-Omics for Precision Medicine (TOPMed) contains a substantial number of admixed African and Hispanic/Latino samples. As a result, these populations are imputed with nearly the same efficacy as European cohorts. However, imputation for ethnic minorities primarily residing outside of North America still falls short in performance due to persisting underrepresentation. To illustrate this point, we curated genome-wide array data from 28 publications published between 2008 to 2021. In total, we imputed over 30k individuals across 145 populations around the world. We identified a number of populations where the imputation accuracy paled in comparison to that of European populations. For instance, the mean imputation r^2 for variants with minor allele frequency (MAF) between 1.0-5.0% for Saudi Arabian (N=1061), Vietnamese (N=1264), Thai (N=2435), and Papua New Guineans (N=776) were 0.79, 0.78, 0.76, and 0.62, respectively. In contrast, the mean r^2 ranged from 0.90 to 0.93 for comparable European populations matched in sample size and SNP content. In addition to overall lower imputation quality, minority populations experienced a steeper drop in imputation accuracy for rarer variants. For example, the mean r^2 for Filipinos (N=1779) declined from 0.93 in alleles with 5-10% MAF to 0.59 in alleles with 0.5-1.0% MAF. In contrast, alleles in the same frequency classes imputed relatively well in Europeans with a decrease from 0.97 to 0.79. Despite tremendous effort in generating large imputation panels like TOPMed, our results suggest that global populations still incur at least a 10% drop in study power due to imputation accuracy alone and even greater power loss for rarer variants. While strategies leveraging smaller population-specific reference panels in conjunction with meta imputation may increase imputation quality, ultimately, reference panels must strive to increase diversity to promote equity within genetics research.

Evolutionary and Population Genetics Posters - Thursday

PB2803. Imputation of ancient genomes.

Authors:

B. Mota^{1,2}, S. Rubinacci^{1,2}, D-I. Cruz-Dávalos^{1,2}, C. Guerra Amorim³, M. Sikora⁴, M. Allentoft^{4,5}, E. Willerslev^{4,6,7,8}, A-S. Malaspina^{1,2}, O. Delaneau^{1,2}; ¹Univ. of Lausanne, Lausanne, Switzerland, ²Swiss Inst. of Bioinformatics (SIB), Lausanne, Switzerland, ³California State Univ., Northridge, CA, ⁴Univ. of Copenhagen, Copenhagen, Denmark, ⁵Curtin Univ., Bentley, Australia, ⁶Univ. of Cambridge, Cambridge, United Kingdom, ⁷Wellcome Sanger Inst., Cambridge, United Kingdom, ⁸MARUM, Univ. of Bremen, Bremen, Germany

Abstract Body:

Ancient DNA (aDNA) studies have unraveled significant aspects of our past that help explain the genetic variation we observe today. aDNA is affected by extensive damage, including fragmentation and C-to-T substitutions. As a result, ancient genomes often have low sequencing depth, which is an impediment to confident genotype calling. Imputation has been proposed as a solution to this problem. However, it is unclear whether ancient genomes can be accurately imputed and how imperfect imputation affects downstream analyses. To address these questions, we downsampled 42 high-coverage (>10x) ancient human genomes from different times and continents to low coverage and subsequently imputed them with GLIMPSE, an imputation and phasing tool, using 1000 Genomes as a reference panel. We found that ancient and modern DNA imputation accuracies were comparable. For most 1x genomes, we recovered original high-coverage genotypes with low error rates (<5% for genotypes with at least one copy of the alternative allele). Error rates were higher for African genomes, likely due to underrepresentation in the reference panel. There were no measurable differences between imputation accuracy of transversion and transition sites, although the latter are known to be affected by C-to-T substitutions. To study imputation effects on downstream analyses, we analyzed high-coverage and imputed genomes using principal component analysis (PCA), genetic clustering estimating ancestry contributions for the European individuals, and runs of homozygosity (ROH, an inbreeding measure typically requiring diploid data). For these applications, we obtained similar results with high-coverage and imputed genomes when depth of coverage was at least 0.5x. Finally, we showed that aDNA imputation accuracy can be further improved by making use of methodological improvements, i.e., GLIMPSE2, and larger reference panels, i.e., UK Biobank with 150K whole-genome sequences. For example, in the case of four Viking genomes, error rates at alternative allele sites went down to below 5% at 0.25x, which was previously only achieved at 0.75x. Altogether, our results show that, with an appropriate reference panel and depending on the genome coverage, imputation can be reliably applied to expand and improve aDNA studies.

Evolutionary and Population Genetics Posters - Thursday

PB2804*. Inference of natural selection on epigenetic marks: implications for the evolution of gene regulation and germline mutation rates

Authors:

L. Boukas, A. Razi, H. Bjornsson, K. Hansen; Johns Hopkins Univ., Baltimore, MD

Abstract Body:

Natural selection is one of the fundamental forces shaping the evolution of living organisms, and detecting its presence is a major goal of evolutionary biology. Thus far, focus has been on detecting selection on the DNA sequence. However, natural selection acts on phenotypes; footprints of selection within the DNA sequence only reflect the extent to which the latter causally determines these phenotypes. In turn, the causal effect of the DNA sequence on phenotypes is mediated via intermediate functional molecular features, such as epigenetic marks. Therefore, it is of interest to move beyond DNA, and identify selective pressures on specific molecular features. This represents the main objective of “evolutionary cell biology”.

Here, to the best of our knowledge, we develop the first framework for detecting selection on epigenetic marks. We formalize a notion of neutrality, and use it to derive a test for selection. Our test captures the key aspect of the underlying biology: that epigenetic marks are controlled in trans by transcription factors and chromatin modifiers. We illustrate the generality of our approach by applying it to 3 different modifications (DNA methylation, H3K4me3, H3K36me3) and a broad range of genomic compartments (promoters, gene bodies, transcriptional end regions). Our findings reveal the action of selection on several marks, including an unexpected finding of positive selection on the size of the hypomethylated region around proximal promoters, even though this has a negligible correlation with expression. In addition, we find that the selective pressure on proximal promoter DNA methylation/H3K4me3 and gene-body H3K36me3 is not a passive consequence of selection on gene expression. This provides evolutionary support for an active, causal role of these marks on gene regulation.

Finally, we provide evidence (using comparative analyses, analyses of trio genomic sequencing data, and population genetic simulations) that - in addition to their involvement in gene regulation - the selective pressure on DNA methylation and H3K36me3 may partly be explained by their effect on regional mutation rates in the germline (on promoter CpGs and exons, respectively). There is currently little empirical support for selection on regional mutation rate modifiers, and it has been proposed that such secondary selection is dominated by random drift. However, epigenetic marks affect the mutation rate at multiple locations simultaneously; we show this cumulative effect may overcome drift.

Our framework for selection inference is simple but general, and we anticipate its core idea to be useful for other molecular features beyond epigenetic marks.

Evolutionary and Population Genetics Posters - Wednesday

PB2805. Inference of the distribution of fitness effects using local genealogical trees

Authors:

D. Ortega-Del Vecchyo, A. Izarraras-Gomez; Natl. Autonomous Univ. Of Mexico, Queretaro, Mexico

Abstract Body:

The distribution of fitness effects of new mutations (DFE) measures the proportion of new variants that have a particular selection coefficient. The DFE is a fundamental determinant of levels of genetic and phenotypic variation. Previous methods that estimate the DFE in humans have mainly used information from summary statistics, such as the site frequency spectrum or metrics of linkage disequilibrium. All of the summary statistics are ultimately a function of the collection of local genealogical trees across the genome. Due to this, the collection of genealogical trees should contain all the necessary information to perform an inference of the DFE that could be more accurate than previous approaches that use summary statistics. Here we propose a new maximum likelihood method to infer the DFE using information from genome-wide local trees. This method uses information from the number of lineages seen at different times in the past in a collection of local genealogical trees. We show that our method is accurate using simulations. Furthermore, we show an application of our method in data from the 1000 genomes project to infer the DFE in nonsynonymous sites in the genome and sets of variants predicted to be deleterious based on various algorithms such as CADD and EVE. We show how our method can be applied to infer the distribution of fitness effects of all the variants found in the genome by leveraging functional information that include data from CADD and EVE.

Evolutionary and Population Genetics Posters - Thursday

PB2806. Inferences about archaic introgression and positive selection from Lithuanian whole-genome sequences

Authors:

A. Urnikyte, A. Masiulyte, A. Molyte, V. Kucinskas; Dept. of Human and Med. Genetics, Inst. of BioMed. Sci., Faculty of Med., Vilnius Univ., Vilnius, Lithuania

Abstract Body:

The sequencing of archaic and modern human genomes has demonstrated gene flow between modern and archaic populations. To understand the functional and evolutionary consequences of introgression we need to identify the genetic variants inherited from archaic hominin ancestors in local populations. Here we present results of identified introgressed archaic DNA fragments in the genome of present-day Lithuanians and determined whether positive selection has acted on these fragments.

To identify candidate introgressive regions, we analysed whole-genome data (~36.27x coverage) of 50 Lithuanian individuals by using two statistical approaches: Sprime and ArchIE. Signatures of positive selection were investigated using three statistics: Tajima's D, FST, and XP-EHH. Annotation performed with Ensembl Variant Effect Predictor and ANNOVAR in GRCh37 (hg19), RefSeqGene, gnomAD, dbSNP147, and CADD version 1.3.

We found 80 Mb of the Lithuanian genome covered by putative introgressed segments detected by ArchIE, and 0.64 Mb by Sprime. The biggest number of Neanderthal introgressed fragments found with ArchIE was identified in chromosome 6, which has HLA genes, important in the acquired immunity. Also, ArchIE identified a fragment in chromosome 9, that has the BNC2 gene, which has an impact on the level of skin pigmentation in Europeans, and a fragment in chromosome 11, comprising the OR4X1 gene, which encodes olfactory receptor.

We have identified strong genomic regions, already described previously, introgressed from Neanderthals, and under positive selection. The most interesting regions include HLA-DRB1 and BNC2 genes.

Comparative analysis of detected introgressed archaic fragments in the Lithuanian population with Vindija Neanderthal genome identified matching SNPs, which are residing in genes, related to adaptive immunity, mitochondrial translation, and DNA reparation.

This research was funded by the European Social Fund under the No 09.3.3-LMT-K-712 "Development of Competencies of Scientists, other Researchers and Students through Practical research Activities" measure.

Evolutionary and Population Genetics Posters - Wednesday

PB2807. Inferring between-population genetic variation using admixed populations: challenges and limitations.

Authors:

J. Huang¹, M. D. Shriver¹, A. A. Zaidi²; ¹Penn State Univ., University Park, PA, ²Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

The extent to which phenotypic differences between populations are driven by a difference in mean genetic value is determined using common garden experiments where the environment can be controlled. In humans, where this is neither practical nor ethical, we often analyze admixed populations, where ancestry and environment are decoupled. Here, we describe, using theory and simulations, the quantitative behavior of genetic variance underlying complex traits in admixed populations and our ability to leverage admixture to infer how source populations differ in their genetic values.

First, we show that ancestry-phenotype correlations, which are often used to illustrate that genetic effects underlie phenotypic differences between populations, are mediated both by trait architecture (e.g. heritability and F_{st}) and admixture dynamics (e.g. time(s) since admixture, ancestry-stratified mating, and degree of gene flow). For traits differing in mean genetic value between the source populations, the ancestry-phenotype correlation is highest when the populations first meet and decays over time. The rate of decay is slowed in the presence of ancestry-stratified mating and continued gene flow from source populations.

We show that while ancestry-phenotype correlations can be informative about genetic effects, at least within a few generations after admixture, they can also result from ancestry-related environmental exposures. Thus, ancestry-phenotype correlations alone are not indicative of a mean genetic difference between the source populations.

It has been suggested that in principle, the genetic variance due to local ancestry in admixed populations can be informative about mean differences in genetic value between source populations. However, the local-ancestry genetic variance is mathematically equivalent to the genetic variance between populations but only at the time of admixture. This equivalence decays in subsequent generations at a rate that depends on admixture dynamics and assortative mating.

We provide a quantitative treatment of the genetic variance components of complex traits in admixed populations as a function of their demographic history and the distribution of environmental effects. In doing so, we highlight the challenges and limitations of using admixed populations to make inferences about the difference in mean genetic value between the source populations. Our work also provides a quantitative description of the effects of admixture dynamics on the predictive accuracy of polygenic scores in admixed populations.

Evolutionary and Population Genetics Posters - Thursday

PB2808*. Inferring germline CpG methylation signature accumulated along the human history from genetic variation catalogs

Authors:

Y. Si, S. Zöllner; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

DNA methylation is a key to understanding gene regulatory mechanisms and modeling mutation rate heterogeneity. Genome wide methylation levels are measured by whole genome bisulfite sequencing (WGBS), but it is currently restricted to small sample sizes especially for human germline cells. Moreover, these sequencing measures depend on individual genotypes of the samples, where C to T mutation carriers do not have methylation signature at that site but the ancestral alleles could be methylated, resulting in a systematic bias masking highly methylated thus often mutated sites. Here we present an alternative approach to estimate historical methylation patterns. We estimate mutation rates of germline methylated CpG sites to be 2.8 times that of non-methylated CpG and 11.4 times that of non-CpG sites, consistent with experimental studies. As a large number of whole genomes have been sequenced, at least one mutation event has been recorded in nearly all methylated thus highly mutable CpG sites. Those methylated CpG sites often contain multiple identical alleles that have originated independently in history, distorting the local sample frequency spectrum from the expectation under the infinite site assumption. CpG methylation status is highly localized, a feature distinct from other major factors affecting mutation rates. Here we leverage this locality and introduce a Hidden Markov Model to infer the germline mutation burden and historical germline methylation patterns based on allele frequencies from large genetic variant catalogs. We applied our method to publicly available allele frequencies from the gnomAD and the TOPMed databases. Our estimates are consistent with methylation levels observed in human germ cells by WGBS in 93.7% of measured CpG sites, but we also identified ~442,000 methylated CpG sites that were not captured due to sample genetic variation and inferred methylation status for ~721,000 CpG sites that were missing from WGBS. We observed that due to technical limitations, missingness of WGBS is enriched by 4.8 fold in GC-rich regions such as CpG islands, where our method has the highest accuracy and methylation status are of the most interest to identify active chromatin regions. Thus combining our results and WGBS identifies active non-methylated regions with higher specificity, and characterizes the germline methylation landscape more comprehensively overall than using experimental measures alone. Our estimated historical methylation status can be leveraged to enhance all bioinformatic analysis of methylation status such as predicting mutation constrain, and annotating regulatory and inactivated genomic regions.

Evolutionary and Population Genetics Posters - Wednesday

PB2809. Insights into non-equilibrium population genetics under strong natural selection from a new theoretical formalism

Authors:

D. Balick; Harvard Med. Sch., Boston, MA

Abstract Body:

Natural populations are virtually never observed in equilibrium: no estimate of human demographic history thus far has resulted in a static effective population size long enough to reach equilibrium. Yet, equilibrium assumptions are both ubiquitous in statistical tools and comprise much of our understanding of population genetics. Here, I develop a field theoretic formalism for population genetics, using standard tools from stochastic and quantum field theory, that describes the temporal evolution of the allele frequency probability distribution in non-equilibrium settings. I focus on strong purifying selection, where most human disease variation lies, to understand the response to a range of demographic scenarios observed in human history. The dynamics of equilibration to mutation-selection-drift balance in a population with constant size are considered, along with three demographic scenarios: exponential growth (e.g., recent human history), a population experiencing a bottleneck (e.g., the Out of Africa), and oscillatory population size (e.g., seasonal changes in *Drosophila*). This formalism provides a prescription for attaining closed-form approximations for arbitrary moments of the frequency distribution in any demography, or for time-dependent selection or mutation, without time-intensive simulations. The first four moments are presented: dynamics of the mean frequency, variance (and homozygosity), skew, and excess kurtosis, the latter two generally ignored in analysis of the frequency distribution. These moments equilibrate at slightly different rates, but generically take longer than the naïve estimate of order $1/s$. A transient increase in the variance of the frequency distribution is detailed prior to equilibration, which can be intensified by exponential growth, indicating that equilibrium does not maximize the possible variance of frequency distribution. Surprisingly, in a constant population, the skew and excess kurtosis approach fixed values, independent of selection (provided $2Ns > 5$) and dependent only on $\theta = 4N\mu$. Under exponential growth, this can be temporarily attained in a transient quasi-equilibrium, prior to the rapid decay of the drift coefficient; thus, even nearly lethal disease can be temporarily harbored in the high frequency tail of the distribution due to this transient state. This field theory of population genetics provides highly accurate approximations in a range of human-relevant regimes under strong selection, enabling a deeper understanding of non-equilibrium scenarios ubiquitous in human history.

Evolutionary and Population Genetics Posters - Thursday

PB2810. Integrative approaches linking divergent selection to disease association studies

Authors:

G-W. Shaw¹, T. L. Mosbrugger¹, Y. Li², D. Monos³, T. Hayeck³; ¹Div. of Genomic Diagnostics, Dept. of Pathology and Lab. Med., Children's Hosp. of Philadelphia, Philadelphia, PA, ²Div. of Genomic Diagnostics, Dept. of Pathology and Lab. Med., Children's Hosp. of Philadelphia; Dept. of Pathology and Lab. Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ³Div. of Genomic Diagnostics, Dept. of Pathology and Lab. Med., Children's Hosp. of Philadelphia, Philadelphia; Dept. of Pathology and Lab. Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Differences and similarities in genetic variation across world populations help us understand adaptations to evolutionary pressures in response to disease. This study leverages the latest release of 1000 Genomes Project (1KGP), high coverage whole genome sequencing with hundreds of complete trios for improved phasing, to characterize genetic diversity and evolutionary selection across and within world populations. A new test statistic was developed looking at balancing and positive selection, to detect regional differentiation in patterns of linkage disequilibrium and density of polymorphism across populations. Strong evidence of population specific differentiation was mapped throughout the 1KGP both at the super and sub-populations levels. We linked selection signal in 1KGP to summary statistics from large-scale genome-wide association studies (GWAS), matching ancestral group in both. This ties potential divergent evolution to corresponding genes and diseases in specific populations. Divergent selection in and around the BTNL2 gene was seen in East Asians at variants tied to both liver and lung function. Whereas in Africans strong divergent selection signal is observed in and around ALG10B at variants related to cardiomyopathy, along with signal in HLA-C and SPOCK3 related to immune response. Interestingly, consistent patterns enrichment of selection signal across populations were observed in both unprocessed pseudogenes and polymorphic pseudogenes pointing to potential biological functions favoring diversification and development of novel genes. Our results provide insight into both divergent and common patterns of evolutionary selection across world populations with implications in adaptation to disease response

Evolutionary and Population Genetics Posters - Wednesday

PB2811. Introgressed sequences and positive selection profiles of the Japanese population revealed by the analysis of 3,256 Japanese whole-genome sequencing (WGS) data.

Authors:

K. Tomizuka¹, X. LIU¹, S. Koyama¹, Y. Ishikawa¹, M. Horikoshi², S. Ikegawa³, Y. Kamatani^{1,4}, K. Ito⁵, K. Matsuda⁶, Y. Momozawa¹, C. TERAOKA^{1,7,8}; ¹RIKEN, Yokohama, Japan, ²RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, ³IMS, RIKEN, Tokyo, Japan, ⁴The Univ. of Tokyo, Tokyo, Japan, ⁵RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, ⁶Osaka Univ. Graduate Sch. of Med., Tokyo, Japan, ⁷Shizuoka Gen. Hosp., Shizuoka, Japan, ⁸Univ. of Shizuoka, Shizuoka, Japan

Abstract Body:

Genetic studies conducted in the Japanese population contributed enormously to our understanding of the genetic basis of various diseases and complex traits. However, the lack of nationwide Japanese whole-genome sequencing (WGS) resources hinders further genetic and translational studies. To fill this gap, we generated a high-depth WGS dataset of 3,256 individuals from different regions of Japan, termed as Japanese Encyclopedia of Whole/Exome sequencing library (JEWEL). We identified 15.4 million unreported variants including rare and functional variants. This valuable resource allowed us to examine several unexplored characteristics of the Japanese population. First, we applied a recently developed reference-free method to detect sequences likely introgressed from Neanderthal or Denisovan, and we identified 44 archaic segments that were significantly associated with clinically relevant phenotypes in the Japanese. Notably, majority of them were specific to East Asian populations. Furthermore, we identified 22 genetic loci under positive selection at the genome-wide significance level including known targets such as *ADH*, *ALDH2*, *MHC* and a dozen of novel loci harboring immune related genes. Collectively, our work revealed new insights into the genetic characteristics of the Japanese population.

Evolutionary and Population Genetics Posters - Thursday

PB2812. Investigating human migrations and genetic diversity in Northeast India over the last ~4,000 years using ancient genomics

Authors:

E. Bandyopadhyay¹, C. de la Fuente¹, D. Witonsky¹, T. Jamir², D. Tetso³, N. Rai⁴, M. Raghavan¹; ¹The Univ. of Chicago, Chicago, IL, ²Dept. of History and Archaeology, Nagaland Univ., Kohima Campus, Nagaland, India, ³Dept. of Anthropology, Kohima Sci. Coll. (Autonomous), Nagaland, India, ⁴Birbal Sahni Inst. of PalaeoSci.s, Lucknow, India

Abstract Body:

Northeast India occupies a strategic position between South Asia, East Asia, and Southeast Asia. Archaeological evidence from cave and rockshelter sites along the Nagaland-Myanmar border suggests that hunter-gatherer societies survived in the forested uplands of Northeast India spanning the mid-Holocene to as late as ~3,500 years before present (yBP). The temporal context to the transition to agriculture in the state of Nagaland is yet to be firmly established. More broadly, across Northeast India, archaeological records reveal that prominent features of farming culture, or the Neolithic, postulated to date back as early as ~4,000 yBP, such as edge ground tools, cord-mark pottery and rice/millet cultivation, display similarities with East Asia and Southeast Asia. Additionally, genetic affinities have been reported between present-day Northeast Indian populations and Eastern Eurasians, though the temporal context and source of this affinity remain unconfirmed by genetic data. This raises an important question about the genetic relationship between hunter-gatherers, early farmers, and present-day inhabitants of the region as well as regional affinities over time with Eastern Eurasian populations. In this study, we aim to characterize the timing and sources of past human movements and interactions in Northeast India by analyzing genome-wide data from samples excavated from the rockshelter site of Photangkhun Longkhap (PTL) in Nagaland State in Northeast India, and archaeologically dated to the hunter-gatherer phase *ca.* 3,700 - 4,300 yBP. Results from this study, representing the first ancient genomic data from this region of the world, will shed light on past genetic diversity and the human movements into Northeast India over the last ~4,000 years.

Evolutionary and Population Genetics Posters - Wednesday

PB2813. Investigating selection and the genetics of height in Central African Rainforest Hunter-gatherers

Authors:

D. Ju¹, M. McQuillan¹, M. Hansen¹, D. Harris¹, C. Zhang¹, A. Ranciaro¹, W. Beggs¹, S. Chanock², C. Fokunang³, A. Njamnshi⁴, M. Yeager⁵, I. Mathieson¹, S. Tishkoff¹; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Natl. Cancer Inst, Rockville, MD, ³Univ. of Bamenda, Bamenda, Cameroon, ⁴Univ. of Yaounde I, Yaounde, Cameroon, ⁵FNLCR/NCI, Rockville, MD

Abstract Body:

Shorter stature has been hypothesized to have been under selection in Central African Rainforest Hunter Gatherer (RHG) populations, but the genetic loci involved remain unclear. We studied a cohort of RHG groups encompassing Baka and Bagyeli (N=207) and an agricultural population, Tikari (N=246), from Western Cameroon that were genotyped on the H3Africa array. We first identified genomic signatures of positive selection in the RHG and then quantified the height variation within RHG that can be explained by global ancestry and European GWAS height-associated SNPs. Finally, we inferred local ancestry tracts in the RHG, comparing two methods, Rfmix and MOSAIC. We used the selscan tool in the Ohana package to identify loci exhibiting the most allelic differentiation in the Baka compared to Tikari, San, and 1000 Genomes CEU. A locus on chromosome 1 that encompasses a cluster of cytochrome P450 genes, involved in drug metabolism and lipid synthesis, showed the strongest signal of selection ($P=2.28E-10$), which was recapitulated as the top hit in a separate population branch statistic (PBS) test. A previously reported shared selective sweep across RHG groups upstream of *TRPS1*, which encodes a transcription factor regulating genes involved in bone and cartilage development, also exhibited an elevated signal in our Ohana ($P=3.34E-7$) and PBS (99.97 percentile) scans. Our integrated haplotype score scan identified a selective sweep in the Baka (normalized $iHS = -7.06$) that harbors a stop gain variant in the last exon of *TLR5*, suggesting adaptation related to immunity. We replicated a previous observation of RHG ancestry being associated with shorter height ($P=0.002$), but it only explained 5.5% of the variance in height in a regression of height on global ancestry, as inferred by ADMIXTURE ($K=2$), controlling for sex and age. We then constructed polygenic scores using SNPs from the UK Biobank GWAS conducted in European-ancestry individuals and found polygenic scores have even less explanatory power within RHG ($R^2=0.025$; $P=0.024$) and do not reflect differences in phenotypic height between RHG and Tikari. Finally, we called local ancestry tracts using Rfmix and MOSAIC and observed MOSAIC systematically called more RHG ancestry in the Baka than Rfmix. Rfmix tended to infer less overall RHG ancestry compared to ADMIXTURE, but concordance was high ($r^2=0.90$). This observation indicates that traditional approaches for inferring local ancestry may be inaccurate in populations with complex demographic histories, especially with old admixture events. Funding: NIH grants R35GM134957-01,5T32DK007314-39, 1F31HG011813-01A1; ADA Grant 1-19-VSN-02

Evolutionary and Population Genetics Posters - Thursday

PB2814. Kidd Lab 55 AISNP panel shows a better performance than the SNP*for*ID 34-plex for genetic ancestry estimates in a Brazilian population sample.

Authors:

C. Mendes-Junior¹, V. M. S. Moraes¹, A. L. E. Pereira², L. Marcorin², G. Debortoli², T. M. T. Carratto¹, G. Valle-Silva¹, N. C. A. Fracasso², A. B. C. Silva¹, E. A. Donadi², A. L. Simoes², E. C. Castelli³, M. G. Oliveira²; ¹Univ.e de São Paulo (FFCLRP-USP), Ribeirao Preto, Brazil, ²Univ.e de São Paulo (FMRP-USP), Ribeirao Preto, Brazil, ³Univ.e Estadual Paulista (FMB/UNESP), Botucatu, Brazil

Abstract Body:

Ancestry informative markers are being used for biogeographic ancestry inference, at the individual or population levels, for medical, forensic and anthropologic purposes. Due to the use of specific parameters to select a given set of AIMs, coupled with the loss of accuracy introduced by the efforts to reduce the number of markers in a set, additional studies are necessary for assessment of their limitations. In this context the SNP*for*ID 34-plex was developed for a single-tube analysis aiming to differentiate Sub-Saharan African, European and East-Asian populations. The Kidd's Lab set of 55 AISNPs was set to differentiate the Sub-Saharan Africa, Europe, Southwest-Asia, South-Asia, East-Asia, Oceania and the Americas. However, their effectiveness for Latin American admixed populations is still poorly assessed. Moreover, despite the complex and countrywide heterogeneous process of peopling, the majority of studies consider the Brazilian population as tri-hybrid, composed by African, European, and Amerindian inputs. Thus, we aim to analyze these two panels of AIMs in an admixed Brazilian population sample from Ribeirão Preto (Southeastern Brazil), and to evaluate the outcomes of models that consider Ribeirão Preto either as tetra- or tri-hybrid (for this latter case, exploring the interchange between Amerindians and East Asians as the third component). DNA was extracted from 513 individuals by salting-out. These two sets of markers were included into a customized Next-Generation Sequencing assay using HaloPlex Target Enrichment System (Agilent) and the MiSeq Personal Sequencer platform (Illumina). PCA analysis for the 34-plex panel clustered Amerindians with East-Asians populations, while the 55 AIMs provided better separation between them, particularly when considering the third component. This issue has influenced populational and individual ancestry inference: three different methods (Admix 2.0, Admix 95 and STRUCTURE) used to assess the three models revealed that EUR (ranging from 67.4% to 70.2%) and AFR (ranging from 19.9% to 24.4%) contributions are consistent between all analyses. However, AMR ancestry is underestimated while EAS ancestry is overestimated in the sampled individuals when the 34-plex is taken into account. Overall, given the capacity of differentiating EAS from AMR, the Kidd's Lab set of 55 AISNPs is better suited for analyzing the ancestry composition at the individual and population levels when both AMR and EAS ancestries are concerned. FINANCIAL SUPPORT: FAPESP (Grant 2013/154470), CNPq/Brazil (Grant 448242/2014-1 and Fellowship 312802/2018-8), and CAPES/Brasil (Finance Code 001).

Evolutionary and Population Genetics Posters - Wednesday

PB2815. Large-scale analysis of 6,000 whole genomes from Qatar uncovers genetic structure of Arab and Middle Eastern populations and establishes a valuable resource for understanding personalized disease risk and causality

Authors:

H. Naeem¹, R. Razali¹, J. Rodriguez-Flores², M. Ghorbani¹, W. Aamer¹, E. Aliyev¹, N. Syed¹, F. Vempalli¹, H. Almabrazi¹, R. TEMANNI¹, L. Li¹, G. Wang¹, M. Hashmi¹, A. Khouly¹, S. Poolat¹, T. Zaid¹, S. Tomei¹, S. Lorenz¹, R. Al-Ali¹, Q. Consortium³, A. Clark⁴, K. Fakhro¹, Y. Mokrab¹; ¹Sidra Med., Doha, Qatar, ²Regeneron Genetics Ctr. LLC, TARRYTOWN, NY, ³List of consortium authors and their affiliations will be provided upon request, Doha, Qatar, ⁴Cornell Univ, Ithaca, NY

Abstract Body:

Arab populations are largely understudied, notably their genetic structure, history as well as underlying disease risk and architecture. We present an in-depth analysis on a novel set of 6,218 whole genomes sequenced at 30x coverage from the Qatar Genome Program (QGP). This revealed extensive diversity as well as genetic ancestries representing the main founding Arab genealogical lineages of Qahtanite (Peninsular Arabs) and Adnanite (General Arabs and West Eurasian Arabs). Peninsular Arabs were found to be the closest relatives of ancient hunter-gatherers and Neolithic farmers from the Levant, and founder Arab populations experienced multiple splitting events 12-20 kya, consistent with the aridification of Arabia and farming in the Levant, giving rise to settler and nomadic communities. As far as recent genetic flow is concerned, these ancestries were found to contribute significantly to European, South Asian as well as South American populations, likely as a result of Islamic expansion over the past 1400 years. In terms of runs of homozygosity (ROH) which are known to be a risk for recessive disease, we found unprecedented lengths ROH particularly amongst Peninsular Arabs, extending to 60-70 Mb. Furthermore, we characterize a cohort of 1,491 men with the ChrY J1a2b haplogroup, identifying 29 unique sub-haplogroups (based on 103 novel SNVs), representing the largest set of subjects sequenced to date with this haplogroup. Based on this QGP dataset of predominantly healthy adults, we built an imputation panel containing 12,432 haplotypes and 69,018,172 SNVs. This provides an unprecedented panel to expand genomic knowledge of Middle Eastern populations, empowering precision genomics studies in the region and worldwide.

Evolutionary and Population Genetics Posters - Thursday

PB2816. Late Night Eating or Circadian disruption could be more dangerous than our think Late Night Eating or Circadian disruption could be more dangerous than our think

Authors:

Q. Naz, N. Verma, K. Usman, A. Mahdi; KGMU, Lucknow, India

Abstract Body:

Aim & Objective: The aim of this study is to investigate whether there was a relationship between morningness(MC) (110n) intermediate(IC) (100n) & eveningness chronotype(EC) (35n) in T2DM. **Methods:** A total of 245 subjects' age 18 to 60 years were recruited in Clinical OPD of General Medicine, KGMU. We have tested FBG & PP level, lipid profile HbA1c, Insulin, Leptin and Cortisol level, 48 hours ABPM. **Result:** When we compared these 3 groups, Significant Different parameters found in FBG (P = 0.01) Postprandial (P = 0.03) HbA1c (P = 0.001) TG (P = 0.0001), Total Cholesterol (P = 0.01) & VLDL (P = 0.005). It also shows the complete inversion of the cortisol level (0.003). Insulin, IL-1 beta & IL-6 also show significant change in late night eating T2DM Patients. Systolic / Diastolic readings of ABPM shows significant change between MC and IC (0.005) but not b/w EC & IC (0.007). And for reliability of sleep by actigraphy shows MC (6:15 ± 1:35) & EC (8:18 ± 1:23) take complete sleep but IC total sleep hours (5:10 ± 1:05) are very less. Disruption of Rev Erb (0.003) & Ror α (0.001) gene expression is also a risk factor for Cardio metabolic Diseases in T2DM Patients. **Conclusion:** Eating late at night are more likely to have late night eaters & are associated with greater risk of Metabolic Disorders like Blood Pressure Variability, Dyslipidemia, T2DM & CVD.

Evolutionary and Population Genetics Posters - Wednesday

PB2817. Leveraging Base Pair Mammalian Constraint to Understand Genetic Variation and Human Disease

Authors:

S. Gazal¹, M. Christmas², M. Dong², D. Genereux³, I. Kaplow⁴, J. Meadows², P. Sullivan⁵, E. Karlsson⁶, K. Lindblad-Toh²; ¹Univ. of Southern California, Los Angeles, CA, ²Uppsala Univ., Uppsala, Sweden, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Carnegie Mellon Univ., Pittsburgh, PA, ⁵Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁶Univ. of Massachusetts Med. Sch., Worcester, MA

Abstract Body:

Although thousands of genomic regions have been associated with heritable human diseases, attempts to elucidate biological mechanisms are impeded by a general inability to discern which genomic positions are functionally important. Evolutionary constraint is a powerful predictor of genome function obtained from genome sequences alone. Previous studies of mammalian constraint were limited by species number and reliance on human-referenced alignments. Here, we used a reference-free whole-genome alignment of 240 species to resolve evolutionary constraint to single base resolution and leveraged this information to understand genetic variation and human diseases and complex traits.

First, we computed single base-level phyloP scores from the whole genome alignment of 240 placental mammals and identified 3.5% of the human genome as significantly constrained (5% FDR). Second, we illustrated that these scores annotate coding variants with a base pair resolution, thus providing a powerful tool to predict deleterious effects of coding variants. Third, we showed that more than 80% of constrained bases are non-coding and that nearly half are unannotated by resources like ENCODE, suggesting yet-to-be-discovered functional elements. Fourth, we performed SNP-heritability of constraint common variants on 63 independent European ancestry GWAS and observed that Zoonomia constraint scores are far more informative than other key functional annotations (i.e. GTEx, ENCODE) on human diseases and complex traits. Specifically, constrained variants were more enriched than key functional annotations, such as non-synonymous coding variants or fine-mapped eQTL-SNPs, validating the ability of constraint to dissect human diseases and complex traits. Finally, we illustrated how Zoonomia constraint scores increase power and interpretation of functionally informed fine-mapping results when using the PolyFun method on UK Biobank traits.

Our results demonstrate the utility of constraint as a functional annotation that can be leveraged to deepen our understanding of heritable human diseases, but also highlight that the regulatory landscape of the human genome still needs to be further explored and potentially used to help decipher mechanisms underlying disease.

Evolutionary and Population Genetics Posters - Thursday

PB2818. Long-range regulatory effects of Neandertal DNA in modern humans

Authors:

D. Yermakovich, V. Pankratov, U. Võsa, B. Yunusbayev, Estonian Biobank Research Team, M. Dannemann; Inst. of Genomics, Univ. of Tartu, Tartu, Estonia

Abstract Body:

The admixture between modern humans and Neandertals has resulted in ~2% of the genomes of present-day non-Africans being composed of Neandertal DNA. Association studies have shown that introgressed DNA significantly influences skin and hair traits, immunity and behavioral phenotypes in people today. Several of the phenotype-associated archaic variants have links to regulatory effects as well. In general, analyses of allele-specific expression, regulatory sequence composition and cis-eQTL have demonstrated a significant contribution of this introgressed DNA to the transcriptomic landscape of modern humans. However, little is known about the impact of Neandertal DNA on long-range regulatory effects that have been shown to explain ~20% of expression variation.

In this study, we investigated one particular mechanism where the trans effect is mediated through cis-eQTLs expression regulation of transcription factors (TFs) genes. We identified 60 TFs genes with their top cis-eQTL SNP being of Neandertal ancestry in GTEx and predicted long-range Neandertal DNA-induced regulatory effects by screening for the predicted target genes of those TFs. We found elevated levels of interactivity between these TFs suggesting a potential for dependency in the functional processes they are involved in. We show that genes in regions devoid of Neandertal DNA are enriched among the target genes of some of these TFs. Furthermore, archaic cis-eQTLs for these TFs included multiple candidates for local adaptation and have associations with various immune traits, schizophrenia, blood cell type composition, and anthropometric measures. Finally, we show that our results can be replicated in empirical trans-eQTLs with Neandertal variants.

Our results suggest that the regulatory reach of Neandertal DNA goes beyond the 40% of genomic sequence that it still covers in present-day non-Africans and that via this mechanism Neandertal DNA additionally influences the phenotypic variation in people today.

Evolutionary and Population Genetics Posters - Wednesday

PB2819. Lower allele frequency among males is a marker for pathogenicity among non-pseudoautosomal X-chromosome variants

Authors:

T. Ciesielski, J. Bartlett, S. Iyengar, S. Williams; Case Western Reserve Univ., Cleveland, OH

Abstract Body:

Pathogenic variants on the X-chromosome can have more severe consequences for hemizygous males, while heterozygote females can avoid severe consequences due to diploidy and the capacity for nonrandom expression. Thus, an allele being more common in females could indicate variant pathogenicity, and large-scale genomic data now makes it possible to compare allele proportions between the sexes. To discover pathogenic variants on the X-chromosome, we analyzed exome data from 125,748 ancestrally diverse participants in the Genome Aggregation Database (gnomAD). After filtering out duplicates and extremely rare variants, 44,606 of the original 348,221 remained for analysis. We placed each variant into one of three a priori categories: 1) Reference (Primarily synonymous), Unlikely-to-be-Tolerated (Primarily missense), and Least-likely-to-be-tolerated (Primarily frameshift). Then we divided the proportion of variant alleles in females by the proportion in males for all variant sites. To assess the impact of ploidy, we compared the distribution of these ratios between pseudoautosomal and non-pseudoautosomal regions. In the non-pseudoautosomal regions, mean female-to-male ratios were lowest among Reference (2.40), greater for Unlikely-to-be-Tolerated (2.77) and highest for Least-likely-to-be-tolerated (3.28) variants. Corresponding ratios were lower in the pseudoautosomal regions (1.52, 1.57, and 1.68, respectively), with the most extreme ratio being 11. Because pathogenic effects in the pseudoautosomal regions should not drive ratio increases, this maximum ratio provides an upper-bound for baseline noise. In the non-pseudoautosomal regions, 319 variants had a ratio over 11. In sum, we identified a metric with a data-set specific threshold for identifying pathogenicity in non-pseudoautosomal X-chromosome variants: the female-to-male allele proportion ratio.

Evolutionary and Population Genetics Posters - Thursday

PB2820*. *MaLAdapt* reveals novel targets of adaptive introgression from Neanderthals and Denisovans in worldwide human populations

Authors:

X. Zhang¹, B. Kim², A. Singh¹, S. Sankararaman³, A. Durvasula⁴, K. Lohmueller⁵; ¹Univ. of California Los Angeles, Los Angeles, CA, ²Stanford Univ., Palo Alto, CA, ³UCLA, Los Angeles, CA, ⁴Harvard Med. Sch., Cambridge, MA, ⁵Univ California Los Angeles, Los Angeles, CA

Abstract Body:

Adaptive introgression (AI) facilitates local adaptation in a wide range of species. Many state-of-the-art methods detect AI with *ad-hoc* approaches that identify summary statistic outliers or intersect scans for positive selection with scans for introgressed genomic regions. Although widely used, these outlier-based approaches are vulnerable to a high false-negative rate as the power of different methods vary, especially for complex introgression events. Moreover, population genetic processes unrelated to AI, such as background selection or heterosis, may create similar genomic signals as AI, compromising the reliability of methods that rely on neutral null distributions. In recent years, machine learning (ML) methods have been increasingly applied to population genetic questions. Here, we present an ML-based method called *MaLAdapt* for identifying AI loci from genome-wide sequencing data. Using an Extra-Trees Classifier algorithm, our method combines information from a large number of biologically meaningful summary statistics to capture a powerful composite signature of AI across the genome. In contrast to existing methods, *MaLAdapt* is especially well-powered to detect AI with mild beneficial effects, including selection on standing archaic variation, and is robust to non-AI selection sweeps, heterosis, and demographic misspecifications. Further, *MaLAdapt* outperforms existing methods for detecting AI based on the analysis of simulated data and on a validation of empirical signals through visual inspection of haplotype patterns. We apply *MaLAdapt* to the 1000 Genomes Project human genomic data, and discover novel AI candidate regions in non-African populations, including genes that are enriched in functionally important biological pathways regulating metabolism and immune responses.

Evolutionary and Population Genetics Posters - Thursday

PB2821. Modeling recurrent mutations predicts allele frequencies and enables precise inference in large samples.

Authors:

E. Koch¹, V. Seplyarskiy¹, D. Lee¹, J. Wakeley², L-T. Fan³, S. Sunyaev¹; ¹Dept. of BioMed. Informatics, Harvard Med. Sch., Boston, MA, ²Dept. of Organismic and Evolutionary Biology, Harvard Univ., Boston, MA, ³Dept. of Mathematics, Indiana Univ., Bloomington, IN

Abstract Body:

As sample sizes of human genome sequences reach the range of millions of individuals, previously reliable assumptions in population genetics have started to break down. An important instance is the assumption that each mutation occurs only once in the history of the sample. When sample sizes are large enough and mutation rates sufficiently high, identical variants shared by multiple individuals may represent independent mutational events rather than inheritance from a recent common ancestor. We develop theoretical and computational machinery to account for recurrent mutations in the site frequency spectrum (SFS). The SFS describes variant counts at each possible sample frequency and is a useful summary of sequencing data. We apply our model to a sample of ~120k exomes (gnomAD v2.1.1) using new per-bp estimates of mutation rates (Roulette) and show that the SFS is well-predicted when mutation rates are known with sufficient accuracy. We next demonstrate how recurrent mutations can be used to estimate the remaining residual variance in mutation rate estimates. A new formula, which accounts for recurrent mutation, is presented for the SFS of rare variants in large samples from exponentially growing populations. This expression greatly improves computational efficiency relative to stochastic simulations and numerical integration. Finally, we demonstrate the importance of modeling the recurrent SFS when estimating selective constraint. Current methods applicable to contemporary human sample sizes largely only model the presence or absence of mutations. A model using the frequencies of loss-of-function variants significantly improves the detection of disease genes over presence/absence models.

Evolutionary and Population Genetics Posters - Wednesday

PB2822. Mosaic chromosomal alterations and longevity

Authors:

A. Leshchyk¹, Q. Xiang¹, A. Gurinovich², Z. Song¹, S. Andersen¹, S. Monti¹, K. Christensen³, A. Yashin⁴, M. Wojczynski⁵, K. Schwander⁶, J. Lee⁷, T. Perls¹, P. Sebastiani⁸; ¹Boston Univ., Boston, MA, ²Tufts Med. Ctr., Boston, MA, ³Univ. of Southern Denmark, Odense, Denmark, ⁴Duke Univ., Durham, NC, ⁵Washington Univ Schl Med, St Louis, MO, ⁶Washington Univ. Sch. of Med., St Louis, MO, ⁷Taub Inst. for Res. on Alzheimer's Disease and the Aging Brain, Columbia Univ., New York City, NY, ⁸Inst. for Clinical Res. and Hlth.Policy Studies, Tufts Med. Ctr., Boston, MA

Abstract Body:

Mosaic chromosomal alterations (mCAs) are structural alterations that are associated with mortality, age, cancer, cardiovascular disease, and diverse infections. The distribution of mCAs in long-lived subjects and individuals with familial longevity is not well described. We applied MOsaic CHromosomal Alteration (MoChA) caller on genome-wide genotype samples of 2025 centenarians, their siblings, and offspring and 273 unrelated controls from the New England Centenarian Study (NECS) and 3642 subjects with familial longevity and 920 controls from the Long-Life Family Study (LLFS). MoChA utilizes a Hidden Markov Model to detect mCA-induced deviations in allelic balance at heterozygous sites with Log R Ratio and B-allele frequency (BAF) with phased genotype information. We analyzed somatic mCAs in samples with genome-wide BAF phase concordance less than 0.51, LOD score greater than 10, and estimated cell fraction less than 50% are evidence of mCAs. The results of the two studies showed that autosomal mCAs spanning over 100 kbase pairs increase with older age until approximately 102 years. However, the prevalence of the subjects with mCAs tends to plateau after that age, suggesting that the accumulation of mCAs is less prevalent in long-lived subjects. We also found that offspring and siblings of centenarians accumulate fewer autosomal mCAs (fixed-effect meta-analysis for NECS and LLFS: RR=0.78, p=0.033) compared to unrelated controls. In addition, consistent with results from other studies, mCAs are associated with increased risk for mortality (HR=1.08, p=0.02) and sex (Male RR=1.37, p=4.15e-05), and impact incident events of cancer, dementia, diabetes, and cardiovascular diseases even at extreme old ages.

Evolutionary and Population Genetics Posters - Thursday

PB2823*. Neanderthal introgression modifies anti-inflammatory response in modern humans.

Authors:

C. Boye¹, Y-L. Lin², A. Findley¹, A. Alazizi¹, A. Dumaine², L. Barreiro^{2,3,4}, R. Pique-Regi^{1,5}, F. Luca^{1,5}; ¹Ctr. for Molecular Med. and Genetics, Wayne State Univ., Detroit, MI, ²Section of Genetic Med., Dept. of Med., Univ. of Chicago, Chicago, IL, ³Dept. of Human Genetics, Univ. of Chicago, Chicago, IL, ⁴Committee on Immunology, Univ. of Chicago, Chicago, IL, ⁵Dept. of Obstetrics and Gynecology, Wayne State Univ., Detroit, MI

Abstract Body:

Approximately 2% of the modern Eurasian genome contains variants that humans obtained through interbreeding with Neanderthals. Neanderthal genetic variants may have been beneficial for modern humans (adaptive introgression), for example by providing a selective advantage in new environments. Studies indicated that Neanderthal-introgressed variants may be important for response to pathogens. We previously showed that Neanderthal-introgressed variants regulate genes that are important for the transcriptional response to other environmental challenges. As we cannot study Neanderthal-derived cells directly, we used a massively parallel reporter assay called Biallelic Targeted STARR-seq to study the gene regulatory function of Neanderthal-introgressed variants in human cells. We designed a library of 40,000 synthetic DNA constructs containing introgressed variants predicted to alter gene regulation. We transfected them into human cells and measured their regulatory activity by quantifying self-transcribed enhancer sequences via RNA-seq. To assess the effect of Neanderthal introgressed variants on the anti-inflammatory response, we treated cells with dexamethasone, a synthetic glucocorticoid that inhibits Nuclear Factor Kappa B (NF- κ B) and binds to the glucocorticoid receptor (GR). For 7,129 variants we identified significant differences in regulatory function between the Neanderthal and the modern human alleles. While 5,403 Neanderthal-introgressed variants directly modify the gene expression response to glucocorticoids (response variants, adjusted $p < 0.1$). Constructs containing these response variants and those differentially expressed in response to glucocorticoids were enriched for NF- κ B and GR binding motifs. Strikingly, the introgressed Neanderthal allele resulted in higher expression in cells treated with glucocorticoids in the majority of the constructs containing response variants. Overall, our results are consistent with a role of Neanderthal introgression in modern human anti-inflammatory response, likely through modification of GR and NF- κ B binding, and motivate future studies to identify adaptive advantage of these introgression events.

Evolutionary and Population Genetics Posters - Wednesday

PB2824. Newfoundland and Labrador: A mosaic founder population of an Irish and British diaspora from 300 years ago.

Authors:

E. Gilbert^{1,2}, H. Zurel³, M. MacMillan⁴, S. Demiriz⁴, S. Mirhendi⁵, M. Merrigan⁶, S. O'Reilly⁶, A. Molloy⁷, L. Brody⁸, W. Bodmer⁹, R. Leach¹⁰, G. Mugford⁴, R. Randhawa³, C. Stephens¹¹, A. Symington⁴, G. Cavalleri^{1,12}, M. Phillips⁴; ¹Royal Coll. of Surgeons in Ireland, Dublin, Ireland, ²FutureNeuro SFI Res. Ctr., RCSI, Dublin, Ireland, ³Sequence Bioinformatics, Inc., St. John's, NL, Canada, ⁴Sequence Bioinformatics, Inc., St John's, NL, Canada, ⁵Sequence Bioinformatics, Inc, St John's, NL, Canada, ⁶Genealogical Society of Ireland, Dún Laoghaire, Ireland, ⁷Sch. of Med., Trinity Coll., Dublin, Ireland, ⁸NHGRI (NIH), Bethesda, MD, ⁹Weatherall Inst Molec Med, Oxford, United Kingdom, ¹⁰Sequence Bioinformatics, Inc., St. John's, NL, Canada, ¹¹Sequence Bioinformatics, Inc., St John's, NL, Canada, ¹²FutureNeuro SFI Res. Ctr., Dublin, Ireland

Abstract Body:

The founder population of Newfoundland and Labrador (NL) is a unique genetic resource, in part due to geographic and cultural isolation, where historical records describe a migration of European settlers primarily from Ireland and England to NL in the 18th and 19th centuries. Whilst its historical isolation, and increase prevalence of certain monogenic disorders, have been appreciated, the fine-scale genetic structure and ancestry of the population has not been well described. Understanding the genetic background on which functional, disease causing, genetic variation resides on would aid informed genetic mapping efforts in the Province. Here, we leverage dense genome-wide SNP data on 1,807 NL individuals to reveal fine-scale genetic structure in NL that is clustered around coastal communities and correlated with Christian denomination. We show that the majority of NL European ancestry can be traced back to the south-east and south-west of Ireland and England, respectively. We date a substantial population size bottleneck approximately 10-15 generations ago in NL, associated with increased haplotype sharing and autozygosity. Our results elucidate novel insights into the population history of NL, and demonstrate evidence of a population conducive to further genetic studies and biomarker discovery.

Evolutionary and Population Genetics Posters - Thursday

PB2825. No reduction in diagnostic yield of exome sequencing in prenatal and pediatric patients with non-European ancestries.

Authors:

Y. Mavura, N. Sahin-Hodoglugil, M. Kvale, J. Van Ziffle, W. P. Devine, U. Hodoglugil, P-M. Martin, B. A. Koenig, S. Ackerman, A. Slavotinek, P-Y. Kwok, M. Norton, N. Risch; Univ. of California San Francisco, San Francisco, CA

Abstract Body:

Background: It has been hypothesized that the diagnostic yield (DY) from Exome Sequencing (ES) is higher among patients of European ancestry compared to those with non-European ancestries because of poor representation of racial/ethnic minorities in clinical genomic sequencing studies. **Methods:** A diverse cohort of 845 pediatric and prenatal patients with suspected genetic disorders underwent ES for potential diagnosis. Using patients' ES data, continental genetic ancestry proportions were obtained with 1KG and HGDP samples as reference. Race and ethnicity information was obtained by self-report from the parents. We examined the effects of genetic ancestry and race/ethnicity on the odds of receiving a positive versus negative diagnosis by logistic regression. The analyses were conducted overall and stratified by mode of inheritance (dominant de novo, dominant inherited, recessive-compound heterozygous, recessive-homozygous, X-linked). **Results:** Overall, the parents were 41% Latinx, 18% white, 5% East Asian, 4% African American, 3% South Asian, 3% Central Asian, 2% Middle Eastern, 1% Native American, 1% Pacific Islander, 7% multiethnic, and 15% with missing information. The average genetic ancestry proportions in the patients were 47.7% European, 23.2% Native American, 9.8% East Asian, 8.2% Middle Eastern, 6.0% African, 5.4% South Asian. Of the 845 patients, 23.9% had a positive (definitive or probable) diagnosis. The majority of positive cases resulted from autosomal dominant inheritance. For all modes of inheritance combined, there was no difference in DY associated with any race/ethnicity or genetic ancestry. However, a significant increase in DY for autosomal recessive homozygous inheritance was observed for Central Asians ($P=.001$), and a suggestive increase for South Asian and Middle Eastern subjects; a similar statistically significant increase was associated with South Asian and Middle Eastern genetic ancestry ($P=.001$ for both). However, when a child consanguinity coefficient (estimated from the exome data) was included in the logistic model, both the race/ethnicity and genetic ancestry terms were no longer significant, while the consanguinity coefficient was significant ($P=.017$). **Conclusions:** ES in this setting of prenatal and pediatric undiagnosed diseases had a comparable DY for all studied race and ethnicity groups, in particular for autosomal dominant and compound heterozygous recessive inheritance, and DY was also not associated with any genetic ancestry. Recessive homozygous inheritance was increased in those of Middle Eastern, South Asian and Central Asian background due to an increase in parental relatedness.

Evolutionary and Population Genetics Posters - Thursday

PB2826. On the Genes, Genealogies, and Geographies of Quebec

Authors:

L. Anderson-Trocme¹, D. Nelson¹, S. Zabad¹, A. Diaz-Papkovich¹, N. Baya², B. Jeffery³, C. Dina⁴, H. Vézina⁵, J. Kelleher³, S. Gravel¹; ¹McGill Univ., Montreal, QC, Canada, ²Broad Inst. of MIT and Harvard, Cambridge, MA, ³Big Data Inst., Li Ka Shing Ctr. for Hlth.Information and Discovery, Oxford, United Kingdom, ⁴Nantes Univ., Nantes, France, ⁵Université du Québec à Chicoutimi, Chicoutimi, QC, Canada

Abstract Body:

The joint analysis of large-scale genetic and genealogical records provides a unique opportunity to bridge the gap between idealized population genetic models and the actual, intricate demographic histories that they seek to capture. This study considers the combination of a pedigree compiled from four million parish records capturing primarily French-Canadian (FC) population history dating back to the first arrivals of French settlers four centuries ago, and genotype data from 2,276 French individuals and 20,451 genotyped FC individuals, of which 4,882 have been linked to the pedigree. Though genetics reflect known differential contributions from French regions during colonization, we show that these differences are shared across Quebec regions. Most population structure within Quebec thus occurred after settlement. To trace the appearance of this population structure over time, we extend the msprime simulation software to handle large-scale pedigree-aware simulations and applied it to the parish record data. We obtain an excellent quantitative description of most aspects of fine-scale FC population genetic structure. We highlight three regional founding events in the regions of Saguenay-Lac-Saint-Jean, the Bas-Saint-Laurent and Beauce that have led to major axes of genetic variation and followed spatial dynamics reminiscent of distinct classical population genetic models. Despite these distinct regional founding patterns, geographic features played an important role in shaping migrations throughout, and we find enrichments for migration, genetic and genealogical relatedness patterns within river networks across Quebec regions. Finally, we provide a freely accessible simulated whole-genome sequence dataset with spatiotemporal metadata for 1,426,749 individuals linked to a genealogy and reflecting realistic and complex population structure.

Evolutionary and Population Genetics Posters - Wednesday

PB2827. Patterns of selection on human gene regulatory variation

Authors:

L. Colbran, I. Mathieson; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

The majority of variants associated with phenotypes in modern human populations, and identified in genome-wide scans for selection, are non-coding. Therefore, interpreting the effects of these variants and understanding the mechanisms behind phenotypic variation are limited by our understanding of gene regulatory processes and difficulty in linking non-coding variants to genes. To overcome these challenges, we developed a gene-by-gene test for population-specific selection based on combinations of regulatory variants.

We extended the Qx test for polygenic selection to gene expression models trained using joint-tissue imputation, a transcriptome-wide association method trained on paired genotype and RNA-seq data. We first identified the most accurate expression model to use for each gene by comparing predicted expression with observed expression in lymphoblastoid cell lines (LCLs) from 447 individuals from the 1000 Genomes (1kG) project. We find that, while LCL-specific predicted expression models were not correlated with actual LCL expression (median $\rho = 0.00081$, maximum $\rho = 0.19$), the models that explained the highest variance in expression during training were more highly correlated (median $\rho = 0.021$, maximum $\rho = 0.93$), regardless of the training tissue. Therefore, we used effect sizes and variants from these best performing models to calculate Qx statistics for 27,000 genes based on allele frequencies in the 26 1kG populations. We found that Qx was correlated with the number of variants and other technical characteristics of the gene regulation models. To control for these effects, we calculated p-values by permuting the effect sizes of variants for each gene while holding allele frequencies and linkage disequilibrium structure constant.

We identified 10 genes showing significant evidence ($FDR < 0.1$) for population-specific selection, including *STXBP4*, *MYO5B*, and *HDLBP*. Despite having one of the highest raw Qx scores and previously characterized regulatory selection, *FADS1*, is not significant. This suggests that the large Qx score for *FADS1* is driven by differing allele frequencies at relevant variants (as detected by other tests for selection), rather than consistent selection on a large number of regulatory variants. This interpretation is supported by the low correlation of significance with haplotype-based selection tests (gene rank correlation of 0.02 for iHS vs Qx P-value). Thus, our test may be useful in combination with more traditional selection tests. These results demonstrate the potential in combining population-level information with basic biology to understand human traits and evolution.

Evolutionary and Population Genetics Posters - Thursday

PB2828. Phased, long read assemblies from Central African participants represent a more comprehensive, inclusive future for human pangenomics

Authors:

J. LoTempio¹, D. Spencer¹, K. Mosema², K. Karume³, M. Almalvez¹, Y. Fu¹, C. Musasa¹, J. Nsibu², M. Bramble¹, E. Kamangu³, D. Tshala-Katumbay⁴, D. Mumba², E. Vilain¹; ¹Children's Natl., Washington, DC, ²INRB, Kinshasa, Congo, Democratic Republic of the, ³Univ. of Kinshasa, Kinshasa, Congo, Democratic Republic of the, ⁴Oregon Hlth.and Sci. Univ., Portland, OR

Abstract Body:

Background:

The challenge of representation of diverse populations in genomics is often discussed, and indeed precision medicine cannot be achieved without global participation. However, the solution, on the ground collaboration, can be hard to achieve. While the cost of generating a telomere-to-telomere genome assembly is out of reach for most laboratories, it is feasible to generate human assemblies which approach reference quality. This brings the goal of these collaborations, reference quality genomes, closer. To this end, our team has translated ideas and concepts of equity and inclusion to realized recruitment and consent of participants from genetically diverse trios from the Democratic Republic of the Congo for high quality genome sequencing.

Methods:

Our ethics protocol was based on the 1000 Genomes Project protocol. Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood from participants. Each member of the trio was sequenced with Illumina to a depth of 30x. DNA was extracted from PBMCs and was sequenced on four PacBio HiFi flowcells for each proband, representing approximately 300x coverage. Assemblies were trio binned with yak, assembled with HiFiASM, and scaffolded with Bionano optical maps.

Results:

Trios were recruited in Kinshasa through our network of physician scientists who prioritized geographic and ethnolinguistic diversity. We chose four trios for sequencing by selecting those thought to be historically distant from one another. With these assemblies, we show that unmapped reads from experiments on novel genomes from Congo, aligned to GRCh38p13, can indeed be placed. We further show that African Pangenomic Contigs from (Sherman et al 2019) can be placed on these genomes, resolving outstanding issues in missing data.

Conclusions:

These HiFiASM assemblies, which will be available as a broad use open resource, represent a strong advancement for placement of known, unmapped genomic data, and thus a step forward in precision medicine. Bionano hybrid scaffolds, while an intermediate product in need of polishment, may help groups achieve more contiguous assemblies at low cost. Together, the accessibility of these technologies will contribute to an environment where concepts can be translated to grassroots collaborations that result in the next era for pangenomics.

Evolutionary and Population Genetics Posters - Wednesday

PB2829. Population-specific Recombination Maps Reduce False Positive Findings in Genomic Scans of Adaptations

Authors:

B. Dinh¹, L. Wilkens², L. Le Marchand², C. Haiman¹, C. W. K. Chiang¹; ¹Univ. of Southern California, Los Angeles, CA, ²Univ. of Hawaii, Honolulu, HI

Abstract Body:

Knowledge of the recombination landscape informs patterns of haplotypic structure in a population and is used in a number of population and statistical genetic analyses such as imputation, local ancestry inference (LAI), identity-by-descent (IBD) inference, and detection of adaptation. However, populations not represented by the 1000 Genomes Project (1KGP) typically do not have a corresponding map. Yet, fine-scale differences in the recombination landscape are known to exist between populations and it may be suboptimal to utilize the recombination map from a distantly related population. The lack of representation may thus contribute to obstacles in effectively extending genomic analyses to understudied populations. To begin to bridge this gap, we create LD- and IBD-based recombination maps using genome-wide array data for the Native Hawaiian population (NH), a minority population encompassing only 0.5% of the U.S. population, with the hope to help improve future studies in NH and other Polynesian cohorts. We find NH maps to be less correlated with those from other 1KGP populations (Spearman's $\rho = 0.79$ between NH and CEU, compared to 0.92 between YRI and CEU at 1Mb resolution) and share the least amount of recombination hotspots with other populations from 1KGP (at 50kb scale, 8% of the top 1 percentile of hotspots are shared between NH and 1KGP CEU, compared to 27% shared between 1KGP YRI and CEU). Additionally, we test the impact of the maps on downstream analyses. We find that NH maps have relatively little impact on analyses such as LAI and imputation. The local ancestry calls are > 95% concordant when comparing the results of the NH map and the omnibus map. Imputation accuracies of variants with minor allele frequency > 5% are 92% for both NH map and the omnibus map when assessed against targeted exon sequencing data of the same individuals. Distribution of called IBD segments and the subsequent relatedness inference are also similar when using the NH-specific map or the omnibus map. However, for genome scans of signals of adaptation using statistics such as integrated haplotype score (iHS), we find a number of loci with apparent genome-wide significant signals (Z -score > 4) that would not be exceptional when analyzed using NH-specific maps. Examination of genes near these ostensibly adaptive loci also do not provide a functional basis for selection. Our results thus suggest that population-specific recombination maps can improve the robustness of haplotype-based statistics such as those used in selection scans, which can help us better characterize adaptive loci that may underlie Hawaiian-specific health conditions today.

Evolutionary and Population Genetics Posters - Thursday

PB2830. Presenilin-1 G206A mutation in Puerto Ricans with Alzheimer Disease reveals it is as a founder event on the African haplotype

Authors:

K. Celis¹, S. Slifer², K. Hamilton-Nelson², F. RAJABLI¹, N. R. Gomez¹, P. Whitehead³, M. Contreras³, S. Tejada³, P. Mena¹, J. J. Sanchez¹, J. Arvizu³, C. G. Golightly¹, L. Adams¹, T. Starks⁴, S. Concepcion¹, M. Cuccaro³, J. Vance¹, G. Byrd⁵, J. Haines⁶, G. Beecham¹, B. Feliciano⁷, A. Griswold¹, M. Pericak-Vance²; ¹Univ. of Miami, Miami, FL, ²Univ. of Miami Miller Sch. of Med., Miami, FL, ³John P. Hussman Inst. for Human Genomics, Miami, FL, ⁴Wake Forest Sch. of Med., Winston-Salem, NC, ⁵NC A&T State Univ, Greensboro, NC, ⁶Case Western Reserve Univ., Cleveland, OH, ⁷PRADI, UCC, SJB Sch. of Med., Caguas, PR

Abstract Body:

Background: Variants in the presenilin-1 gene (*PSEN1*) are known to be pathogenic for Alzheimer disease (AD). The variant G206A in *PSEN1* has been identified in AD Caribbean Hispanic families from Puerto Rico with variable ages of onset and incomplete segregation (Athán et al., 2001). Puerto Ricans have an admixed genetic background which includes European, African and Amerindian ancestry. We investigated the role of G206A in the genetic etiology of AD in Puerto Rico, the ancestral background and founding haplotype in this admixed population. **Methods:** We screened 896 individuals (182 from families; 714 unrelated cases/controls) from the Puerto Rican Alzheimer Disease Initiative (PRADI) using whole genome sequencing (WGS) and identified carriers of G206A. Genotyping data were phased using SHAPEIT to identify local ancestry of the *PSEN1* haplotype followed by RFMix to estimate the genetic ancestry. Haplotype modeling was performed using MERLIN software. In addition, we tested p-Tau181 levels from plasma protein using Simoa chemistry assays (Quanterix HD-X) and data was analyzed using HD-X Analyzer Software v1.6. **Results:** We identified 38 carriers of the G206A variant. 25 carriers (20 AD, two MCI, one neuropsychiatric disorder, and two cognitively unimpaired individuals under 65 years old) were identified in eight families, with the largest family having nine carriers. G206A did not completely explain AD in these eight families as eight other AD cases in the families did not carry the variant. The additional 13 carriers identified were part of the PRADI cases/controls cohort (10 AD, one MCI, and two cognitively unimpaired under 65 years old). Clinically, G206A carriers with AD varied in age of onset (55% with onset before age 65), severity measured by CDR (42% CDR>2) and *APOE* genotype (60% were *APOE3* carriers). Local ancestry indicated that the mutation arose on an African ancestral haplotype. However, no other carriers were identified while screening 4,348 WGS of individuals from Nigeria and African American cohorts. ~37% of the G206A carriers were located in small area of (PR Zone 2), suggesting this region contained the original founding individual. We evidence an increase of p-Tau181 levels in G206A carriers with AD compared to carriers without AD. **Conclusion:** Our results support that G206A contributes to AD in the Puerto Rican population but does not completely explain the genetic risk in multiplex PRADI AD families. We show that the G206A variant occurs on a common haplotype across carriers representing a founder event on an African haplotype background in Puerto Rico.

Evolutionary and Population Genetics Posters - Wednesday

PB2831. Primate comparative genomics and infectious diseases; COVID-19, Monkey Pox and what comes next

Authors:

M. Roodgar¹, X. Shen¹, R. Purohit¹, L. Ramirez¹, M. Snyder²; ¹Stanford Univ., Palo Alto, CA, ²Stanford Univ., Stanford, CA

Abstract Body:

Nonhuman primates (NHP), specifically macaques, share approximately 93% genome homology with humans, making them the most appropriate animal models to study human infectious diseases such as COVID-19, tuberculosis (TB), and HIV. With the emergence of infectious diseases like COVID-19 and recently monkey pox, there is an increasing need in better understanding primate immunogenomic features that correlates with human disease susceptibility and/or disease transmission. In this study, we perform comparative genomics and comparative single-cell transcriptome immune profiling of peripheral blood mononuclear cells (PBMCs) from four macaque species and humans. Rhesus macaque, cynomolgus macaque, pig-tailed macaque, and African green monkeys are NHP species that are the most commonly used animal models to study human infectious diseases including COVID-19. The comparative genomics and immune cells transcriptomic of these macaque species reveals high homology of several immune cell subtypes of pig-tailed macaque with those of humans. Additionally, the comparative analysis of immune gene orthologous isoforms leveraging long-read sequencing reveals novel species-specific immune orthologs and primate repetitive elements. These novel isoforms and repeat elements play a crucial role in resistance to infectious diseases. The genomics information from our study has broad application in understanding human immune system and development of new therapies leveraging primate comparative immunogenomics.

Evolutionary and Population Genetics Posters - Wednesday

PB2833. Recovering signatures of ghost admixture using ancestral recombination graphs.

Authors:

A. Biddanda¹, Y. Zhang², C. O'Dushlaine¹, P. Moorjani²; ¹54 Gene, Washington, DC, ²UC Berkeley, Berkeley, CA

Abstract Body:

With the sequencing of archaic hominin (e.g. Neanderthal and Denisovan) genomes, the contributions of these groups to modern human populations have been well-characterized over the past decade in non-African populations. However, the understanding of archaic ancestry in other modern genomes, particularly contemporary Africans, remains poorly understood due to technical challenges in DNA preservation. Recent evidence suggests that most African groups have ~2-20% of ancestry from an unknown “ghost” archaic hominin (Durvasula and Sankararaman, 2020), and another plausible explanation poses that these patterns can be explained by weak population structure in the distant past of African populations (Ragsdale et al, 2022). Despite these broader results, the model of gene flow remains unclear, since no archaic reference genome for this “ghost” ancestry or outgroup without admixture is available for current methods.

Recent advances in methods to reconstruct the ancestral recombination graph (ARGs) (e.g. Spiedel et al, 2019, Kelleher et al 2019) from large numbers of sequenced genomes has enabled more detailed reconstruction of evolutionary history, including historical gene flow. Importantly, the inferred ARGs from these methods have not been applied to evaluate different scenarios of gene flow between human populations, particularly in the case of ghost ancestry.

We introduce a new approach to characterize the history of archaic ancestry using estimated ARGs from modern genomes alone. Our method is a hidden markov model (HMM) that leverages two key features of genealogies that are distorted in the face of introgression: “long branches”- deep lineages in the coalescent tree, and “long haplotypes”- characteristic of recent gene flow unbroken by recombination. By conducting simulations under a range of demographic scenarios, we characterize the sensitivity and specificity of our method to previously developed methods using ground-truth ARGs and inferred ARGs. Our method works reliably with true ARGs, outperforming available methods to detect archaic admixture, but exhibits lower sensitivity when using inferred ARGs. We also formally compare the models of ancestry in African populations within the HMM - “ghost” introgression or weak population structure, to quantify the ability of ARG-based summary statistics to evaluate competing hypotheses regarding the dynamics of gene flow.

While we have focused on scenarios relevant to human evolutionary history, particularly in Africa, the proposed approach is generalizable to any organism for which reliable ARGs can be estimated and where appropriate archaic references may not be available.

Evolutionary and Population Genetics Posters - Thursday

PB2834. Relatedness inference up to 3rd degree from low coverage ancient genomes in presence of contamination, long runs of homozygosity and ascertainment bias

Authors:

D. Popli, S. Peyrégne, B. M. Peter; Max Planck Inst. for Evolutionary Anthropology, Leipzig, Germany

Abstract Body:

Identifying related individuals from genetic data is a common task in population genetic studies. Inferring relatedness from ancient DNA data is often difficult because sequence coverage is typically low, and present-day DNA contamination is common. In addition, many ancient populations have long runs of homozygosity (ROH) that could impact analyses. Here, we present KIN, a Hidden-Markov-Model-based method to infer relatedness by identifying the parts of the genomes that are shared by descent between a pair of individuals, taking contamination and ROH in each individual into account. Using simulations, we show that we can accurately infer up to 3rd degree relatedness from a sequence coverage as low as 0.1x. We first applied KIN to genetic data from a set of 16 Neandertal specimens from Chagyrskaya Cave, Siberia, generated by targeted enrichment for 713,000 SNPs across the genome. The coverage ranges from 0.01x to 12.34x, and several libraries show evidence of contamination. We find long ROH in most individuals, and are able to detect three specimens that came from the same individual, a father-daughter pair, and a pair of 2nd degree relatives. We also apply Kin to 118 Bronze age individuals from Germany, and compare it two other methods, READ and lcMLkin. We find that the methods generally agree, but KIN is often able to make inference where lcMLkin would require more data, and unlike READ, we are able to distinguish between parent-child and siblings.

Evolutionary and Population Genetics Posters - Wednesday

PB2835. Relating the evolutionary fitness costs of loss-of-function mutations to their pathogenic consequences in humans

Authors:

I. Agarwal^{1,2}, Z. L. Fuller², S. Myers¹, M. Przeworski²; ¹Univ. of Oxford, Oxford, United Kingdom, ²Columbia Univ., New York City, NY

Abstract Body:

A number of recent disease studies have reported an enrichment of causal variants in “mutation intolerant genes”. Here, we consider this relationship explicitly, by relating pathogenic consequences at present-day to inferred fitness effects of loss of function mutations. To this end, we first infer posterior distributions for the evolutionary fitness costs of loss-of-function (LOF) mutations for 17,744 autosomal and 679 X-linked genes, given the observed frequencies of LOF mutations in 56,855 individuals. Estimated fitness costs are on the order of 1% on average; they are typically largest for X-linked genes, whether or not they have a Y homolog, followed by autosomal genes, and smallest for genes in the pseudoautosomal region. We then compare the distribution of fitness effects (DFE) for all possible de novo LOF mutations to the distribution for mutations identified in pedigree studies of individuals diagnosed with one of six severe complex diseases. All six cohorts are enriched for mutations with estimated fitness effects in excess of 10%, indicating that causal mutations tend to be highly deleterious, and are likely to be quite recent in origin. We see an effect of disease severity, with a larger enrichment of such highly deleterious variants for early onset developmental disorders than Schizophrenia or Tourette syndrome. We also detect influences of the pedigree study design on the types of mutations detected for a given diagnosis: thus, we find that in autism studies, mutations identified in probands of simplex family studies are more deleterious on average than those in multiplex family studies, and mutations found in female probands are more deleterious on average than those in males. More generally, this approach allows us to characterize the types of mutations that contribute to complex disease risk and how they are transmitted in populations.

Evolutionary and Population Genetics Posters - Thursday

PB2836. Repeats R/T ratio is associated with Human Y chromosome Haplogroups

Authors:

T. Puurand¹, T. Kivisild²; ¹Univ. of Tartu, Tartu, Estonia, ²KU Leuven, Leuven, Belgium

Abstract Body:

SNP based Y chromosome haplogroup determination for ancestry and relatedness analyses is challenging from ultra-low coverage (<0.1x) sequence data that is often achievable from poorly preserved human remains. These challenges are due to low probability of critical clade-defining markers to be covered by data. While most previous research on genetic ancestry in the last two decades has focused on single nucleotide variants, the completion of human genome sequencing is opening new avenues for the use of structural variation for the analysis of genetic ancestry. More than one half of the human genome contains repeated sequences and the completed sequence of human Y chromosome includes extensive heterochromatin regions. We show how specific k-mers can be used to predict Y chromosome haplogroups and to assess variation in tandemly repeated and single copy regions for the assessment of genetic ancestry as well as, potentially for genetic relatedness. Random repeat (Line) k-mer frequency / tandem repeat k-mer frequency ratios (R/T ratios) of 198 Estonian males are graphically visualized <https://bioinfo.ut.ee/randomtandem/EGV.html>. We show that the clustering of the male samples by the counts of the three Y chromosome-specific tandem repeats (vntr, centromere, heterochromatin) clearly separates the main haplogroups (N3a, I, R1a, R1b, E). We have explored the potential for R/T ratios-based method for assessing specific tandem repeat length ratios in Y chromosome and using these ratios as predictors of Y chromosome haplogroups. We have tested the approach on three Y chromosome-specific k-mers using high coverage (~20x Y chromosome) Illumina WGS reads as a proof of principle. The method can be improved by using more and shorter k-mers for low coverage short read sequencing applications like in NIPT, forensic and ancient DNA studies. Variation inside haplogroups shows potential for ancestry detection from non-imputed data for individual identification, relatedness assessment and for the basic understanding of the rate of structural variation in human genome.

Evolutionary and Population Genetics Posters - Wednesday

PB2837. Resolution of structural variation in diverse mouse genomes reveals dynamic chromatin remodeling due to transposable elements

Authors:

A. Ferraj^{1,2}, P. Audano¹, P. Balachandran¹, J. Flores¹, I. Walawalkar¹, A. Czechanski³, L. Reinholdt³, C. Beck^{1,2}; ¹The Jackson Lab. for Genomic Med., Farmington, CT, ²Univ. of Connecticut Hlth.Ctr., Farmington, CT, ³The Jackson Lab., Bar Harbor, ME

Abstract Body:

Inbred laboratory mice are one of the most popular biomedical research models. Diverse substrains of mice have been utilized to create populations that exhibit genotypic and phenotypic heterogeneity and have enabled mapping of phenotypes to genomic loci. The underlying causes of phenotype variation and the discovery of genetic causes are dependent on high-quality reference genomes and variant catalogues for each of the progenitor strains of mice. Despite recent efforts to construct strain-specific reference genomes, much of the genomic landscape that is variable between these mice remains incomplete, which is attributable to the technological limitations of short-read whole genome sequencing. Recent advances in long-read sequencing have since surpassed the longstanding inability to accurately resolve repetitive DNA and genetic variants of large size. Although large scale studies have utilized long-read sequencing in distinct human populations, to date there have been no efforts to fully resolve and detect structural variation (variants ≥ 50 bp) among diverse mouse genomes with these technologies. We performed Pacific Biosciences (PacBio) long-read sequencing in order to sequence-resolve structural variants across 20 genetically diverse inbred laboratory mouse strains. All variants were merged into a complete catalogue of non-redundant SV calls spanning three major subspecies of *mus musculus*: *domesticus*, *castaneus*, and *musculus*. We identified 413,969 SVs which occur at unique sites in the reference, cumulatively affecting $\sim 13.5\%$ (368 Mb) of the mouse genome. Furthermore, we have detected subspecies-specific as well as inter-subspecies SVs that exhibit high impact consequences, including 654 coding sequence SVs and variants that affect the landscape of mESC chromatin accessibility and gene expression. Our data comprehensively identify the extensive SVs present in diverse mouse genomes and strongly suggest that SVs contribute to phenotypic diversity between recently diverged mice.

Evolutionary and Population Genetics Posters - Thursday

PB2838. Resurrecting the alternative splicing landscape of archaic hominins using machine learning.

Authors:

C. Brand, J. A. Capra; Univ. of California San Francisco, San Francisco, CA

Abstract Body:

While some phenotypic differences between archaic hominins and modern humans are known from the fossil record, most archaic tissues have not survived to the present. Thus, most phenotypes cannot be directly compared. Divergence in gene expression is thought to be a major driver of the phenotypic differences between these closely related groups. Alternative splicing, which increases proteomic diversity, has been shown to contribute to adaptation in various species. Alternative splicing can be measured in extant taxa and machine learning can be used to predict splicing from sequence alone, unmasking this previously unobservable mechanism and associated phenotypes. Here, we apply SpliceAI, a machine learning algorithm that predicts the probability a variant is splice altering, to high-coverage genomes from three Neanderthals and a Denisovan. We identify 5,950 putative splice altering variants, of which 2,343 are archaic-specific and 3,607 also occur in modern humans via introgression or shared ancestry. Collectively, archaic-specific variants are enriched in genes that contribute to many traits such as excessive sweating, platelet volume, and spinal rigidity. We also recovered strong, distinct phenotype enrichment among lineage-specific splice variants. We anticipate that most splice variants have large effects and are likely to be removed by purifying selection. Indeed, archaic-specific variants occur in sites under weaker selection than shared splice variants. Further, we find that high-probability archaic splice variants are more frequently present in more tissue specific genes. We also find that archaic lineages with low effective population sizes are enriched for splice variants compared to shared variants, consistent with selection against most splice variants. We quantified the contribution of introgression to splicing in modern humans, finding that introgressed splice variants were largely shared across all three Neanderthals, suggesting older splice variants were most tolerated in modern human genomes. Our results quantify the previously unobservable splicing landscape of archaic hominins and suggests that this mechanism may underlie phenotypic differences between archaics and modern humans.

Evolutionary and Population Genetics Posters - Wednesday

PB2839. Risk of venous thromboembolism in individuals with supernumerary sex chromosome aneuploidies in two large population-based cohorts

Authors:

A. Berry¹, B. M. Finucane¹, S. M. Myers¹, D. H. Ledbetter², C. L. Martin¹, M. T. Oetjens¹; ¹Geisinger Hlth.System, Lewisburg, PA, ²Univ. of Florida, Gainesville, FL

Abstract Body:

Venous thromboembolism (VTE), which includes deep venous thrombosis (DVT) and pulmonary embolism (PE), is a preventable and treatable condition with high morbidity and mortality. An increased risk for VTE has previously been reported in men with Klinefelter syndrome; however, the prevalence of VTE in other sex chromosome aneuploidies (SCA) has not been well-characterized. Here, we performed an observational study of VTE and SCA by analyzing X- and Y-chromosome dosage and VTE prevalence in 642,544 individuals from two population-scale biobanks, Geisinger's MyCode Community Health Initiative and the UK Biobank. We sought to determine if VTE risk is associated with SCA. The prevalence of SCA was 1 in 528 adults in MyCode and 1 in 832 adults in the UK Biobank, in line with previous estimates of 1 in 450 to 1 in 1,500. Relative to adults with two sex chromosomes, the odds of having a VTE diagnosis in MyCode was four- to six-fold as high among individuals with 47,XXX (OR, 4.5; 95% CI, 2.5-7.7, $P < 1 \times 10^{-6}$), 47,XYY (OR, 3.7; 95% CI, 1.8-7.2, $P < 1 \times 10^{-3}$), and 47,XXY (OR, 6.2; 95% CI, 3.8-10.0, $P < 1 \times 10^{-12}$). Replication in the UK Biobank yielded similar results: the odds of having a VTE diagnosis before the baseline interview was four- to seven-fold as high among individuals with 47,XXX (OR, 4.4; 95% CI, 2.3-7.6, $P < 1 \times 10^{-6}$), 47,XYY (OR, 6.7; 95% CI, 4.0-10.7, $P < 1 \times 10^{-13}$), and 47,XXY (OR, 5.7; 95% CI, 3.6-8.5, $P < 1 \times 10^{-14}$) compared to adults with two sex chromosomes. Additionally, the presence of a supernumerary sex chromosome was associated with a 4.2-fold (95% CI, 2.8-6.3, $P < 1 \times 10^{-11}$) increased risk of VTE in the 10 years following initial assessment in the UK Biobank, compared to individuals with two sex chromosomes. This increased risk was mediated in part by known anthropometric factors and blood biomarkers associated with both VTE and supernumerary SCA, including height and cystatin C; however the majority of the association between VTE and supernumerary SCA remains unexplained (95% CI, 50-74%, $P < 1 \times 10^{-4}$). Our data suggest that supernumerary SCA are underappreciated risk factors for VTE, comparable in magnitude to heritable forms of thrombophilia. Therefore, VTE evaluation and prophylaxis should be considered for adults diagnosed with these relatively common chromosomal disorders. Conversely, the medical work-up of adults who experience VTE should prompt consideration of diagnostic genetic testing for SCA, particularly if other clinical SCA characteristics (e.g., tall stature) are present, to establish an underlying etiology for VTE.

Evolutionary and Population Genetics Posters - Thursday

PB2840. SALAI-Net: Species-agnostic local ancestry inference network.

Authors:

B. Oriol Sabat¹, D. Mas Montserrat², X. Giró-i-Nieto³, A. Ioannidis⁴; ¹Univ. of California Los Angeles, Los Angeles, CA, ²Stanford Univ., Woodside, CA, ³Univ. Politecnica de Catalunya, Barcelona, Spain, ⁴Stanford Univ., Palo Alto, CA

Abstract Body:

Local Ancestry Inference (LAI) is the high resolution prediction of ancestry categories across a DNA sequence. LAI is becoming increasingly important in the study of human history, migrations, and precision medicine applications, including genome-wide association studies (GWAS) and polygenic risk scores (PRS).

Most of modern LAI models do not generalize well between species, chromosomes, or even ancestry groups, requiring re-training for each different setting. Furthermore, such methods can lack interpretability, which is a paramount element in biomedical applications.

We present SALAI-Net, a portable statistical LAI method that can be applied on any set of species and ancestries (species-agnostic), requiring only haplotype data and no other biological parameters. Inspired by Identity By Descent (IBD) methods, SALAI-Net estimates population groups by performing a reference matching approach, which leads to an interpretable and fast technique. We benchmark our models on whole genome data of humans and we test its ability to generalize to dog species when trained on human data. SALAI-Net outperforms previous methods in terms of balanced accuracy while generalizing between different settings, species, and datasets. Moreover, it is up to two orders of magnitude faster and uses considerably less RAM memory than competing methods. We provide an open source implementation and there is publicly available data and trained models on <https://github.com/AI-sandbox/SALAI-Net>.

Evolutionary and Population Genetics Posters - Wednesday

PB2841. Sequence variants affect the genome-wide rate of germline microsatellite mutations.

Authors:

S. Kristmundsdóttir, H. Jónsson, M. T. Hardarson, G. Palsson, D. Beyter, H. Eggertsson, A. Gylfason, G. Sveinbjornsson, G. Holley, O. Stefansson, S. Olafsson, G. Árnadóttir, P. Olason, O. Eiriksson, G. Masson, U. Thorsteinsdóttir, T. Rafnar, P. Sulem, A. Helgason, D. Gudbjartsson, B. Halldórsson, K. Stefansson; deCODE Genetics, Reykjavik, Iceland

Abstract Body:

Microsatellites are polymorphic tracts of short tandem repeats (STRs) with one to six base-pair (bp) motifs and are some of the most polymorphic markers in the genome. Using 6,084 Icelandic parent-offspring trios we estimate 63.7 (95% CI: 61.9-65.4) microsatellite de novo mutations (mDNMs) per offspring per generation. Paternal mDNMs occur at longer repeats, while maternal mDNMs affect more bp. mDNMs increase by 0.97 (95% CI: 0.90-1.04) and 0.31 (95% CI: 0.25-0.37) per year of father's and mother's age at conception, respectively. We found two independent coding variants associated with an increased number of mDNMs transmitted to offspring; A missense variant (allele frequency (AF) = 1.9%) in *MSH2*, a mismatch repair gene, increases transmitted mDNMs from both parents (effect: 13.1 paternal and 7.8 maternal mDNMs). A synonymous variant (AF = 20.3%) in *NEIL2*, a DNA damage repair gene, increases paternally transmitted mDNMs (effect: 4.4 mDNMs). Thus, the microsatellite mutation rate in humans is in part under genetic control.

Evolutionary and Population Genetics Posters - Thursday

PB2842. Sequencing of 230 primate species identifies conserved enhancers underlying complex human disease

Authors:

S. Rashid¹, L. Kuderna¹, A. Cox¹, J. Ulirsch¹, Primate Sequencing and Conservation Consortium, J. Rogers², T. Marques-Bonet³, K-H. Farh¹; ¹Illumina, Foster City, CA, ²Baylor Coll. of Med., Houston, TX, ³Inst. of Evolutionary Biology (UPF-CSIC), Barcelona, Spain

Abstract Body:

Novel functional genomic elements in the primate lineage are prime candidates for understanding the changes that have contributed to the uniqueness of our own species. Although comparisons between the human genome with those of other mammal and vertebrate species have revealed an extensive catalog of conserved genes and regulatory elements, the conserved sequence elements in human that arise particularly from the primate lineage have been very challenging to identify due to the short evolutionary distances separating primate species. At these short timescales, it is unclear whether the absence of changes between species is due to evolutionary constraint, or simply because there has not been the opportunity for random mutations to arise. Compared to the mammalian lineage, which includes over 6000 species separated by ~200 million years of evolution, the primate lineage only consists of 521 species that are separated by a fraction of this time. Prior studies sequenced and aligned only a fraction (43/521) of these primate species constituting a total phylogenetic branch length of only 10% that of the placental mammal alignment. These challenges make it clear that a successful strategy to identify conserved sequence elements in human that stem solely from primate lineage will require improved algorithms for detecting conservation at short evolutionary distances, as well as sequencing data from a far greater number of primate species. To address these challenges, we constructed a multiple sequence alignment of 230 primate species, representing nearly half of all extant species in the primate lineage. Even with 230 primate species, prior methods to calculate per base conservation genome-wide provided insufficient resolution due to the short branch lengths in the primate lineage. To address this, we developed a method for finding conserved nucleotides at very short evolutionary distances, using deep learning to model mutation rates. We identified over 208,000 conserved DNase-hypersensitivity sites and 887,000 conserved transcription factor binding sites in human genome that arise exclusively from the primate phylogenetic branch and not across other placental mammals. However, we observed that only a small fraction of protein-coding exons and genes in the human genome are exclusively conserved in primates. Our results highlight the central role of novel regulatory sequence elements rather than protein-coding sequence in the emergence of humans from other non-primate placental mammals.

Evolutionary and Population Genetics Posters - Wednesday

PB2843. Signatures of Mutational Processes in Human DNA Evolution

Authors:

D. Ebrahimi¹, H. Hamidi², A. Khosh³, H. Alinejad-Rokny⁴, T. Coorens⁵, R. Sanghvi⁶, S. Lindsay⁵, R. Rahbari⁷; ¹Texas BioMed. Res. Inst., San Antonio, TX, ²Univ. of Calgary, Calgary, AB, Canada, ³Yazd Univ., Yazd, Iran, Islamic Republic of, ⁴UNSW, Sydney, Australia, ⁵Wellcome Sanger Inst., London, United Kingdom, ⁶Wellcome Sanger Inst., CAMBRIDGE, United Kingdom, ⁷Wellcome Trust Sanger Inst., Cambridge, Cambridgeshire, United Kingdom

Abstract Body:

The human genome contains over 100 million SNPs, most of which are C/T (G/A) variations. The type and sequence context of these SNPs are not random, suggesting that they are caused by distinct mutational processes. Deciphering the mutational signatures is a crucial step to discovering the molecular processes responsible for DNA variations across human populations, and potentially for causing genetic diseases. Here, we used four quantitative biology approaches to deconvolute germline mutational signatures using data from diverse human populations. We report that at least four mutational processes are responsible for human genetic variations. One process is European-specific and is no longer active. The signature of this process is only found within genic regions and certain endogenous elements. The remaining three processes are currently active in all human populations. Two of the active processes co-occur and leave a single mixed mutational footprint in human nuclear DNA. The third active process is unique in that it is specific to mitochondrial DNA, and inflicts mostly C-to-T mutations at non-CpG sites and T-to-C mutations at non-TpG sites.

Evolutionary and Population Genetics Posters - Thursday

PB2844. Simple scaling laws control the genetic architectures of complex traits

Authors:

Y. Simons¹, H. Mostafavi², C. Smith¹, J. Pritchard¹, G. Sella³; ¹Stanford Univ., Stanford, CA, ²Stanford Univ., Mountain View, CA, ³Columbia Univ., New York, NY

Abstract Body:

Genome-wide association studies (GWAS) have revealed that the genetic architectures of complex traits vary widely, including in the numbers of independent signals, the magnitudes of the biggest hits, and the frequency distributions of significant signals. However, at present we lack a coherent framework for understanding the similarities and differences among traits. Here, we describe a probabilistic model that combines mutation, drift, and stabilizing selection at individual sites with a genome-scale model of phenotypic variation. We performed inference under this model for 95 diverse quantitative traits from the UK Biobank. Our estimates show that selection plays a major role in shaping trait architectures. One implication is that GWAS hits should be much younger than frequency-matched neutral SNPs; indeed, the distribution of allele ages is well-predicted by our model. Remarkably, we find surprisingly similar distributions of selection coefficients across all traits. Therefore, differences in trait architectures arise almost exclusively from just two parameters: heritability and mutational target size, which vary by 1-2 orders of magnitude across traits. When these two scale factors are accounted for, the architectures of all 95 traits are nearly identical.

Evolutionary and Population Genetics Posters - Wednesday

PB2845. Simulating effects from genetic and experimental perturbations to gene regulatory networks

Authors:

M. Aguirre¹, G. Sella², J. K. Pritchard³; ¹Stanford Univ., Stanford, CA, ²Columbia Univ, New York, NY, ³Stanford, Stanford, CA

Abstract Body:

Gene regulatory networks (GRNs) govern many of the core developmental and biological processes underlying human complex traits. Even as genome-wide transcriptomic resources have grown in scope, it remains challenging to interpret how the structure of GRNs impacts the distribution of genetic effects on gene expression. Learning the genetic architecture of gene expression traits is a key aim in quantitative genomics, but there is an unmet need to model the statistical basis of these traits in the mechanistic context of GRNs. Here, we propose a simple approach to simulate the structure and model the function of GRNs, making use of techniques from small world network theory and dynamical systems models of gene regulation. Specifically, we model gene expression regulation using a stochastic differential equation: the steady-state expression of each gene is given by a baseline transcription term and contributions from its parent nodes (transcription factors) in the GRN. This formulation permits variation in regulatory parameters due to genetic effects on gene expression (i.e., expression quantitative trait loci (eQTLs)) or intracellular interventions (e.g., targeted gene knock-down or knock-outs). We use this approach to generate synthetic samples of gene expression data, and systematically characterize the effects of gene knockouts, finding that perturbation effects flow through key regulatory pathways in the network. We further this analysis by simulating a population of individuals that harbor a substantial burden of eQTLs, investigating the distribution of gene expression heritability introduced by different factors in the network, and exploring how structural properties of the GRN affect the relative contributions of variation in cis and in trans. Finally, we conclude by discussing implications of our work towards understanding the architecture of heritable genetic variation in complex traits beyond gene expression.

Evolutionary and Population Genetics Posters - Thursday

PB2846. Simulating Genome-phenome Dataset of 1,000,000 Individuals for the European 1+ Million Genomes Initiative

Authors:

T. Hiekkalinna¹, V. Heikkinen¹, M. Perola¹, J. D. Terwilliger^{2,3,4}; ¹Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, ²Columbia Univ., New York, NY, ³Dept. of Psychiatry, Dept. of Genetics and Dev., Gertrude H. Sergievsky Ctr., New York, NY, ⁴Div. of Med. Genetics, New York State Psychiatric Inst., New York, NY

Abstract Body:

“The '1+ Million Genomes' (1+MG) initiative aims to enable secure access to genomics and the corresponding clinical data across Europe for better research, personalized health care and health policy making.” For this initiative, we are in the process of creating truly anonymous data set from mosaic of publicly available sequence data sets from European populations. This data is needed to set a founding population (original chromosomes) and parameters for the simulation of million genomes. Simulated data can be used to benchmark the national service implementations, technical standard compatibility and best practices in sensitive human data management. Purpose is to create a truly anonymized data set of million genomes in different formats that can be used and shared without fear for data security issues.

In this simulation we are using chromosome-based rapid simulation method (Terwilliger, 1993) to drop chromosomes through a population for a few hundred generations of drift, while keeping track of recombination events in every meiosis, so that at the end of the simulation we will have a set of families whose genomes will be highly recombined mosaics of the original chromosomes. The data set will have all the characteristics and data issues that any "true" genome set would.

Phenotypes will be simulated based on epidemiological models, comprised of parameters such as heritability, contribution of environmental factors, prevalence, and others which must be hypothesized (number of genes, allelic complexity, relative effect size). Based on such parameters, a genotype > phenotype relationship is simulated. This simulated genotype-phenotype data also allows testing of various genetic models, study designs and testing of statistical approaches to correctly infer the basic genetic architecture of traits of interest (important to public health). And finally, it will help us to guide appropriate sample designs, sample sizes, hypotheses, and analytic methods.

We are already simulating hundreds of thousands individuals with phenotypes and we utilize supercomputer resources of Finnish CSC IT Center for Science. This project is organized by the Finnish Institute for Health and Welfare and funded by the Ministry of Social Affairs and Health, Finland.

Evolutionary and Population Genetics Posters - Wednesday

PB2847. South Asian and Greater Middle Eastern Enrichment of *NOTCH3* p.Arg1231Cys Variant and Associated Stroke Risk

Authors:

J. Rodriguez-Flores¹, S. Khalid^{2,3}, A. Rasheed², S. A. DiGioia¹, B. Ye¹, H. Martinez¹, J. Backman¹, Regeneron Genetics Center, A. Shuldiner¹, D. Saleheen^{2,3}, M. Kapoor¹; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Ctr. for Non Communicable Disease, Karachi, Pakistan, ³Columbia Univ., New York, NY

Abstract Body:

South Asians represent less than 2% of participants in published genetic studies, despite comprising 25% of the world population. Exome sequencing of 4,882 Pakistani stroke cases and 6,094 controls identified p.Arg1231Cys in *NOTCH3* to be associated with a ~3-fold increased risk of stroke (p value 2.18e-8, odds ratio 2.97 per allele, 95% confidence interval 2.03 to 4.35). In order to better understand the population frequencies of this variant and its implications for stroke risk across South Asia (SAS), a regional survey of p.Arg1231Cys was conducted. We first performed analysis of public datasets such as gnomAD, Greater Middle East (GME) Variome, Iranome, and QChip Knowledgebase. Pooled SAS from gnomAD revealed an alternate allele frequency (AAF) of 0.00536, highest in the world. The variant was not detected in Syrian Desert nor Israeli populations. Elevated AAF was observed in Qatar (AAF = 0.0075), Northwest Africa (AAF = 0.0051), Northeast Africa (AAF = 0.0013), Turkish Peninsula (AAF = 0.0030), and Central Asia (AAF = 0.0076). Multiple Iranian populations had the variant, including Lur (AAF = 0.01), Baloch (AAF = 0.005), Kurd (AAF = 0.005), Persian (AF = 0.005), and Turkmen (AF = 0.005). Further analysis was conducted using 1000 Genomes Phase 3 (1000G) and UK Biobank (UKB) participants reporting birth in SAS or GME. The variant was observed in all 1000G SAS populations (AAF 0.00485 to 0.00521), except for Bangladesh. Within UKB, the variant was absent from participants originating in Egypt but was present in those originating from Turkey, Iran, Afghanistan, Pakistan, and India (AAF 0.0036 to 0.0056). Combined analysis of 37 thousand Pakistani exomes from the Pakistan Genome Resource (PGR), 1000G and UKB identified three major clusters of Pakistanis, and ethnic/caste data was used to label them as Baloch/Pathan (AAF = 0.0074), Punjab/Rajput (AAF = 0.006), and Ansari (AAF = 0.0056). The Punjab/Rajput clustered with 1000G Punjab, the Ansari formed a distinct cluster closer to populations from Eastern India, and Baloch/Pathan clustered with Afghans. Based on this analysis, the p.Arg1231Cys variant appears to be found across a broad geographic range with AAF > 0.003 in South Asia, Central Asia, North Africa and Turkish Peninsula. In contrast, AAF in Europe and other continents is consistently much lower (AAF < 0.0004). This result has precision medicine implications for inhabitants of these understudied populations and their diaspora.

Evolutionary and Population Genetics Posters - Thursday

PB2848. Structure-function analysis of a tRNA-derived SINE.

Authors:

E. Koch, L. E. Kirby, G. Mies, S. Emery, J. B. Moldovan, J. V. Moran, J. M. Kidd; Univ. of Michigan Med. Sch., Ann Arbor, MI

Abstract Body:

Transposable elements contribute to the phenotypic diversity present among many mammalian genomes through their ability to insert into new genomic locations. While some transposable elements, such as autonomous Long Interspersed Elements (LINEs), encode proteins required for their own mobilization (i.e., retrotransposition), other transposable elements, such as Short Interspersed Elements (SINEs), are non-autonomous and require the activity of LINE-encoded proteins to mediate their retrotransposition. In canine genomes, SINEs are present at ~1.1-1.3 million copies and are more abundant than their autonomous LINE counterparts.

Retrotransposition of the canine SINE, SINEC_Cf, is responsible for breed-specific phenotypic differences such as merle coat color, cleft palate, narcolepsy, and centronuclear myopathy. As a result, SINEC_Cf is implicated in serving an important evolutionary role in the canine genomic landscape. Unlike 7SL RNA-derived human Alu elements, SINEC_Cf is derived from a tRNA. Although the structures of tRNAs are well established, the structure and functionally important domains of SINEC_Cf RNA require elucidation. To define the relationships between SINEC_Cf structure and function, we determined the putative secondary structure of SINEC_Cf RNA using SHAPE-MaP, a chemical probing technique. The analysis of SHAPE-MaP-informed putative RNA structures, and comparisons among SINEC sequences, identified candidate secondary structural elements that may be important for SINEC_Cf retrotransposition. A sequence difference relative to the RNA polymerase III B-box promoter consensus sequence, in a region predicted to form a stem structure, was tested for retrotransposition activity in a cell-culture based assay. Our initial findings suggest that this mutation in the SINEC_Cf B box, which restores identity to the promoter consensus, results in a 35% reduction of retrotransposition activity compared to wild-type SINEC_Cf ($p=0.0004$ in a paired sample t-test). Furthermore, retrotransposition activity can be restored by a second-site compensatory mutation ($p=0.00015$, paired sample t-test compared to the B box mutant), which is predicted to restore base pairing in stem structure. In sum, our findings help clarify how variation in SINE RNA sequence and secondary structure impacts the function of SINEs and set the stage for a comparative analysis of evolutionarily diverged SINEs in other species to reveal general principles for how SINEs can evolve from tRNAs.

Evolutionary and Population Genetics Posters - Wednesday

PB2849. The differential identification of cross-species lung scRNA expression via integrative analysis

Authors:

W. Lin¹, T. Li², Y. Gong¹, Y. Chen³; ¹Tulane Univ., New Orleans, LA, ²Central South Univ., Changsha, China, ³Hunan Normal Univ., Changsha, China

Abstract Body:

The lung is an essential open organ that acts as the primary site for the respiratory process. Due to their damaging consequences, respiratory and pulmonary diseases continue to be a significant cause of morbidity and mortality on a global scale. For example, the Covid-19 virus has caused millions of deaths and widespread infection around the globe. Due to the difficulty of conducting clinical studies and acquiring human tissues, the study of in vitro cells and model organisms has considerably contributed to our understanding of molecular pathways. In the context of biomedical research laboratories, mice are the species most frequently exploited. It is probable that the evolutionary conservation of functional gene expression, which specifies unique cellular function and structure, predominates between these two species, but it is also conceivable that there may be significant genomic differences between them. Using scRNA-seq data acquired from the GEO database, our purpose was to identify genomic areas whose expression levels differ between humans and mice. The analysis of the study includes a total of 14985 human cells and 14994 mouse cells. We created an impartial atlas of cell types across species by utilizing Seurat. The single-cell weighted gene co-expression network analysis (scWGCNA) was used to identify the modules that are associated with each species. The results reveal that two of these modules have a significantly strong link with their respective species. We conducted a GO enrichment analysis on these two modules to discover the primary function of each, which reveals that these two modules are related to the activity of the Alveolar Type II Cells. These findings may help researchers to better understand the difference between mice and humans at the molecular level, which serves as the foundation of clinical advancement.

Evolutionary and Population Genetics Posters - Wednesday

PB2850. The genetic diversity of Japanese populations inferred from the whole genome sequencing data

Authors:

Y. Kawai¹, Y. Watanabe^{1,2}, Y. Omae¹, R. Miyahara^{1,3}, S-S. Khor¹, E. Noiri¹, K. Tokunaga¹, National Center Biobank Network; ¹Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, ²The Univ. of Tokyo, Tokyo, Japan, ³Natl. Inst. of Infectious Diseases, Tokyo, Japan

Abstract Body:

The Japanese archipelago is located in the east of Eurasia that is one of the endpoints of human migration since Out-of-Africa. The genetic diversity of the modern Japanese population is influenced by the admixture of hunter-gatherers (Jomon people) who migrated about 16,000 years ago and agriculturalists (Yayoi people) who migrated about 3,000 years ago. Genome studies have revealed genetic differentiation between the mainland (Hondo) and the Ryukyu Islands (e.g., Okinawa), but it is not clear when and where this genetic differentiation occurred. Furthermore, the evolutionary background of population structure within Hondo has not been clarified. In this study, we examined the genetic characteristics of the Japanese population by whole genome sequencing data. We performed a high-coverage WGS analysis of 9,850 samples from the National Center Biobank Network (<https://ncbiobank.org/en/>), which is a federation of biobanks of national hospitals in Japan, and identified more than 122 million variants. The demographic analysis revealed that the Hondo and Ryukyu populations experienced stagnation, but the time and mode are different between them. The analysis of gene genealogy from haplotype analysis identified genes related to alcohol and lipid metabolism that were affected by positive natural selection. Two genes related to alcohol metabolism were found to be 12,500 years out of phase with the time when they began to be affected by positive natural selection.

Evolutionary and Population Genetics Posters - Thursday

PB2851. The genetic making of Bangladesh

Authors:

P. Singh¹, A. Pathak², G. Sultana³, M. Metspalu², G. Chaubey¹; ¹Banaras Hindu Univ., Varanasi, India, ²Univ. of Tartu, Tartu, Estonia, ³Univ. of Dhaka, Dhaka, Bangladesh

Abstract Body:

In terms of global genetic diversity of humans, South Asia holds a central role as a major reservoir of genetic variability second only to that in Africa. However, most of the genetic information is coming from Pakistan and India, whereas Bangladesh remained largely unexplored in spite of being one of the most densely populated nations. Bangladesh is a country that is geographically surrounded on three sides by India, joins a land bridge to Southeast Asia via Myanmar, and in the South, it joins the Indian Ocean with the Bay of Bengal. The populations of Bangladesh constitute mostly Bengali speakers (Indo-Aryan), while the remaining are tribal groups affiliated with Dravidian, Austroasiatic and Tibeto-Burman language families. The number of genetic studies in this region is very limited and mainly restricted to forensic markers for a few caste and tribal populations. To bridge this gap, in the current study, we generated new genomic data from tribal and caste populations sampled across all geographic regions of Bangladesh and analyzed them to the previously published genomes to infer the genetic ancestries, both at the uniparental and biparental genome levels. About 827 individuals were analyzed for haploid analysis and 67 individuals were randomly selected for autosomal genome-wide analysis. We particularly identified the ancestry components, their admixture time, sharing, and range distribution among various caste and tribal populations. We found distinct ancestral origins of caste and tribal populations of Bangladesh. The tribal populations associate well with their linguistic affiliation. Also, we aimed to know the genetic origin of Rohingya people as they are a most debatable minority of South & South-East Asia. We sampled 25 individuals for each group viz. Rohingya, Bangladesh Muslim, and Bangladesh Hindu populations. We performed a comparative analysis with published world data. Our genome-wide analyses suggest that they belong to the South Asian stock with notable East Asian-specific ancestry. We have calculated the admixture dating to detect the admixture events which indicate very recent admixing with East Asian population ancestors. Overall, this is the first comprehensive genetic study on Bangladeshi and Rohingya populations showing their genomic landscape.

Evolutionary and Population Genetics Posters - Wednesday

PB2852. The genomic history of human populations in the North American Central and Southern Plains.

Authors:

L. Sykora¹, J. Tackney¹, B. Morris², C. Kisielinski¹, K. G. Beaty¹, L. E. Y. Norman¹, M. Adair¹, A. Rutherford³, P. Skoglund⁴, A. Reynolds², J. Raff¹; ¹Univ. of Kansas, Lawrence, KS, ²Baylor Univ., Waco, TX, ³Univ. Coll. London, London, United Kingdom, ⁴The Francis Crick Inst., London, United Kingdom

Abstract Body:

The Great Plains of North America have been continuously occupied by humans for over 13,000 years. Although there is extensive archaeological evidence for a dynamic history of adaptation, migration, and sociopolitical interactions among the many populations who have inhabited the area prior to European Contact, our understanding of the biological histories of these peoples is severely limited by a lack of ancient human genomes.

In order to explore the population histories in this region, we performed genome-wide targeted SNP captures on ancient individuals residing in the Plains between 1100-1700 CE (850-250 calibrated years before present, YBP). These individuals lived at several sites in the present-day states of Kansas and Texas, and date to periods spanning key social, environmental, and political changes. This project includes the origin of the descendant tribe, who gave us permission for this research and participates as a collaborating partner.

These Ancient Plains individuals share more genetic drift with present day members of the “Southern Native American” (SNA) lineage, particularly with the Mixe and Pima, than with members of the “Northern Native American” (NNA) lineage. Intriguingly, they also share genetic drift with an ancient member of the NNA lineage, suggesting that there was greater complexity in past population dynamics than are currently accounted for in existing models. Finally, we identified a second-degree kin relationship between two contemporaneous individuals from sites in Kansas and Texas, suggesting long-distance interactions between populations in the Central and Southern Plains, during this time period. This project will highlight the importance of collaborative research with tribal partners in documenting genetic variation in past populations and generating new models of population histories in the Great Plains.

Evolutionary and Population Genetics Posters - Thursday

PB2853. The impact of background selection on deleterious alleles and its consequences on trait evolution and DFE inference

Authors:

X. Li¹, J. Berg¹, J. Novembre²; ¹Univ. of Chicago, Chicago, IL, ²Univ Chicago, Chicago, IL

Abstract Body:

In humans, background selection has been estimated to be a dominant factor explaining neutral variation. It also distorts the neutral site frequency spectrum by reducing variation and leaving an excess of rare variants. However, the impact of background selection on deleterious alleles and its downstream effects on quantitative traits are not well studied.

First, we studied the site frequency spectrum of focal deleterious alleles with different selection coefficients under background selection. We found that for weakly deleterious alleles, background selection reduces the genetic diversity of and leaves an excess of rare variants, similar to its impact on neutral alleles. However, for strongly deleterious alleles, because they are short-lived, background selection has minimal effect.

Second, we investigated how background selection impacts the evolution of complex traits using theoretical models and simulations. We find that background selection shifts the genetic architecture contributing to phenotypic variance toward rare alleles. This agrees with the result in our first section, in that background selection removes weakly selected, more common variants, and has less effect on strongly selected, rare variants. In addition, background selection reduces heritability for traits with a substantial contribution from weakly selected variants. Furthermore, by including background selection in a disease trait model (the liability threshold model), we show that the reduction in diversity caused by background selection leads to a corresponding increase in disease prevalence in order to maintain mutation-selection balance. Using plausible parameter values that generate a diversity reduction of 20% due to background selection (as estimated in humans), we find background selection leads to an increase in disease prevalence by 20%.

Third, while a site frequency spectrum is widely used to infer a distribution of fitness effects (DFE) at functional sites, the distortion of a site frequency spectrum under background selection could potentially bias DFE estimates. We extensively tested the performance of DFE inference under background selection with the 'moments' software (Jouganous et al., 2017). We found that by using synonymous sites with a misspecified demographic model to control for background selection, one can robustly estimate DFE under different demographic histories.

More generally, our work addresses the role of background selection as a factor impacting phenotypic variation, including disease prevalence, and informs best practices for DFE inference from population genetic data.

Evolutionary and Population Genetics Posters - Thursday

PB2854. The influence of demographic history and genetic architecture on complex phenotypes via runs of homozygosity

Authors:

Z. Szpiech; Pennsylvania State Univ., University Park, PA

Abstract Body:

Runs of homozygosity are long stretches of identical-by-descent (IBD) haplotypes inherited from parents with a recent common ancestor. Their abundance and distribution within a population are affected by numerous factors such as population bottlenecks and isolation, founder effects, recent inbreeding, and natural selection, and it has been shown that long ROH are enriched for deleterious homozygotes. Although ROH abundance has been associated with increased risk for various complex traits, it remains unclear the extent to which population history and genetic architecture influence ROH associations with phenotypes. Here we take a simulation approach to characterize the relationship between demographic history, genetic architecture, and a generic complex phenotype. We perform forward-in-time simulations of a three-population human demographic history roughly representing African, European, and Asian continental populations. We simulate a 100 Mb chromosome region with exon structure and a variable recombination map based on the first 100 Mb of human chromosome 1 and allow deleterious mutations with selection coefficients drawn from a gamma distribution. Phenotype is modeled as a function of selection coefficients, with parameters that allow us to vary the relative importance of rare versus common variants in their contribution to the total phenotype. Our results show that demographic history, ROH length, and dominance coefficient are important factors contributing to how much variation in a trait are explained by ROH. We show that ROH influence the simulated phenotype most strongly when alleles are recessive and rare, and that, under these conditions, long ROH (comprised of IBD haplotypes inherited from a very recent ancestor) have more influence than short ROH. These results suggest that the role of ROH in contributing to complex phenotypes may be largely due to the pairing of rare alleles of recessive effect and that incorporating ROH into disease mapping approaches may help the identification of recessive effects.

Evolutionary and Population Genetics Posters - Wednesday

PB2855. The landscape of tolerated genetic variation in humans and primates

Authors:

H. Gao¹, T. Hamp², J. Ed², J. Schraiber¹, J. McRae³, M. Berk⁴, Y. Yang¹, A. Dietrich¹, P. Fizev⁵, P. Fizev¹, L. Kuderna¹, L. Sundaram⁶, Y. Wu¹, A. Adhikari¹, Y. Field¹, C. Chen¹, S. Batzoglou⁷, F. Aguet¹, Primate Sequencing Consortium¹¹, M. Lek⁸, S. Sunyaev⁹, A. O'Donnell-Luria⁴, H. Rehm¹⁰, J. Xu¹¹, J. Xu¹, J. Rogers¹², T. Marques-Bonet¹³, K-H. Farh¹; ¹Illumina, inc., Foster City, CA, ²Illumina, inc., Cambridge, United Kingdom, ³Illumina, Inc, Foster City, CA, ⁴Broad Inst. of MIT and Harvard, Boston, MA, ⁵San Jose, CA, ⁶Illumina, San Diego, CA, ⁷Stanford Univ, Stanford, CA, ⁸Yale Univ. Sch. of Med., New Haven, CT, ⁹Harvard Med. Sch./Brigham & Women s Hosp., Boston, MA, ¹⁰Massachusetts Gen. Hosp., Boston, MA, ^{11,12}Baylor Coll. Of Med., Houston, TX, ¹³Inst. de Biologia Evolutiva (UPF-CSIC), Barcelona, Barcelona, Spain

Abstract Body:

Personalized genome sequencing has revealed millions of genetic differences between individuals, but our understanding of their clinical significance remains largely incomplete. To systematically decipher the effects of human genetic variants, we obtained whole genome sequencing data for 809 individuals from 233 primate species, and identified 4.3 million common protein-altering variants with orthologs in human. We show that these variants can be inferred to have non-deleterious effects in human based on their presence at high allele frequencies in other primate populations. We use this resource to classify 6% of all possible human protein-altering variants as likely benign and impute the pathogenicity of the remaining 94% of variants with deep learning, achieving state-of-the-art accuracy for diagnosing pathogenic variants in patients with genetic diseases.

Evolutionary and Population Genetics Posters - Thursday

PB2856*. The lingering effects of Neanderthal introgression on human complex traits

Authors:

X. Wei¹, C. Robles², A. Pazokitoroudi³, A. Ganna⁴, A. Gusev⁵, A. Durvasula², S. Gazal⁶, P-R. Loh⁷, D. Reich⁸, S. Sankararaman²; ¹Cornell Univ., Ithaca, NY, ²UCLA, Los Angeles, CA, ³UCLA, LA, CA, ⁴Inst. for Molecular Med. Finland, Helsinki, Finland, ⁵Dana-Farber Cancer Inst., Boston, MA, ⁶USC, Los Angeles, CA, ⁷Brigham and Women's Hosp. / Harvard Med. Sch., Boston, MA, ⁸Harvard Univ, Boston, MA

Abstract Body:

The mutations introduced into the ancestors of modern humans from interbreeding with Neanderthals have been suggested to contribute an unexpected extent to complex human traits. However, testing this hypothesis has been challenging due to the idiosyncratic population genetic properties of introgressed mutations. We developed rigorous methods to assess the contribution of introgressed Neanderthal mutations to heritable trait variation relative to that of modern human variants. We applied these methods to analyze 235,592 introgressed Neanderthal mutations and 96 distinct phenotypes measured in about 300,000 unrelated white British individuals in the UK Biobank. Introgressed Neanderthal mutations have a significant contribution to trait variation consistent with the polygenic architecture of complex phenotypes (contributing 0.1% of heritable variation averaged across phenotypes; $p = 9.59 \times 10^{-9}$). However, the contribution of introgressed mutations tends to be significantly depleted relative to modern human mutations matched for allele frequency and linkage disequilibrium (about 57% depletion on average), consistent with purifying selection on introgressed mutations. Different from previous studies (McArthur 2021), we find no evidence for elevated heritability across the phenotypes examined. We identified 348 independent significant associations of introgressed Neanderthal mutations with 64 phenotypes ($p < 1 \times 10^{-10}$). Previous work (Skov 2021) has suggested that a majority of such associations are likely driven by statistical association with nearby modern human variants that are the true causal variants. We therefore developed a customized statistical fine-mapping methodology for introgressed mutations that led us to identify 112 regions (at a false discovery proportion of 16%) across 47 phenotypes containing 4,303 unique genetic variants where introgressed mutations are highly likely to have a phenotypic effect. Examination of these mutations reveal their substantial impact on genes that are important for the immune system, development, and metabolism. Our results provide the first rigorous basis for understanding how Neanderthal introgression modulates complex trait variation in present-day humans.

Evolutionary and Population Genetics Posters - Wednesday

PB2857. The NHGRI Sample Repository for Human Genetic Research: biospecimens and a new genomic data search tool

Authors:

A. Liautaud, D. Kusic, G. Calendo, J. Madzo, B. Kerr, S. Sampson, L. Willis, E. Horton, M. Mitchell, N. Gharani, L. Scheinfeldt; Coriell Inst. for Med. Res., Camden, NJ

Abstract Body:

The NHGRI Sample Repository for Human Genetic Research (NHGRI Repository) housed at the Coriell Institute for Medical Research facilitates studies of human genetic and genomic variation by establishing, characterizing, and distributing a large (over 3,700) and diverse publicly available collection of renewable biospecimens obtained from human populations living around the world, including biospecimens associated with the 1000 Genomes Project. Participants that generously donated to the NHGRI Repository consented to their biospecimens being used for a wide range of general research and to broad sharing of largescale, associated genomic data collected from their biospecimens. Through the 1000 Genomes Project, the majority of this collection has been characterized with publicly available whole genome sequencing data and other large-scale genomic data. We recently developed a user-friendly and integrated genomic data search tool that allows catalog users to dynamically query 1000 Genomes Project whole genome sequencing and RNA-Seq data. The whole genome sequencing data can be searched by individual SNP (rsid), or by gene (HUGO symbol), and the RNA-Seq data can be searched by gene. The SNP search returns an interactive table of each subject, subject affiliation, genotype, and sex that can be filtered or exported to an excel file. The Gene search returns an interactive table of each variant identified in the dataset, chromosome location, and human reference genome build GRCh37 position, alleles, and functional annotation that can be filtered or exported to an excel file. The Gene Expression search returns an interactive table that can be viewed or downloaded (to a .CSV file) with gene expression levels (in counts per million (CPM), fragments per kilobase of exon model per million reads mapped (FPKM), reads per kilobase of transcript per million mapped reads (RPKM), transcripts per million (TMP), or by a variance stabilizing transformation (VST) of the data), and also including the chromosome and chromosomal start and end positions of the gene, and the type of gene (e.g., protein coding). Researchers can view a histogram plot of gene expression levels, and mark where any given sample expression levels fall within the distribution. These tools are intended to enable researchers a fast and simple way to identify biospecimens with the most appropriate genetic variation and gene expression profiles for their research goals. Our Genomic Data Search tools can be accessed at <https://www.coriell.org/1/Browse/Genomic-Data-Search>, and more information about the NHGRI Repository can be found at <https://catalog.coriell.org/NHGRI>.

Evolutionary and Population Genetics Posters - Thursday

PB2858. The spatial distribution of rare deleterious alleles: implications for study design in human genetics

Authors:

M. Steiner¹, D. Rice¹, C. Porras^{2,1}, J. Novembre¹; ¹Univ. of Chicago, Chicago, IL, ²Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Sampling design is a key question in the construction of large-scale sequencing cohorts, which are a priority for human genetics due to the potential of downstream studies to uncover rare, large-effect variants. The geographic component of sampling design can be viewed on a spectrum between two extremes: either sampling many individuals from a single location or fewer individuals from each of multiple locations. The consequences of such "narrow" vs. "broad" sampling for detecting rare variants - which tend to be geographically localized - are largely unknown. To this end, we develop a population genetic model for the distribution of deleterious alleles in a spatially structured population - accounting for dispersal, drift, selection, and mutation simultaneously - while incorporating spatially-variable sampling. Using techniques from the theory of superprocesses, we compute the expected site frequency spectrum (SFS) of a sample as a function of the population genetic parameters and sampling scheme. We then use the SFS to estimate the power to detect and characterize rare variants. In doing so, we identify trade-offs between narrow and broad sampling. Analysis of our model suggests that, as a result of the low population frequency and spatial clustering of rare, deleterious variants, deep sequencing of one or a few populations can be a more efficient means to find multiple carriers of a rare variant, as is needed to understand phenotypic effects. Yet, we find that in the case of association studies or individual risk assessment, sampling individuals from only a few geographic regions limits transferability across populations, with potential consequences for the deployment of precision medicine techniques such as polygenic risk scores. Our approach allows us to quantify these trade-offs and thus we expect the results will inform existing discourse regarding optimal strategies for designing sequencing cohorts. Moreover, improving the ascertainment and characterization of deleterious alleles has the potential to unveil novel disease mechanisms and drug targets, refine precision medicine approaches, and improve understanding of genetic architectures of complex traits.

Evolutionary and Population Genetics Posters - Wednesday

PB2859. Tracing the genetic link of Mizo people of India: Genetic & archaeological evidence

Authors:

B. Bankura^{1,2}, P. Singh³, G. Chaubey³, M. Das²; ¹Med. Coll., Kolkata, Kolkata, India, ²Univ. of Calcutta, Kolkata, India, ³Banaras Hindu Univ., Varanasi, India

Abstract Body:

The Himalayan region played a crucial role in the human dispersal and colonization of South Asia. It forms a unique and narrow passage that connects the Indian sub-continent to East and South East Asia. Archaeological and anthropological studies suggest migration and cultural diffusion in this region with the Tibetan plateau in the north. Mizoram is the southernmost state in the northeast region, sharing its borders with Tripura, Assam, and Manipur, along with Myanmar and Bangladesh. The Mizo people are believed to be a part of the Mongoloid feature and Mizoram has the highest number of tribal people among all states of India. Therefore, our study is an attempt to clarify the origins of the Mizo people and their migration roads to their present settlements. Thus, issues of ethnicity and migration are complex but highly relevant to the modern state. A more concentrated effort is required to track these population movements and dispersals linking archaeology and genetics. Plausible migration histories of these groups have been addressed based on linguistic, ethnographical, historical, and folkloristic data however; less attention has been given to genetics or modern genome mapping. Given the vastness of the region and considerable diversity in the topography, a specific region has been selected for the study: the region adjoining the Indo-Myanmar border in the Champhai district of Mizoram. Therefore, our study is an attempt to clarify the origins of the Mizo people and their migration roads to their present settlements. In the present study, we have investigated 110 individuals belonging to different clans of Mizo people for a genome-wide autosomal marker study. Our results suggest that they are a unique and isolated tribal group and have a very distinct population history. Haplotype and allelic frequency-based analyses indicate that they substantially shared ancestry primarily with East Asian populations and Tibeto-Burman speakers of India but also differentiated from them and became a distinct population with time.

Evolutionary and Population Genetics Posters - Thursday

PB2860. Understanding ancestry-specific disease allelic effect sizes by leveraging multi-ancestry-matched single-cell RNA-seq and GWAS datasets

Authors:

J. Wang, S. Gazal; Univ. of Southern California, Los Angeles, CA

Abstract Body:

Genome-wide association studies (GWAS) have highlighted that human diseases and complex traits are dominated by common regulatory risk variants. However, some of these variants have ancestry-specific effects, as suggested by the poor transportability of polygenic risk scores across populations. One explanation would be that regulatory differences across ancestries (due to different genetics and/or environments) lead to ancestry-specific disease effect sizes. Here we leveraged multi-ancestry-matched single-cell RNA-seq (scRNA-seq) and GWAS datasets to detect genes differentially expressed across ancestries (ANC-DE genes) at the cell-type level and to test if genetic variants around those genes tend to be enriched in GWAS ancestry-specific effects. First, we analyzed scRNA-seq data in peripheral blood mononuclear cells (PBMCs) from 45 individuals of East-Asian (EAS) or European (EUR) ancestry (Perez et al. 2022 Science). We detected 320 genes significantly differentially expressed (FDR 5%) across EAS and EUR ancestries in at least one cell type. The majority of these genes (85%) tend to be differentially expressed in a single cell type, and were enriched in genes involved in the immune response to the environment. Second, we leveraged 15 GWAS datasets of blood and immune-related traits available in both EAS (average N = 53K) and EUR populations (average N = 55K), and ran S-LDXR (Shi et al. 2021 Nat Com) to test if genetic variants surrounding ANC-DE genes were enriched in ancestry-specific GWAS effects. We determined that the squared multi-ancestry genetic correlation is 0.66 ± 0.05 for SNPs surrounding PBMCs ANC-DE genes, which represents the lowest existing correlation obtained by S-LDXR (vs 0.94 ± 0.01 for SNPs surrounding any gene). We observed a (non-significant) lower depletion around lymphoid ANC-DE genes (0.59 ± 0.06) compared to myeloid ANC-DE genes (0.65 ± 0.06); these depletions were driven by B cells ANC-DE genes (0.28 ± 0.06) and by conventional dendritic cells ANC-DE genes (0.26 ± 0.09) in lymphoid and myeloid cells, respectively. Finally, we leveraged our findings to interpret different discordant results between EAS and EUR Lymphocyte Count GWAS around the MCL1 gene. Altogether, our results highlight that genes with ancestry-specific levels of regulation in PBMCs are enriched in genes involved in the immune response to the environment and enriched in genetic variants with different effect sizes in blood and immune-related traits. These results suggest that ancestry-specific regulation leads to ancestry-specific genetic effect sizes and disease risk.

Evolutionary and Population Genetics Posters - Wednesday

PB2861*. Understanding natural selection in the European Holocene using Ancient DNA.

Authors:

D. Pandey^{1,2}, M. Harris³, N. R. Garud³, V. Narasimhan¹; ¹The Univ. of Texas at Austin, Austin, TX, ²Indian Inst. of Technology Kharagpur, Kharagpur, India, ³Univ. of California, Los Angeles, CA

Abstract Body:

Ancient DNA (aDNA) has been a revolutionary technology in understanding human history but has not been used extensively to study natural selection as large sample sizes to study allele frequency changes over time have thus far not been available. Previous methods to study selection using aDNA have been limited to examining the most recent time period or looking at specific alleles already known to be under selection. Here, we examined a time transect of 885 published samples over the past 12,000 years using haplotype homozygosity based selection scans that do not require additional demographic models to uncover selective events. We used f_4 -statistic, date of the sample, archeological context, and geographic location to group the data into five time periods: the Mesolithic (M), Neolithic (EN), Bronze Age (BA), Iron Age (IA), and Historical period (H), with sample sizes of 177 individuals each. As aDNA data is affected by high missingness, ascertainment bias, DNA damage, random allele calling, and is unphased, we first validated our selection scan (a modified version of the G12 statistic) on simulated data resembling aDNA under a demographic model that captures broad features of the allele frequency spectrum of European genomes. Our scan was sensitive enough to detect events that occurred at different time depths, selection intensities, and modes of selection. By utilizing samples from the H period, which are 1000 years older than modern samples, we also validated our scan on several positive controls that have been previously identified and functionally validated in modern European datasets. We then applied our statistic to real data and used our time transect to examine the timing of selection on well-studied loci (*LCT*, *SLC22A5*, and *SLC24A5*) and show that we detect selection on *LCT* in the H period and for *SLC24A5* in the BA, while *SLC22A5* appears to be under selection in the EN. We also found that functional categories of SNPs showed a significantly higher signal than SNPs that were not annotated as being functionally relevant, with the HLA region the most elevated. In the earliest time period studied (M), we obtained an elevated signal for loci that had previously not been identified by any selection scans in modern European data, including in the gene *ALDH*, which is the principal catalyst for the oxidation of acetaldehyde during alcohol metabolism as well as in *AOAH* which prevents prolonged and damaging inflammatory responses from bacteria. Our results suggest that applying selection scans to aDNA uncovers putative selection signals at loci in the ancient past that might have been masked in modern samples due to population structure, admixture, and genetic drift.

Evolutionary and Population Genetics Posters - Thursday

PB2862. Understanding the structure of regional populations with founder effects.

Authors:

L. Gagnon, C. Moreau, A. Girard, J. Bouchard, C. Laprise, H. Vézina, S. Girard; Univ. of Quebec at Chicoutimi, Chicoutimi, QC, Canada

Abstract Body:

Background/objectives: Quebec is an example of a young and non-uniform founder population. By combining genetic and genealogical data from different regions in Quebec, it's possible to study in detail the regional subdivisions and better understand the associated founder effects. **Methods:** The genetic and genealogical data are available for 666 individuals from 8 Quebec regions. We have characterized the identity-by-descent (IBD) sharing, the genealogical kinship, the number of most recent common ancestors (MRCA), the inbreeding of MRCA and the ancestors' occurrences per generation using bioinformatic tools such as refinedIBD and genlib R package. **Results:** The IBD length distribution analysis reveals more numerous and longer segments in individuals from Saguenay and Acadians from Gaspésie. However, genealogical information, such as kinship, MRCA and ancestors' occurrences, indicates that despite the similar IBD distribution, the demographic and historical phenomena behind are specific to each subpopulation. Last, using the genealogical kinship of ancestors, it is possible to follow the evolution of the population structure at each generation from the beginning of the colony until today. **Conclusion:** These results demonstrate a fine and specific regional genetic structure in Quebec subpopulations. Notably, Saguenans and Acadians from Gaspésie have different historical particularities that lead to similar patterns of genetic sharing: a strong founder effect for Saguenay and high levels of kinship among Acadians' ancestors. This difference is however observable within the genealogical data. Finally, this approach integrating population genetics and genealogical structure will be used for studying complex diseases, such as asthma, epilepsy, and schizophrenia, in future projects.

Evolutionary and Population Genetics Posters - Wednesday

PB2863. Unsupervised Learning of Ancestry Informative Markers and Genetic Admixture Coefficients From Biobank Data

Authors:

S. Ko, E. Sobel, H. Zhou, K. Lange; Univ. of California, Los Angeles, Los Angeles, CA

Abstract Body:

Admixture estimation plays a crucial role in ancestry inference and genomewide association studies (GWAS). Computer programs such as ADMIXTURE and STRUCTURE are commonly employed to learn the admixture proportions of sample individuals. However, these programs can be overwhelmed by the computational burdens imposed by the 10^5 to 10^6 samples and millions of markers commonly found in modern biobanks. An attractive alternate strategy is to run these programs on a small set of ancestry informative SNP markers (AIMs) that exhibit substantially different frequencies across populations. Unfortunately, existing methods for identifying AIMs require knowing ethnicity labels for a subset of the sample. This supervised learning approach creates a chicken or the egg scenario. Here, we present an unsupervised, scalable framework that carries out AIM selection and estimation of admixture coefficients seamlessly. Our simulated and real data examples show that this approach is scalable to biobank size. Our implementation of this method is called OpenADMIXTURE. For the first part of our pipeline, AIM selection, we use the SKFR (sparse K-means with a feature ranking) method. Our procedure is highly parallelized and effective in AIM selection. SKFR's unsupervised clustering is insensitive to a small fraction of labeling errors and admixed samples. SKFR also delivers an explicit ranking of AIMs. Our experiments suggest that 100,000 informative SNPs deliver better clusters than full SNP sets. Uninformative SNPs simply constitute noise that slows clustering and obscures subpopulations. For the second part of our pipeline, admixture estimation, we created an open-source re-implementation of ADMIXTURE in the Julia programming language. The original paper announcing ADMIXTURE has garnered over 5000 citations. Our implementation in OpenADMIXTURE is 8 times faster than the original ADMIXTURE on CPUs with multithreading, and 35 times faster on an Nvidia V100 GPU. Both our results and our total computation time is comparable to SCOPE, the only previous admixture method known to be scalable to biobank data. However, our method results in a ranking of AIMs and is based on a likelihood model that incorporates basic population genetics concepts. Both SCOPE and OpenADMIXTURE can estimate admixture coefficients for a biobank data set with 500,000 subjects and 600,000 SNPs in about 2 hours with a 64-thread CPU and a V100 GPU. However, the peak memory demands of OpenADMIXTURE is less than 30% that of SCOPE; for this data set, under 75 GB of RAM versus 250 GB. OpenADMIXTURE's efficiency is partly due to its systematic exploitation of Plink's binary format for both storage and computation.

Evolutionary and Population Genetics Posters - Thursday

PB2864. Using a population-specific reference panel improves genotype imputation accuracy in individuals of African ancestry

Authors:

R. Mayanja^{1,2}, A. B. Kamiza³, O. Soremekun¹, K. Hatzikotoulas⁴, W. N. Rayner⁴, A. P. Morris⁵, M. Crampin³, T. Chikowore⁶, S. Fatumo¹; ¹Uganda Virus Res. Inst./Med. Res. Council/London School of Hygiene and Tropical Medicine, Entebbe, Uganda, ²Makerere Univ., Kampala, Uganda, ³Malawi Epidemiology and Intervention Res. Unit, Lilongwe, Malawi, ⁴Inst. of Translational Genomics, München, Germany, ⁵Ctr. for Genetics and Genomics Versus Arthritis, Ctr. for Musculoskeletal Res., Div. of Musculoskeletal and Dermatological Sci., The Univ. of Manchester, Manchester, United Kingdom, ⁶Sydney Brenner Inst. for Molecular BioSci., Faculty of Health Sci., Univ. of the Witwatersrand, Johannesburg, South Africa

Abstract Body:

Genotype imputation uses densely genotyped haplotypes to predict genotypes of untyped variants in a target dataset. It boosts the power of genome-wide association studies (GWAS) and can be used for meta-analysis, fine mapping, and developing genetic risk scores. Factors such as the reference panel's representation of the target population influence the quality of imputation. We investigated the best performing imputation reference panel in people of African ancestry by comparing the performance of the five most commonly used imputation reference panels containing African ancestry data (Trans omics for Precision Medicine (TopMed), Haplotype Reference Consortium (HRC), 1000 Genomes Project (1000G), Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA), and African Genome Resource (AGR)). Data from 288 Malawi Epidemiology Intervention Research Unit (MEIRU) cohort members who were genotyped on the H3A genotyping array were used in this study. The total number of SNPs imputed with an imputation quality > 0.3 and mean imputation quality scores across a range of minor allele frequency (MAF) bins were used to evaluate imputation performance. TopMed and AGR reference panels imputed the most SNPs (73,553,040 and 26,131,641, respectively), while HRC reference panel imputed the fewest (17,998,233). TopMed and AGR had similar mean r^2 values and outperformed all other reference panels with $MAF > 0.2$. However, AGR had a higher mean r^2 than all other reference panels for SNPs with $MAF > 0.2$. In conclusion, when performing genotype imputation on people of African ancestry, AGR and TopMed should be prioritized to improve imputation coverage and accuracy.

Evolutionary and Population Genetics Posters - Wednesday

PB2865. Variable representation of duplicated sequence in recent canine genome assemblies.

Authors:

A. Nguyen¹, J. Kidd²; ¹Univ. of Michigan, Ann Arbor, MI, ²Univ MICHIGAN, Ann Arbor, MI

Abstract Body:

Gene duplications arise through two means: the reverse transcription and integration of RNA transcripts, or the duplication of DNA segments. DNA-level duplications take the form of either whole genome duplications, or smaller and restricted segmental duplications, which can create additional copies of genes or gene clusters. When these additional copies differ in number between individuals or populations, they are referred to as copy number variants (CNVs), which can lead to the evolution of novel gene functions as well as alteration of normal gene expression patterns. Over time, changes in CNVs can result in the evolution of gene families and are associated with a range of adaptive and detrimental phenotypes. However, despite the importance of duplication, systematically identifying and analyzing these variants has been challenging. The recent explosion of new canine long-read genome assemblies allows for a finer investigation of duplicated sequences in the canine model than ever before. However, the extent to which duplicated sequences are accurately represented in these new assemblies remains unclear. By conducting genome assembly self-alignment using the computational program BISER, and read depth analysis using fastCN and QuicK-mer2, we have assessed the representation of genomic sequence in nine published, long-read, high-quality, reference-level canine genome assemblies. Our analysis has three major conclusions. The first is that recently published genome assemblies, including canFam4, the current canine reference genome, include several notably-misassembled areas. The second, a subset of genes appear to be proliferating through the genome at extremely high levels in some assemblies, and based on exon-exon coverage, these gene duplicates appear to be retrogenes. Finally, certain differences between assemblies are representative of polymorphic duplications that are segregating among canines.

Omics Technologies Posters - Wednesday

PB2866. A bioinformatics pipeline to aid in the design of allele-specific antisense oligonucleotides for patients with rare genetic conditions

Authors:

Q. Zhang¹, T. Huang², T. Cole³, T. Truong², K. Cung¹, R. Wang¹, A. Crawford¹, R. Taft¹; ¹Illumina Inc, San Diego, CA, ²Illumina Inc, Foster City, CA, ³n-Lorem Fndn., San Diego, CA

Abstract Body:

Antisense oligonucleotides (ASO) are short, synthetic oligodeoxynucleotides which are able to modify gene function; they can be developed rapidly, while being highly specific and relatively inexpensive. ASOs are a promising treatment for many rare genetic diseases, especially those caused by dominant gain-of-function mutations, where effective allele-specific ASO design requires discriminating the parental haplotypes to target the corresponding mutant pre-mRNA. However, a standard short-read sequencing approach often fails to confidently phase the two haplotypes, especially at highly homozygous or low-complexity regions. Illumina's Infinity technology, which expands read length to up to 10kb, helps to resolve phasing in the remaining cases. We developed a computational pipeline to aid ASO design with Infinity alignments, while leveraging the highly accurate call set from short-read sequencing. The haplotype with the known de novo pathogenic variant was used as a target to design gene silencing ASO candidates; while the other haplotype, together with haplotypes from functionally associated paralogs, were used as a background database to remove ASOs that have high affinity to them. Next, the specific ASO candidate sequences were compared against a patient-specific genomic or transcriptomic database to further increase specificity. Finally, in-vitro and in-vivo functional screening will be carried out to evaluate the efficacy and specificity of proposed ASOs. Therapy development for rare disease is often hindered by the small number of patients. This work showcased a clinical application of infinity long reads to bridge variant identification with personalized medicine.

Omics Technologies Posters - Thursday

PB2867. A Cloud-scalable Software Suite for Large-Scale Proteogenomics Data Analysis and Visualization

Authors:

A. Gajadhar, H. Auluck, M. K. R. Donovan, K. Motamedchaboki, Y. Lou, T. Platt, S. Batzoglou; Seer, Redwood City, CA

Abstract Body:

Assessment of the flow of genetic information through multi-omic data integration can reveal the molecular consequences of genetic variation underlying human disease. Next-generation sequencing (NGS) can be used to identify genetic variants, while mass spectrometry can be used to assess the proteome. Integration of proteomics and genomics data requires many tools of which require complex workflows that can act as a barrier for researchers to adapt new analysis tools. Proteograph™ Analysis Suite (PAS) software application is a dedicated, cloud-based software solution removes barriers for proteomics researchers by enabling processing, analyzing, and visualizing proteomics data sets generated by liquid chromatography-mass spectrometry (LC-MS). PAS includes an experiment data management system, analysis protocols, analysis setup wizard, and result visualizations. PAS can support both Data Independent Analysis (DIA) and Data Dependent Analysis (DDA) workflows. PAS includes visualizations to assess data quality, as well as enables biological insights through differential protein abundance analyses. Further, we present a new, proteogenomics integration to PAS that enables the identification and exploration of peptides with protein sequence altering genetic variants through the integration of Proteograph proteomics data with NGS variant information. This proteogenomic analysis first enables users to build a customized protein sequence database, which involves uploading a custom or sample-specific variant call file (VCF) to identify genetic variants that may result in single amino acid variants (i.e., peptide variants) not captured in the reference human proteome. Custom peptide variants are then combined with reference human proteome to generate custom peptide sequence databases. Using the customized protein sequence database, DDA data is searched for peptide variants using an MS/MS search engine. Results are then accessible through a peptide variant browser and proteogenomic data explorer. The peptide variant browser summarizes results in an interactive table and plots, which enables users to explore the identified peptides and their intensities variants across samples. The Proteogenomic data explorer is an interactive browser to examine how Proteograph peptide and variant peptide data maps in genomic space (at protein- and at nucleic acid/amino acid- scale resolution). From data to insight, PAS provides an easy-to-use and efficient suite of tools to enable proteogenomic data analysis.

Omics Technologies Posters - Wednesday

PB2868. A Comparative Bioinformatics Analysis to Propose a COVID_19 Candidate Vaccine Epitope at the ACE-Bioinformatics in Mali (West Africa)

Authors:

M. Kouyate¹, M. HAIDARA², B. Sacko³, M. Sangare⁴, S. Doumbia⁵, M. DIAKITE⁶, M. Wele¹; ¹ACE-USTTB, street 171/ door 85, Mali, ²Univ. of Sci., of Techniques and technologies of Bamako, Bamako, Mali, ³Univ. of Pennsylvania Sch. of Med., Pennsylvania, PA, ⁴Univ. of Sci., Techniques and Technologies of Bamako (USTTB), Bamako, Mali, ⁵USTTB, Bamako, Mali, ⁶Malaria Res. and Training Ctr., Bamako, Bamako, Mali

Abstract Body:

Coronaviruses are host-specific unless a mutation occurs in the viral genome. Human coronavirus 229E (HCoV-229E), Middle East respiratory syndrome (MERS-CoV), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are three of six human coronaviruses that are ubiquitous and could persist on inanimate surface. COVID barrier measures and drug treatment are useful, however effective vaccinations are key to stopping the propagation of COVID-19. Our aim was to identify COVID-19 vaccine candidate epitopes using bioinformatics approaches. We obtained SARS-CoV-2 sequence from 2019nCoV, NCBI, GISAID and TAURAU/T-bio-info server. We reviewed the viral origin and compared the sequencing data to identify the consensus motive “**KRSFIEDLLFNKVTLADAGF**” (a target binding site) in the spike domain. This motif has unanimously been found in both newly identified and previous 2019-nCoV strains, which makes it a potential good candidate epitope. The receptor binding domain (RBD) in spike proteins are responsible for the contact between the virus and its host. We found that highly conserved -with 70 % identity- residues in the receptor-interacting motive , five important amino acids (L455, Y473, N479, F486, and Q493) on the RBD in spike proteins and D480, and T487 in the SARS-CoV_ RBD that allow interspecies infection. In SARS-CoV-2, a slight modification of some of these residues could improve the interaction with the human cellular receptor. In SARS-CoV, two main residues N479 and T487, which correspond to Q493 and N501 in SARS-CoV-2 have been associated with the recognition of the human angiotensin converting enzyme 2 (ACE2) receptor. These residues drive energetically favorable changes for a strong interaction with the human receptor. The motif **KRSFIEDLLFNKVTLADAGF** sat on a region around one of the known cleavage sites of the SARS virus, which are believed to be required for virus activation for cell entry. This sequence motif and surrounding variations could be further investigated as a specific synthetic COVID-19 vaccine epitope. Nevertheless, the epitope we are proposing is presumably a good target for a good COVID-19 vaccine candidate in clinical trials. Although new research data are emerging on COVID-19 at an explosive rhythm, our work can easily be replicated by others.

Omics Technologies Posters - Thursday

PB2869*. A comparison of the sensitivity and diagnostic rate of long-read vs. short-read whole genome sequencing in critically ill infants in the SeqFirst Project

Authors:

D. Miller¹, M. Galey¹, T. Wenger², M. Sikes³, A. Scott², C. Davis¹, K. Buckingham¹, J. Chong¹, D. Veenstra⁴, K. Dipple¹, D. Copenheaver⁵, J. Juusola⁶, K. Retterer⁶, M. Kirsty⁶, A. Snook⁶, P. Kruszka⁶, E. Eichler⁴, M. Bamshad¹; ¹Univ. of Washington, Seattle, WA, ²Seattle Children's Hosp., Seattle, WA, ³Seattle Children's, Seattle, WA, ⁴Univ of Washington, Seattle, WA, ⁵Gene Dx, Gaithersburg, MD, ⁶GeneDx, Gaithersburg, MD

Abstract Body:

Evaluation and testing of critically ill neonates for an underlying genetic condition is strongly influenced by a provider's a priori level of suspicion for a genetic condition. SeqFirst is a project whose main aim is to develop and test approaches toward centering equity for obtaining a precise genetic diagnosis at the initial point of care for infants with a critical illness using simple exclusion criteria (i.e., presentation fully explained by prematurity, trauma, or infection) rather than complex inclusion criteria to identify infants eligible for rapid short-read whole genome sequencing (srWGS). While there have been anecdotal reports of using long-read WGS (lrWGS) to detect variants missed by srWGS, no systematic comparison has been performed in a clinical setting. From January of 2021 to January of 2022 the electronic medical records for every infant under 6 months of age (n=411) admitted to the NICU at Seattle Children's Hospital were reviewed and 125/233 eligible families agreed undergo rapid srWGS at GeneDx. A result that fully or partially explained clinical findings was achieved in 70 (56%) families.

To evaluate the sensitivity (i.e., ability to detect previously identified explanatory variants) and added diagnostic value of lrWGS, we sequenced the same families using lrWGS on the Oxford Nanopore platform. Each individual was sequenced to a target depth of 40x coverage and evaluated using a standard analysis pipeline. Single nucleotide and insertion or deletion variants were called with Clair3, phased with LongPhase, and annotated using VEP. Structural variants were called using Sniffles2, CuteSV, and SVIM and prioritized using allele frequencies from publicly available datasets. Methylation was called using Guppy and differentially methylated regions (DMRs) were identified using MeOW. We report a comparison of sensitivity and diagnostic rates as well as highlight explanatory variants that were challenging or not possible to detect using srWGS.

Omics Technologies Posters - Wednesday

PB2870*. A computational approach to design a COVID-19 vaccine against a predicted SARS-CoV-2 variant: high immunogenicity, efficacy and safety of DELLERA vaccine.

Authors:

R. Paone, A. Infascelli, G. Valoti, P. Graziani, R. Ottalevi, N. Sabatini, M. Masciovecchio, C. Ciace, D. Clementoni, A. Lo Iacono, G. Iodice, M. Luciano, M. Brown, J. Caceres, I. Baldini; Dante Labs, L'Aquila, Italy

Abstract Body:

Severe acute respiratory syndrome coronavirus-2 (SARS-Cov-2) caused 6,259,945 deaths worldwide since its first identification in 2019. Even if the development of coronavirus disease 2019 (COVID-19) vaccines has proceeded at an extremely rapid pace, achieving global vaccine coverage remains a major hurdle due to the constant evolution and selection of vaccine escape variants. Therefore the tireless development of new candidate vaccines remains critical. In this study, we use a computational approach for the design of a novel candidate vaccine for a predicted SARS-CoV-2 variant. A total of 393,594 Spike (S) protein genomes of SARS-CoV-2 were analyzed with the aim to find a reference sequence of the virus most widely spread in Europe at the time and predicted (MaxEnt niche-based model) to remain as the dominant clade for the next COVID-19 wave. Therefore, we selected the viral S protein of an UK-Alpha variant of SARS-CoV-2 from a real sample (hCoV-19/England/MILK-B94A53/2020) as the basis for our vaccine design. This predicted reference shares mutations with current circulating variants, such as Omicron BA.1, BA.2, and BA.4&5. Computational screening of amino acid mutagenesis on the hotspot residues at protein-protein interfaces (between SARS-CoV-2 and ACE2 receptor) using relative free energy calculations was performed to increase the strength of the immune response against the target S protein. The preclinical study consists of 3 injections (0, 14, 28 days) per individual with two different dosages (1µg and 10µg). The results were used to evaluate: toxicity (Sprague Dawley rats, n=30), immunogenicity (Sprague Dawley rats, n=30; C57BL/6J and BALB/c mice, n=60) and efficacy (Syrian hamsters, n=24). No treatment-related deaths were reported. Evaluation of physical and clinical parameters showed lack of adverse or toxic effects. High titers of anti-RBD antibodies for ancestral and Omicron (P=.0018) variants and anti-S1/S2 antibodies for UK-Alpha (P=.0016) variant suggested the induction of a robust immune response. Increased IFN-γ (P=.00001) and IL-2 (P=.00001) levels showed a clear polarization of Th1 immune responses. Antibodies developed by the animals demonstrated effective neutralization capability virus (P<.0001). Preliminary data suggest robust protective efficacy against the ancestral SARS-CoV-2 strain as shown by less signs of lung damage in Syrian hamsters challenged with an high intranasal Sars-CoV-2 inoculum. Successful preclinical data on safety, immunogenicity and efficacy of Dellera vaccine suggest that our approach represents a promising choice to design vaccines against new predicted variants of SARS-CoV-2.

Omics Technologies Posters - Thursday

PB2871. A deep learning framework for structural variant discovery and genotyping

Authors:

V. Popic¹, C. Rohlicek¹, F. Cunial¹, K. Garimella¹, D. Meleshko², I. Hajirasouliha²; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²Weill Cornell Med., New York, NY

Abstract Body:

Structural variants (SV) are a major driver of genetic diversity and disease in the human genome and their discovery is imperative to advances in precision medicine and our understanding of human genetics. Existing SV callers rely on hand-engineered features and heuristics to model SVs, which cannot easily scale to the vast diversity of SV types nor fully harness all the information available in sequencing datasets. Since deep neural networks can learn complex abstractions directly from the data, they offer a promising approach for general SV discovery. Here we propose a novel generalizable deep learning framework, Cue, to call and genotype SVs, which effectively leverages deep learning to automatically discover the underlying salient features of different SV types, including complex and somatic subclonal SVs. In particular, we formulate SV discovery as a multi-class keypoint localization task, where keypoints correspond to breakpoints of different SVs in multi-channel images that juxtapose two genome intervals (thus capturing both SV breakpoints regardless of SV size) and simultaneously represent multiple alignment signals (such as read depth and split reads) as separate image channels. We use a stacked hourglass network to predict the type, genotype, and genomic locus of the SVs captured in each image. We allow multiple SVs to be present in the same image, thus enabling our model to easily handle nested or clustered SVs. This formulation allows Cue to be easily generalizable by design to different sequencing technologies and new SV types. To date, we have trained Cue to detect deletions, tandem duplications, inversions, inverted duplications, and inversions flanked by deletions larger than 5kbp using short reads. We have evaluated Cue using synthetic and real short-read whole-genome sequencing datasets and show that Cue outperforms state-of-the-art methods on benchmarks where high-confidence SV calls are available (namely, in simulation and the HG002 GIAB Tier1 benchmark). To further analyze Cue's performance on real data, we have used several long-read callers to orthogonally validate the results of the short-read tools. In particular, we compared the performance of short-read and long-read methods on the CHM1 and CHM13 genome mix using short-read Illumina and long-read PacBio datasets. We show that Cue achieves the highest relative concordance with long-read methods for CHM1 and CHM13 deletions. Finally, we show a proof-of-concept extension of Cue to long-read and linked-read sequencing platforms and its performance gains, especially in complex SV discovery, as compared to several state-of-the-art long-read and linked-read SV callers.

Omics Technologies Posters - Wednesday

PB2872. A Deep-learning based RNA-seq Germline Variant Caller

Authors:

D. Cook¹, A. Venkat², Y. Pouliot², F. De La Vega², P. Chang¹, A. Carroll¹; ¹Google, Mountain View, CA, ²Tempus Labs, Chicago, IL

Abstract Body:

RNA-seq is a widely used method for transcriptome analysis. The technology is often used to study gene expression, but RNA-seq data can be used for other applications including germline variant calling. Here, we introduce DeepVariant RNA-seq, a new method for performing germline variant calling from RNA-seq data. Our method encompasses updates to DeepVariant software that allow for compatibility with RNA-alignment data and a newly trained RNA-seq model.

DeepVariant is a deep-learning based variant caller that turns variant calling into an image classification problem. It does this by extracting sequence features from alignment files and encoding them as a series of "channels" that form an image. These images are then used to train a model capable of accurately calling genotypes. DeepVariant demonstrated high performance in germline variant calling from whole genome sequencing, however, RNA-seq variant calling has not previously been explored.

We demonstrate that DeepVariant RNA-seq outperforms existing methods including GATK and Platypus. We trained and evaluated DeepVariant RNA-seq on Genotype-Tissue Expression (GTEx) RNA-seq samples. In CDS regions, DeepVariant RNA-seq achieves a median F1 of 0.933 across 200 samples vs 0.903 for GATK and 0.898 for Platypus. DeepVariant also outperforms other callers in terms of indel performance, achieving a median F1 of 0.73 (precision=0.897, recall=0.614) which is a substantial improvement over GATK (F1 = 0.485) and Platypus (F1 = 0.414). We also observed improved performance across all other examined stratifications including exonic regions, gene regions, and transcript regions.

We also examine factors that influence DeepVariant RNA-seq performance. For example, we examine performance across a series of coverage thresholds, finding that restricting evaluation to regions with at least 8x coverage allows for high precision (0.985) and sensitivity (0.958). Additionally, we find that RNA integrity number and tissue type do not substantially impact DeepVariant RNA-seq performance. Instructions for downloading and using DeepVariant RNA-seq are available at <https://www.github.com/google/deepvariant>.

Omics Technologies Posters - Thursday

PB2873. A demonstrated workflow to unleash the multiomic potential of a single blood draw.

Authors:

B. Nui¹, H. Latif¹, V. Tumilasci², Y. Han¹, Y. Fan¹, D. Corney¹, P. Vishwanath¹, W. Wei¹, A. Staffa², I. DeVito¹, L. Turner¹, C. Mozdierz¹, P. Nowacki², G. Zhou¹; ¹Azenta, South Plainfield, NJ, ²Azenta, Quebec, QC, Canada

Abstract Body:

The Omics Era has greatly expanded the repertoire of approaches available for researchers and clinicians to unravel the complexity underpinning human health. Molecular biology approaches were disrupted with the advent of Next Generation Sequencing (NGS) which can elucidate the genome, epigenome, and transcriptome. NGS has been coupled with advanced DNA barcoding and microfluidics to enable multiomic characterization of single cells. More recently, highly multiplexed DNA barcoded protein/aptamer panels have been combined with NGS platforms to detail the proteome. Peripheral blood mononuclear cells (PBMCs) offer a window into the immune system that, when combined with omics tools, can provide a near holistic view of immune processes across patient cohorts. Here, we detail a proof-of-principle workflow which utilizes a single blood draw to rapidly produce a diverse set of datatypes interrogating the genome, epigenome, transcriptome, and proteome. To do so, blood draws using heparin tubes were collected and processed within 24 hours of draw to ensure high viability and yield of PBMCs. Plasma was collected and frozen during the processing of blood for PBMCs isolation and ultimately processed using Olink's kit for exploration of the proteome. PBMCs were aliquoted and cryopreserved with different vials processed for whole exome sequencing, 10X single cell sequencing, and epigenetic characterization. All of these datatypes can be produced within days of PBMC collection and in different combinations, as needed, to address the biological question at hand. With increased accessibility of omics approaches, integrative workflows such as the one described here will gain broader adoption and drive greater insights and innovation in human health applications.

Omics Technologies Posters - Wednesday

PB2874. A framework for evaluation of new or modified sequencing technologies for use in human genomics

Authors:

M. Gatz, C. Kachulis, M. Fleharty, M. Cipicchio, S. Low, B. Blumenstiel, K. Larkin, S. Gabriel, N. Lennon; Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract Body:

The rapid pace of development and innovation in sequencing technologies accelerates the genomics community's ability to understand the genetic underpinnings of disease, develop and deploy new screening and diagnostic assays, and advance the creation of new therapeutic modalities. When a new sequencing technology is released, or an existing technology is significantly changed (new instrument, new chemistry) there is a need to quickly understand the performance and limitations of the new platform so that decisions can be made about what applications may be most appropriate. The NIST Genome in a Bottle curated truth reference samples HG001 and HG002 are a critical resource for the community as they represent a standard against which we can compare all new technologies in an unbiased manner. This is particularly applicable to understanding the performance of the platforms for human genomics but also highlights performance characteristics that can be used to predict strengths and weaknesses as it relates to microbial, somatic, and other applications. We have created a framework for assessment of new technologies (or changes to existing technologies) that gives a first pass insight into the performance of the technology across variant types and genomic contexts. Genomic contexts used are those defined by the GA4GH group and include tandem repeats and homopolymers, segmental duplications, low mappability regions and a range of GC content regions. We have also included the 273 genes included in the 'challenging medically-relevant genes' set that has been curated by NIST in HG002. We have applied that framework to a variety of currently available (or soon to be available) technologies including Illumina (NovaSeq), Ultima Genomics (UG100), Element Biosciences (Aviti), Singular Genomics (G4), PacBio (Sequel IIe) and Oxford Nanopore (PromethION). We share here the results of those evaluations as well as the code and contexts required for researchers to apply this framework in their own hands. Additional sequencing metrics-based evaluations and alternative reference sample types for this analysis are also discussed.

Omics Technologies Posters - Thursday

PB2875. A High Throughput Dual Nucleic Acid Extraction Method Enables Whole Genome Sequencing from PAXGene Blood RNA Tubes

Authors:

C. Alba¹, X. Zhang¹, G. Sukumar¹, M. D. Wilkerson², C. L. Dalgard²; ¹Henry M. Jackson Fndn., Bethesda, MD, ²Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD

Abstract Body:

Next-generation sequencing methodologies for genomes and transcriptomes are powerful approaches for population genetic studies. However, it is often challenging to obtain the quantity of starting material required from retrospective sample collections, particularly when the desired outcome is PCR-free library preparation for whole genome sequencing (WGS). PAXgene Blood RNA Tubes contain an additive that stabilizes transcript abundance in blood, making them ideal for long term storage of whole blood specimens intended for isolation of total RNA. While it provides high quality and yield of total RNA, no standard vendor-provided procedure for isolating genomic DNA (gDNA) from this blood tube exists. Standard gDNA extraction methods reported by other studies resulted in a low recovery rate from PAXgene Blood RNA Tubes in part due to this additive. Validating a workflow for isolation of gDNA suitable for PCR-free WGS from non-standard blood tubes would expand the specimens usable in population genomic studies. Here, we present a high-throughput dual nucleic acid extraction method from a single PAXgene Blood RNA tube. When processed on the Qiagen QIASymphony instrument, this method results in a batch of 24 sequencing-grade gDNA and total RNA pairs, from single PAXgene Blood RNA tubes, within hours, in a cost efficient manner. DNA concentrations measured by a fluorescence-based assay resulted in dual method yields which, while lower per mL of blood input, all passed the threshold for PCR-free DNA library preparation by ligation. Sequencing library yield and size distribution were comparable to sources from EDTA blood tube, with an average library size of 575bp and library concentration of 7.3 nM. Libraries were sequenced on an Illumina HiSeq X in single library, single lane topography. All parameters for sequencing quality metrics (clusters passing filter, q30, percent alignment, error rate) were above technical thresholds. More importantly, alignment and variant calling resulted in high concordance rate (>99.9%) between whole genome data from PAXGene Blood RNA tube and EDTA blood tube-derived DNA sources. Variant-level analysis for discordant variant calls did not result in a statistically significant enrichment for a specific base-substitution at specific genomic sites. Further, the data indicated that isolated gDNAs from PAXgene Blood RNA tubes were of sufficiently high molecular weight (>60kb) to potentially provide input for long read sequencing approaches. These results suggest that cohort-scale retrospective genomic studies may be enabled by PAXGene Blood RNA tubes when canonical blood tube sources are not available.

Omics Technologies Posters - Wednesday

PB2876. A Hybrid Machine Learning and Regression Method for Cell Type Deconvolution of Spatial Barcoding-based Transcriptomic Data

Authors:

X. Yan^{1,2}, N. Li^{2,3}, Y. Liu², A. Justet¹, T. Adams¹, A. Balayev¹, N. Kaminski¹, Z. Wang²; ¹Yale Univ. Sch. of Med., New Haven, CT, ²Yale Sch. of Publ. Hlth., New Haven, CT, ³Sch. of Life Sci. and Biotechnology, Shanghai Jiao Tong Univ., Shanghai, China

Abstract Body:

Spatial barcoding-based transcriptomic (ST) technologies unbiasedly measure mRNA expression of cells with physical locations in intact tissue. However, these technologies measure the average expression profile of a few unknown cells in each capture spot, lacking single-cell resolution and requiring cell type deconvolution to enable cellular level downstream analysis. In this study, we developed a novel hybrid method that considers platform effects, spatial information and sparsity in deconvolution of ST data using reference single cell RNA sequencing (scRNA-seq) data of the same tissue type. The hybrid method consists of two steps. In the first step, a conditional variational autoencoder (CVAE) is used to remove platform effects, i.e., the systematic differences between ST data and reference scRNA-seq data. In the second step, a graph Laplacian regularized regression model (GLRM) is fitted which assumes the ST data of each capture spot to follow a Poisson-loglinear model of expression profiles of cell types present in the spot. Spatial information is considered by graph Laplacian regularization to encourage cell type composition of neighboring spots to be similar. Sparsity is enforced using L1 regularization to shrink the number of unique cell types present per spot. To evaluate and compare the performance of our method and six existing methods, we simulated spot data by coarse-graining the image-based *in situ* spatial transcriptomic data with single-cell resolution, which aggregated expression profiles of cells located within a neighborhood of a center location. Method performance was evaluated by comparing the estimated cell type composition to the ground truth in the coarse-grained *in situ* spatial data. The results showed that our method was robust to the platform effects and achieved the most accurate estimation of cell type compositions regardless of whether platform effects are present or not. When platform effects were present, machine learning-based approaches achieved the second-best performance and regression model-based approaches had the worst performance. When there were no platform effects, regression model-based approaches had the second-best performance, and machine learning-based approaches had the least accurate results. Application to real ST data with histology staining images also demonstrated accurate results of our method. In summary, by considering platform effects, spatial information and sparsity using a combination of CVAE and GLRM model, our method achieved significantly more accurate and robust results than existing methods in ST data deconvolution.

Omics Technologies Posters - Thursday

PB2877*. A Machine Learning-based approach to extract the gene-disease association discovery information from OMIM

Authors:

K. Rahit, V. Avramovic, M. Tarailo-Graovac; Univ. of Calgary, Calgary, AB, Canada

Abstract Body:

For decades, researchers have been associating disease-causing variants in genes with specific Mendelian disease (MD) conditions. Online Mendelian Inheritance in Man (OMIM) has been tracking the Gene Disease Associations (GDAs) from the first reported association. This online catalogue is curated manually and updated regularly, and contains information on the number of unrelated individuals identified in the study reporting the GDA, as well as model organisms used to validate the GDAs. However, most of the information in OMIM is textual, and heterogeneous (e.g. summarized by different experts), which complicates automated reading and understanding of the data to extrapolate GDA trends. Here, we have used Machine Learning, Natural Language Processing (NLP) in particular, to make a tool that could semantically understand OMIM text and extract the data of interest. The Gene-Phenotype Association Discovery (GPAD) tool utilizes a language-based model to the text obtained from OMIM API to extract GDA discovery-related information.

Using GPAD, we have observed a recent decline in GDA discovery. Before the era of sequencing technologies, about ~150 new GDAs were described annually. The advent of exome sequencing increased that rate to ~250 per year. It was estimated that we would be able to uncover much of unknowns about rare diseases by 2020, and by 2027 we will be able to diagnose all and might provide therapy for many of the MDs. Still, the genetic origin of thousands of known MDs remains elusive, with many more discoveries anticipated. However, our analyses revealed that in the last three (2019-2021) years, the gene-disease discovery rate decreased significantly to ~120 GDAs per year. Together by also looking into the utilization trends of patient matchmaking and model organisms for GDA discovery over the years, we may be able to speculate about the reasons behind the declining GDA trends.

Given the real-time analyzing capacity of our tool, the GPAD offers an up-to-date OMIM-dependent view of GDA discovery. Therefore, researchers and clinicians could quickly look for their gene of interest and understand the trend that could potentially help in planning and managing the research strategy accordingly; allowing to approach the GDA discovery study more efficiently.

Omics Technologies Posters - Wednesday

PB2878*. A massive proteogenomics screen identifies thousands of novel human coding sequences

Authors:

J. Xing¹, S. Sun¹, X. Cao^{1,2}; ¹Rutgers, the State Univ. of New Jersey, Piscataway, NJ, ²Zhujiang Hosp., Southern Med. Univ., Guangzhou, China

Abstract Body:

Accurate annotation of genes in the human genome is fundamental for biomedical research and genome interpretation. The Ensembl, RefSeq, and GENCODE consortiums continuously update the human genome annotations based on new computational and experimental evidence, and new genes were identified constantly. These new genes are from noncanonical open reading frames (ORFs) of known genes, genes that were previously treated as long non-coding RNAs, or from newly annotated genes. The Genotype-Tissue Expression (GTEx) project generated more than 15,000 RNA sequencing dataset from multiple-tissues of more than 800 donors, which allows to model almost all transcripts and proteins in the human genome. Using proteins translated from the GTEx transcript model, more than 21 million *in-silico* trypsin-digested peptides were generated. To identify high-confidence novel proteins with proteomics support, in this study we screened proteomic projects in the PRIDE database and selected more than 50,000 mass spectrometry (MS) runs from 923 projects. These MS data were used to validate the predicted novel peptides by GTEx. With a stringent standard, we identified almost 20,000 novel peptides and about 2,000 novel proteins that met the Human Proteome Project (HPP) Data Interpretation Guidelines. These peptides and proteins were from noncanonical ORFs, new isoforms of known genes, and un-annotated genes. Our method has shown a promising and remarkable potential in identifying novel proteins with high confidence using large public datasets. These findings will greatly improve our understanding of coding genes in the human genome and facilitate the interpretation of genomic data in biomedical research.

Omics Technologies Posters - Thursday

PB2879. A method for ABO genotyping by Sanger DNA sequencing

Authors:

E. Schreiber; Thermo Fisher Scientific, South San Francisco, CA

Abstract Body:

Same ABO blood group matching between donor and recipient decreases graft versus host disease risk in allogeneic solid organ transplantation. A recent publication (Zhou et al doi: 10.3389/fimmu.2020.608716) has documented a case of early graft dysfunction due to an ABO genotype mismatch that was not evident by conventional serotyping. Here we present a research method for genotyping the seven major ABO alleles A101, A201, B101, cis-AB01, O01, O02, O03 by Sanger-based DNA sequencing exons 6 and 7 of the human ABO gene. The method entails bi-directional sequencing of four PCR-generated amplicons and analyzing the resulting sequence trace files using Applied Biosystems SeqScape software. Deciphering the mixed sequencing traces from heterozygous alleles can often be challenging for genotyping complex loci like the ABO gene. The SeqScape analysis software generates a genotype report of the 15 alleles that determine the seven major blood types. The report is then imported into an Excel-based macro that transforms the genotype information to a searchable (query) code. Here the query is aligned and matched by a simple find operation in a look-up table to a list of codes representing the 28 different homo-and heterozygous genotype scenarios. This workflow enables the determination of common and rare ABO genotypes with possible weak phenotypes that may evade correct typing by serology. For Research Use Only. Not for use in diagnostic procedures.

Omics Technologies Posters - Wednesday

PB2880. A multi-omics approach to study monozygotic twins discordant for amyotrophic lateral sclerosis.

Authors:

M. Tosi¹, M. Zuccalà¹, F. Favero¹, L. Corrado¹, R. Croce¹, C. Basagni¹, N. Barizzone¹, L. Follia¹, A. Costa¹, F. De Marchi², E. Chinni³, L. Mazzini², D. Corà¹, M. Leone³, S. D'Alfonso¹; ¹Univ. of Eastern Piedmont, Novara, Italy, ²ALS Ctr. AOU Maggiore della Carità, Novara, Italy, ³SC Neurologia, Dipartimento di Scienze Mediche, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy

Abstract Body:

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease, characterised by progressive death of upper and lower motor neurons, whose aetiology is still partially understood. The majority of ALS cases are sporadic (sALS), while 10% are familial (fALS). To investigate genetic and epigenetic factors underlying ALS, we studied a monozygotic twin pair discordant for ALS with a multi-omics approach, combining whole exome sequencing with genome-wide methylome- and transcriptome data from whole blood and PBMCs. For methylation, we used the Illumina EPICArray and ChaAMP software for the analyses, while for gene expression study Illumina TruSeq Stranded mRNA sequencing was performed. Results of the three omics were considered independently and in combination. We identified 59 differentially expressed genes (DEGs) ($p_{\text{adj}} < 0.1$; $|\log_2\text{FC}| > 1$) and confirmed the up or downregulation for 6 of them by digital droplet PCR (ddPCR). Functional analyses on DEGs performed by GSEA, IPA and G-Profiler revealed the involvement of adaptive and innate immune system pathways. After QC, we found 2 differentially methylated probes (DMPs) ($p_{\text{adj}} \leq 0.1$) in *CACNA1G* and *VAX1* genes; while filtering by delta beta ($\Delta\beta$) values, we identified 2 probes with $\Delta\beta \leq -0.25$ (in an intergenic region and *RUSCI-AS1*) and 2 probes with $\Delta\beta \geq 0.25$ (in *AARS* and *KPNA4*). None of them fell into the highlighted DEGs. These results were compared with different larger literature datasets that included sALS and fALS patients, non-symptomatic *FUS* and *C9ORF72* mutation carriers and healthy controls, without finding any correspondence at DMP and pathway levels. For exome analyses, ExomeDepth and ClassifyCNV identified 3 deletions and 1 duplication of uncertain significance in the ALS twin. Analyses of SNV, after filtering for frequency ($\leq 0,00005$) and QC (PASS), identified 25 variants classified as VUS (n=18) or likely benign (n=7). In conclusion, we integrated different omics performing functional analyses with several bioinformatic tools that underlined a possible role of the immune system in the disease. Further understanding of these immunological results and the validation of DMPs by methylation-specific droplet digital PCR (ddMSP) combined with methylation-dependent restriction enzymes are ongoing to elucidate possible somatic genetic factors that could underlie susceptibility to sporadic ALS.

Omics Technologies Posters - Thursday

PB2881. A new approach to assess the allele frequency of small insertions and deletions

Authors:

H. Milo Rasouly, S. Bheda, S. Krishna Murthy, A. Gharavi; Columbia Univ., New York, NY

Abstract Body:

Background: Genetic diagnostics and gene discovery efforts rely on accurate estimation of population allele frequency (AF) and identification of rare protein-truncating variants. However, estimation of AF for small insertion and deletions (indels) is often inaccurate due to differences in mapping and variant calling methods. We developed a “regional AF” method to more accurately estimate AF for indels. **Method:** To calculate a regional AF, we created a python script that bins indels within different base pair windows (10, 20, 30 and 40bp) and calculates the regional AF within a bin. Variants included (1) indels with identical genomic position but different insertion or deletion, (2) overlapping indels with different genomic positions, and (3) indels within fixed genomic windows. We next calculated a regional AF for indels with gnomAD AF < 10%, using unrelated individuals from the Columbia Institute for Genomic Medicine (IGM, N= 40,762) and the publicly available gnomAD v2.1.1 exome data set (N=125,748). Rare variants were defined as gnomAD global AF smaller than 10^{-4} . We calculated the statistical difference between the mean number of rare indels in individuals using the pairwise t-test. **Results:** The proportion of rare indels with a regional AF greater than 10^{-4} was 28% for the 10bp, 32% for the 20bp, 35% for the 30bp and 38% for the 40bp windows in the IGM cohort and 14% for the 10bp, 19% for the 20bp, 22% for the 30bp and 25% for the 40bp windows in the gnomAD cohort. Amongst constrained genes ($pLi > 0.5$) with rare indels with regional AF greater than 10^{-4} (10bp window), 1169/4667 (25%) and 860/3463 (25%) are associated with diseases in OMIM in the IGM and gnomAD cohort respectively. All 860 genes identified in the gnomAD cohort were also identified in the IGM cohort. When assessing the impact of the application of a regional AF $< 10^{-4}$ on the mean number of high impact rare indels (Ensembl definitions) in the IGM cohort, the regional AF significantly reduced it from 6.61 per individual to 5.67 for the 10bp window (p -value= 2.77×10^{-08}), 5.26 for the 20bp (p -value= 2.81×10^{-17}), 4.9 for the 30bp (p -value= 2.44×10^{-28}) and 4.56 for the 40bp (p -value= 9.78×10^{-42}). **Conclusion:** Our binning method identified rare indels in regions with a high frequency of indels, enabling better estimation of AF. The differences between the results in the IGM and the gnomAD cohorts may reflect variability in mapping and variant calling. Application of indels regional AF may eliminate false positive signals and facilitate detection of novel gene- disease associations.

Omics Technologies Posters - Wednesday

PB2882*. A new framework for efficient Perturb-seq enables cheap large-scale dissection of the innate immune response and provides insight into regulatory eQTL relationships.

Authors:

D. Yao¹, L. Binan², J. Bezney², B. Simonton², C. Frangieh³, A. Regev⁴, A. Gusev⁵, B. Cleary²; ¹Harvard Univ., Cambridge, MA, ²Broad Inst., Cambridge, MA, ³MIT, Cambridge, MA, ⁴Genentech, South San Francisco, CA, ⁵Dana Farber Cancer Inst., Boston, MA

Abstract Body:

Perturb-seq is a technology that combines droplet-based single-cell RNA sequencing with pooled CRISPR screening, enabling one to study the effects of many CRISPR perturbations on the entire transcriptome simultaneously. Although the technology has been very useful in probing transcriptional regulation in diverse model systems, it remains prohibitively expensive. Moreover, large-scale eQTL studies are a population-based alternative to Perturb-seq for understanding trans-regulatory networks and human disease, but the connection between Perturb-seq and eQTL studies remains mostly unexplored. Here, we developed a new experimental and analytic framework that substantially reduces the cost of Perturb-seq experiments. First, we generate special “composite measurements” (representing random combinations of perturbations) that require only small changes to the conventional Perturb-seq protocol. We accomplish this in two ways: by encapsulating multiple cells in each droplet prior to sequencing, or by knocking out multiple random genes at a time in each cell. Next, we use a novel inference procedure to infer the effect sizes of individual perturbations from these composite measurements. We used our approach to conduct a Perturb-seq screen of 600 genes in a human macrophage cell line stimulated with LPS, which to our knowledge represents the largest Perturb-seq screen conducted in immune cells. For validation, we conducted the same screen using the conventional Perturb-seq approach. By comparing the two experiments, we show that **our efficient Perturb-seq screen produces accurate perturbation effect sizes at only 10-25% the cost of conventional Perturb-seq**. Using the causal effect estimates from our screens, we gained several new insights into the innate immune response, including a novel complex comprising two genes without prior immune annotations, XPR1 and KIDINS220, that act as negative regulators of inflammatory signaling. We also investigated whether cis-by-trans causal links learned from Perturb-seq are concordant with cis-by-trans eQTLs in a population-scale study of the same cell type and treatment as our experiment (Fairfax et al. 2014 Science). Surprisingly, we found that **cis-by-trans gene pairs implicated from Perturb-seq contain a slightly smaller proportion of significant cis-by-trans eQTLs than random gene pairs (1.5% vs. 1.7%)**, suggesting that trans-regulatory relationships learned from Perturb-seq are not reflected in eQTL datasets. Our results demonstrate that large disparities still exist between experimental assays and population-level genetic data, and further work is needed to reconcile these differences.

Omics Technologies Posters - Thursday

PB2883*. A new method for detecting genes differentiated in expression variance using single-cell RNA sequencing data

Authors:

M. Chen, A. Dahl; Univ. of Chicago, Chicago, IL

Abstract Body:

The development of single-cell RNA-sequencing (scRNA-seq) technology offers opportunities to characterize cellular heterogeneity at an unprecedented resolution and scale. This technology has been widely used to identify differentially expressed genes (DEG) that change in expression mean between cell types, such as cells before and after drug treatment. However, there could be functionally important genes that change in expression variance across cell types while remaining relatively stable in mean. Furthermore, expression variance may inform underappreciated form of cellular heterogeneity. We propose mixed models for cell type-specific gene expression (GxCTMM) to characterize variation in gene expression between cell types. We model both overall pseudobulk expression and cell type-specific pseudobulk expression. In simulations, we showed that GxCTMM is powerful and well-calibrated even for modest sample sizes (e.g., dozens of individuals). We also showed cell type-specific pseudobulk analyses add dramatic power compared to pseudobulk (up to 10 fold), illustrating the utility of scRNA-seq data. We applied GxCTMM to scRNA-seq from differentiating human iPSCs; by defining cell differentiation stages as cell types, we found many well-known genes for pluripotency and differentiation had differential variance expression; for example, the strongest differential variance signal was for POU5F1, one of three core transcription factors regulating pluripotency genes. Furthermore, GxCTMM identified 83 genes with differential variance across cell types that did not have differential mean. Finally, we extend GxCTMM to model cell type-specific *heritability* of gene expression, and present preliminary results uncovering cell type-specific genetic regulation in lupus. Overall, GxCTMM is a powerful tool to identify genes differentiated in expression variance between cell types from scRNA-seq data.

Omics Technologies Posters - Wednesday

PB2884. A novel auto-normalizing pooled library construction method that streamlines sample preparation prior to targeted hybrid capture.

Authors:

M. Costello, C. Leong, D. Bays, M. Rahardja, J. Smith, G. Rush, J. Mellor; seqWell, Inc., Beverly, MA

Abstract Body:

Hybrid capture is a well-established method for targeted sequencing, enabling cost effective analysis of larger and more complex regions than amplicon-based approaches. Most hybrid capture methods require pooling of libraries prior to capture to improve robustness and streamline downstream hybridization and capture workflows. However, there is a high cost and labor burden in individually preparing, sizing, quantifying, normalizing, and evenly pooling libraries prior to capture.

Here, we describe an approach for hybrid capture targeted sequencing based on a novel library preparation chemistry that permits pooling of samples immediately following transposase (tn5) mediated tagging, using the seqWell purePlex library preparation technology. In the purePlex process, genomic DNA is fragmented and full-length adapters containing i7 and i5 unique dual indexes (10 bp) are added in a single step. This tagging chemistry has been optimized to enable auto-normalization of libraries and consistent insert sizes across a 4-fold DNA input range between 50-200 ng, enabling equal volume pooling of samples immediately after tagging and before any downstream steps into flexible batch sizes from 8 to 96 based on target design size and desired coverage. In the example of 24-plex pooling, this reduces the number of PCR amplifications, bead based clean ups, and library QCs required for a full plate of samples from 96 at each step to just 4, enabling significant savings in time, labor, reagents, and tips. We demonstrate the suitability of this library construction method on a range of hybrid capture bait designs including human exome and compare the ease of workflow and performance to other commonly used library protocols. We anticipate that the method will have wide applicability to targeted NGS workflows that have significant multiplexing requirements, and that also require the sensitivity and performance benefits of unique dual indexing.

Omics Technologies Posters - Thursday

PB2885. A novel Bayesian factor analysis method improves detection of genes and biological processes affected by perturbations in single-cell CRISPR screening.

Authors:

Y. Zhou, K. Luo, M. Chen, X. He; Univ. of Chicago, Chicago, IL

Abstract Body:

Recent years have seen major efforts in understanding how perturbations in the genome affect gene expression phenotypes. High-throughput methods such as CROP-seq and Perturb-seq that combine multiplexed CRISPR screening with single-cell RNA readout have made it possible to assess the impact of gene knock-outs/downs on the transcriptome at a single-cell level. However, due to the sparsity and complex structure of data, analysis of single-cell CRISPR screening data remains challenging. In particular, standard differential expression analysis methods are often under-powered to detect genes affected by CRISPR perturbations.

Our proposed approach is motivated by the observation that genetic perturbations typically affect expression, not one gene at a time, but many related genes simultaneously. Indeed, transcriptomes often vary across cell types, conditions, and individuals in a modular fashion, reflecting the underlying coordinated regulation of genes in the same or related pathways. These gene modules can be inferred by matrix factorization and related techniques. We propose to infer gene modules from single-cell RNA-seq data, and borrow information across genes to improve the power of detecting differentially expressed genes. Existing factor analysis methods, however, are not readily applied to single-cell CRISPR screening data, as the factors are not directly linked with genetic perturbation, and the effects of perturbation on the expression of individual genes are not assessed.

Here we present Guided Sparse Factor Analysis (GSFA), a novel computational framework for analyzing single-cell CRISPR screening data. GSFA infers latent factors that represent co-regulated genes or gene modules, and by borrowing information from these factors, infers the effects of genetic perturbations on individual genes. We demonstrated through extensive simulation studies that GSFA detects perturbation effects with much higher power than state-of-the-art methods. Using single-cell CRISPR data from human CD8⁺ T cells and neural progenitor cells, we showed that GSFA identified biologically relevant gene modules and specific genes affected by CRISPR perturbations, many of which were missed by existing methods, providing new insights into the functions of genes involved in T cell activation and neurodevelopment.

Omics Technologies Posters - Wednesday

PB2886. A novel enzymatic fragmentation library preparation workflow that prevents sequencing artifacts for production-scale DNA-seq

Authors:

J. Haimes¹, Z. Jaafar¹, T. Harrison¹, L. Peterkin¹, M. Ranik¹, K. Scott², K. Giorda¹, B. Kudlow¹; ¹Watchmaker Genomics, Boulder, CO, ²Colorado State Univ., Dept. of Cell & Molecular Biology, Fort Collins, CO

Abstract Body:

NGS is at a turning point as the cost of sequencing has plummeted while capacity has grown rapidly. In light of these advances, library preparation methods need to be revamped for scalability without compromising quality for clinical and translational research. The clinical utility of whole genome sequencing has been well established and has tremendous potential to improve disease diagnosis and treatment.

Historically, acoustic sonication has been the preferred method for DNA fragmentation as it produces evenly sized fragments and uniform GC sequencing coverage. Yet Covaris shearing is associated with a high upfront investment, expensive consumables, and is a time consuming bottleneck in sample preparation. Here, we utilized sophisticated enzyme engineering and a multidimensional Design of Experiment approach to develop a novel library preparation method that harnesses the process benefits of enzymatic fragmentation while delivering data quality on-par with sonication.

To evaluate the performance of this approach, PCR-free whole genome sequencing (WGS) was performed using the GIAB pilot genome (NA12878) as a direct measurement of library preparation efficiency, sensitivity, and accuracy. Libraries generated using our optimized workflow exhibited a 4 to 10-fold reduction in chimeric reads and terminal hairpin artifacts compared to other enzymatic methods, and reached comparable levels to mechanically sheared DNA controls. All enzymatic fragmentation methods produced low GC-bias libraries and sufficient depth for genotyping analysis. Higher coverage combined with lower sequencing artifacts produced the highest variant calling sensitivity and specificity of the enzymatic library preparation methods, as determined using the genome in a bottle high confidence regions as ground truth. In addition to the data quality benefits, we provide workflow modifications which allow precise insert size tuning and enable rapid protocol customization for existing and emerging short read platforms.

Altogether, this enzymatic fragmentation and library preparation workflow avoids library preparation artifacts that convolute variant calling, is highly scalable, and suitable for production-scale WGS studies.

Omics Technologies Posters - Thursday

PB2887. A platform for high-resolution morphology analysis reveals tumor heterogeneity and enables label-free enrichment of target cell subpopulations

Authors:

A. Jovic¹, V. Lu², C. Johnson¹; ¹DeepCell Bio, Menlo Park, CA, ²Deepcell, Inc, 4025 Bohannon Dr., CA

Abstract Body:

Methods to study cancer at single cell resolution and capture tumor heterogeneity are crucial for better understanding and treatment of cancer. However, current isolation of tumor cells from tissue typically relies on targeting biomarkers, such as EpCAM, resulting in the isolation of tumor cell population subsets used for further molecular and functional analysis. The Deepcell platform performs high-dimensional morphology analysis of single cells using deep learning on high resolution bright-field images captured in microfluidic flow to phenotype and enrich target cells that are label-free in real-time.

We applied the Deepcell platform to train a deep convolutional neural network classifier to identify and enrich for malignant cells from non-small cell lung cancer (NSCLC) dissociated tumor cell (DTC) samples. The enriched NSCLC cells are label-free, unperturbed, and viable, making them amenable to diverse molecular and functional analyses. Importantly, the cell images can be used to generate high-dimensional morphological profiles that can be visualized by Uniform Manifold Approximation and Projection (UMAP) to uncover morphologically heterogeneous cell populations that can also be sorted for and further molecular work performed that can also be sorted out for further molecular work.

We verified enrichment of malignant cells by performing RNA and DNA analysis on the enriched sample. scRNA-Seq analysis showed populations of cells with high levels of EpCAM expression in sorted cells. CNV analysis demonstrated increased amplitude of deletion and amplification peaks relative to the pre-sorted DTC sample. Further, mutational analysis shows increased allele frequency of mutations including P53 and KRAS in sorted compared to pre-sorted samples. Multiple clusters of morphologically unique tumor cell populations were detected by UMAP analysis. We further trained the classifier to detect and enrich for each of these subpopulations; CNV, mutation, bulk RNA-Seq, and scRNA-Seq analysis revealed molecular differences between the morphologically unique subgroups. Additionally, scRNA-Seq identified EpCAM positive and negative populations underscoring the ability of the classifier to identify carcinoma cells in a label-free manner.

Here, we demonstrate the use of the Deepcell platform to perform high-dimensional morphological profiling of NSCLC tumor cells from DTC, revealing a new dimension for identifying and understanding heterogeneous tumor cell populations. Further work is planned to evaluate the link between the morphological, molecular, and functional characteristics of each subpopulation.

Omics Technologies Posters - Wednesday

PB2888. A rigorous benchmarking of methods for SARS-CoV-2 lineage detection in wastewater

Authors:

S. Knyazev^{1,2}, T. Tao², K. Wang³, B. Tyshchenko⁴, W. Ouyang⁵, A. Frolova⁴, P. Baykal^{6,7}, B. Pasaniuc¹, N. Beerenwinkel^{6,7}, N. Wu⁵, A. Zelikovsky⁸, A. Smith⁹, S. Mangul²; ¹Geffen Sch. of Med. at UCLA, Los Angeles, CA, ²Univ. of Southern California, Sch. of Pharmacy, Los Angeles, CA, ³Univ. of Southern California, Keck Sch. of Med., Los Angeles, CA, ⁴Ukraine Kyiv Academic Univ., Inst. of Molecular Biology and Genetics, Kyiv, Ukraine, ⁵Univ. of Illinois at Urbana-Champaign, Urbana, IL, ⁶ETH Zurich, Dept. of Biosystems Sci. and Engineering, Basel, Switzerland, ⁷SIB Swiss Inst. of Bioinformatics, Basel, Switzerland, ⁸Georgia State Univ., Atlanta, GA, ⁹Univ. of Southern California, Astani Dept. of Civil and Environmental Engineering, Los Angeles, CA

Abstract Body:

The COVID-19 pandemic showed that an efficient real-time response to pandemics is required to minimize the economic, social, and public health burdens resulting from pandemics. As the SARS-CoV-2 virus continues to spread, evolve, and mutate, the need for a cost-effective and efficient way to detect the presence of lineages is beyond urgent. Using wastewater-based surveillance for COVID-19 showed its efficacy in numerous countries around the globe for monitoring viral prevalence in the population. Wastewater-based surveillance has numerous advantages including that it does not require interaction with patients and can simultaneously monitor entire communities including underserved and vulnerable populations as well as asymptomatic cases. The vast majority of current SARS-CoV-2 wastewater monitoring facilities are qPCR-based and can only quantify viral load without differentiating viral strains, which prevents monitoring strain prevalence and detecting novel strains. To quantify the novel and existing strains, the sequencing of wastewater samples coupled with advanced computational tools can promise to elucidate the relative abundances of known and novel strains. Some studies show that this approach allowed monitoring not only for the number of cases but also to quantify the prevalence of viral variants including detecting strains that were absent from clinical databases-suggesting that clinical databases may be delayed. However, scalable and effective methods are yet to be developed because wastewater genomic surveillance poses technical challenges including viral genome degradation in wastewater facilities resulting in poor sequencing quality and incomplete genome coverages. We propose benchmarking existing methods for lineage detection in wastewater samples containing SARS-CoV-2. We will get state-of-the-art methods and apply them to our in-house benchmarks. The in-silico benchmarks are based on real wastewater samples and mimicking its lineage composition, genome degradation, and genomic bias identical to real samples. The genome-engineered in-vitro gold benchmarks provide positive and negative control for testing the tools. A total of 26 methods are going to be benchmarked in order to give a reference for wastewater genome-based monitoring.

Omics Technologies Posters - Wednesday

PB2890. A software platform for real-time and automated simultaneous analysis and detection of genetic diseases

Authors:

M. Garrido Navas¹, C. Kyriakidis², G. Madjarov³, D. Galevski³, A. Schack⁴, L. Krych⁴, A. Nikov⁵, Z. Velkoski⁶; ¹GENYO Ctr. for Genomics and Oncological Res.: Pfizer, Univ. of Granada, Andalusian Regional Government. Liquid Biopsy and Cancer Interception Group, Granada, Spain, ²gMendel ApS, Copenhagen, Denmark, ³Univ. Ss Cyril & Methodius, Skopje, Macedonia, The Former Yugoslav Republic of, ⁴Univ. of Copenhagen, Copenhagen, Denmark, ⁵Netcetera, Zurich, Switzerland, ⁶gMendel ApS, Copenhagen, Denmark

Abstract Body:

Eight out of ten rare diseases have a genetic cause. Often, they are chronic and life-threatening, and it takes seven years on average to receive an accurate diagnosis. Because of that, we developed Phivea[®], a software platform for real-time and automated simultaneous analysis and detection of genetic diseases using Oxford Nanopore Technologies (ONT). It enables extensive numerical and visualization analysis and accurate, systematic and timely diagnosis significantly improving disease management. The Phivea[®] platform analyzes the .fastq files generated by GridION x5 (ONT) and accesses them using a shared file system. The analysis process includes four phases:

1. Quality check (read length 900-1200bp with average Phred quality score above
2. Demultiplexing (this is done using Tochlex, a method for real-time demultiplexing ONT reads)
3. Chromosome classification (BLAST algorithm is used to compare the sequence with a library and find the sequence characteristic specific for a particular chromosome)
4. Genetic disease classification and analysis report generation (Phivea[®] performs discriminant analysis of the patients' samples compared to controls and calculates the probability of presence of genetic disorder)

The Phivea[®] platform can be applied directly on the stream of base-called DNA reads generated by the ONT device. It exceeds the limits of the real-time monitoring and analysis per DNA sample, which can significantly reduce the overall costs. The calculated throughput of the analysis pipeline is 4120 reads/s measured on a referent hardware architecture using thread parallelism of 10. Also, the current version of the Phivea[®] platform managed to correctly identify 6 different genetic disorders (Klinefelter, Turner, Down, Edwards, Patau and Prader-Willi/Angelman syndromes) with sensitivity and specificity higher than 90% on a patient level. Our technology was tested and validated on a mix of real and synthetically prepared samples, but the obtained results can be directly extrapolated for analysis and detection of other genetic disorders.

Omics Technologies Posters - Thursday

PB2891*. A spatial map of neurodevelopmental disorder risk in the developing human cortex

Authors:

F. Segato Dezem¹, S. Chen², B. Davis¹, I. Glass³, K. Aldinger⁴, J. Plummer¹; ¹Cedars Sinai Med. Ctr., Los Angeles, CA, ²Cedars-Sinai Med. Ctr., Los Angeles, CA, ³Univ of Washington, Seattle, WA, ⁴Seattle Children's Res. Inst., Seattle, WA

Abstract Body:

The diagnosis, prognosis, and treatment stratification for most neurodevelopmental disorders (NDD) remain uncertain in large part due to their high clinical and genetic heterogeneity. The genes associated with NDD are expressed during early human brain development. To understand these genes' role in pathogenicity, we sought to examine the relationship between NDD genes and their spatial expression in the context of the developing human cortex, an affected structure in NDD. Using single-cell RNAseq paired with spatial transcriptomics, we generated a high-resolution atlas of the developing human prefrontal cortex, neocortex, thalamus, and striatum (130 days old). This multi-omic approach resolved geographical, subcortical localization within developing brain architectures which are not seen by scRNAseq, or spatial RNAseq alone. After defining cell identities from the scRNAseq/spatial atlas, an integrative analysis of known NDD genes with our refined spatial map was conducted for spatial resolution of cell type with pathogenesis. To do so, we employed an empirical binary classifier (ROCit) on 345 syndromic NDD and autism genes, (curated from SFARI high confidence gene list) against gene expression profiles from the four cortical regions. From this analysis, we identified 2 clusters enriched for Glutamatergic Neuron subpopulation (GluN7) highly associated with NDD genes in the pre-frontal cortex (AUC > 0.75). The genes within this cluster most highly enriched included MEF2C, FOXP1, DPYSL2 and MEIS2. In comparison, NDD genes most enriched in the thalamus included the NDD genes: AUTS2, TCF7L2, PTMS, PPP2R1A and ACTB. Partitioning cell type identities revealed that these enriched NDD genes were most highly represented in the GluN3 and MGE Interneuron cell type. Understanding spatial gene expression during brain development and in the context of NDD risk genes provides a framework for understanding which genetic perturbations during neurodevelopment can lead to disease.

Omics Technologies Posters - Wednesday

PB2892. A submodular fairness metric for sequence-to-function models in precision genomics

Authors:

E. Robson, N. M. Ioannidis; Univ. of California, Berkeley, Berkeley, CA

Abstract Body:

Deep learning promises to revolutionize precision genomics, but the creation of clinically relevant sequence-to-function models faces significant challenges in ensuring model accuracy and model fairness, especially when participation in training datasets is largely confined to individuals from particular ancestry groups. To ensure equitable access to precision genomic technology, new tools are necessary to both measure and overcome such algorithmic bias in genomics. Here we present a label-free fairness metric to estimate a priori the portion of algorithmic bias in sequence-to-function models due to generalizability error across individuals. Its intended use case is sequence-to-function datasets in genomic machine learning, and unlike SNP counts against a reference genome, our fairness metric has no need for a pre-specified reference sequence. We provide a comparative risk analysis for individual genotypes and 26 populations in the 1000 Genomes Phase 3 dataset before extending our analysis to a supervised setting on the Geuvadis gene expression subset of 1000 Genomes. We demonstrate the utility of our proposed fairness metric in both evaluating the representativeness of a human genome sequence database and in estimating sequence-to-function model biases, and we outline key risks and preprocessing considerations for fairer genomic AI models.

Omics Technologies Posters - Thursday

PB2893. A systems biology framework to evaluate the contribution of cellular crosstalk in Alzheimer's disease genetic risk and other psychiatric traits

Authors:

R. D'Oliveira Albanus¹, L. Brase¹, S. You¹, C. Soriano-Tarraga¹, C. Cruchaga¹, B. Benitez^{2,3}, C. Karch¹, O. Harari¹; ¹Washington Univ. Sch. of Med., Saint Louis, MO, ²Beth Israel Deaconess Med. Ctr., Boston, MA, ³Harvard Med. Sch., Boston, MA

Abstract Body:

Cellular crosstalk plays an important role in brain physiology, mediating detrimental and protective processes in neurodegeneration. How crosstalk ligand/receptor proteins modulate genes associated with disease genetic risk has not been systematically characterized.

We developed a systems biology framework to characterize cellular crosstalk networks and apply it to estimate the role of crosstalk in the genetic risk of Alzheimer's disease (AD) and other psychiatric traits. We analyzed single-nucleus transcriptomic profiles of ~294K high-quality nuclei from parietal cortex of 67 donors from the Knight ADRC and DIAN brain banks, including neuropathological-free controls and AD cases. We estimated crosstalk patterns among the brain cell types based on the expression of known ligand-receptor pairs (CellPhoneDB). Crosstalk interactions involving genes implicated in AD genetic risk as receptor or ligand were enriched in microglia (OR=2.35, p=1.22e-5). Using control donors, we calculated the enrichment of crosstalk genes in other psychiatric diseases. We observed nominal enrichment of GWAS-nominated risk genes in crosstalk interactions involving oligodendrocyte precursors (schizophrenia; OR=2.6, p=1.91e-7), neurons (depression; OR=1.72, p=0.0036), and astrocytes (Parkinson's; OR=2.23, p=0.03). Focusing on AD, we determined the preferential crosstalk partners for microglia. AD crosstalk interactions were enriched between neurons and microglia (OR=2.74, p=4.41e-15), and most (64.9%) codified for microglial cell membrane receptors. These interactions included semaphorin and the PlexinA1 complex, which is mediated by the AD risk gene TREM2. We identified genes and pathways involved in the crosstalk between the emitter and receptor cells by adding functionality to CytoTalk, which infers co-expression networks using information and graph theories. The microglia network was enriched for known AD genes (OR=3.66, p=5.96e-15) and included subnetworks positively and negatively associated with Braak stage. We predict that the PlexinA1-semaphorin interaction directly modulates a sub-network associated with microglia activation and protective against AD (association with Braak stage GLM $\beta=-0.25$, p=4.6e-38), involving TREM2, APOE, and HLA genes previously implicated in AD. We replicated these results in public snRNA-seq data and are planning experimental validations.

We present a generalizable framework to investigate crosstalk in disease genetic risk quantitatively. We identify neuron-microglia interactions as an important component of AD pathology and predict that this crosstalk modulates genes associated with AD genetic risk.

Omics Technologies Posters - Wednesday

PB2894. A two-channel deep learning framework for accurate prediction of the change in protein folding free energy upon mutations

Authors:

Q. Liu¹, M. Schiller¹, Z. Zhao²; ¹Univ. of Nevada Las Vegas, Las Vegas, NV, ²Univ Texas HSC Houston, Houston, TX

Abstract Body:

Background: A stable folding is critical for a protein to perform its biological functions, and protein folding is frequently affected by various mutations. This mutational effect can be measured via the change of protein folding free energy ($\Delta\Delta G$). Thus, accurate prediction of this mutational effect on protein folding directly measures biological activities of proteins. So far, both structure-based and sequence-based methods have been proposed to predict $\Delta\Delta G$ upon mutations via machine learning techniques. However, both approaches have limitations on handcrafted features of proteins and small size of training data when applying deep learning algorithms. Pre-training of deep learning-based methods have great potential to increase the power in accurate prediction of $\Delta\Delta G$. **Methods:** We develop a two-channel learning framework (DL2ddG) for the estimation of $\Delta\Delta G$ with deep learning-based sequence representation. Sequence presentation of deep learning pretrained on Uniref90 protein sequences generates a novel feature for each position of protein sequence, and also makes it feasible to accurately predict mutation effect on limited training data with $\Delta\Delta G$. To further characterize mutation effect, we design a two-channel learning to predict $\Delta\Delta G$: one channel is for wild-type sequence representation and the other for mutant sequence representation. The difference of the two channel is learned for $\Delta\Delta G$ prediction via a neural network framework. DL2ddG can estimate $\Delta\Delta G$ for both single-point and multiple-simultaneous mutations. **Results:** We train DL2ddG on ProTherm data and evaluate it with both cross-validation and cross-dataset testing strategies on the sequence datasets of Ssym, p53, myoglobin and frataxin mutants. In our comparison with a number of sequence-based prediction methods, we demonstrate that our framework DL2ddG mostly outperforms existing methods with ~ 0.7 Pearson correlation coefficient between experimental $\Delta\Delta G$ and predicted $\Delta\Delta G$. **Conclusion:** We propose an accurate method to predict folding free energy change upon mutations. It does not need human-designed features of protein sequences and performs well on limited training data with deep learning frameworks. Our method will help for deeper understanding of protein functions affected by mutations and variant interpretation of mutant proteins, leading to potential disease marker discovery and therapeutical targets of drug development.

Omics Technologies Posters - Thursday

PB2895. A unified computing environment for genomics data storage, management, and analysis: NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space (AnVIL)

Authors:

S. Mosher¹, M. C. Schatz¹, A. A. Philippakis², A. The AnVIL Team³; ¹Johns Hopkins Univ., Baltimore, MD, ²Broad Inst., Boston, MA, ³The full list of contributors is available at: <https://anvilproject.org/about/team>, NA, MD

Abstract Body:

Recent years have seen astronomical growth in human genomics. Together with single-cell and functional genomics, electronic medical records and other biomedical data, the field is well-positioned to make great advances in human health. However, the complexity of genomic data sharing, where data is downloaded from centralized datastores for local analysis, is unsustainable and cost prohibitive. Furthermore, housing genomic data across redundant institutional compute infrastructures makes assuring data security and compliant usage of protected data a massive challenge.

The NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space, or AnVIL (<https://anvilproject.org/>) was developed to address these and other concerns by providing a unified cloud-based computing environment for genomics data storage, management and analysis. The AnVIL platform inverts the genomics data sharing model by eliminating the need for data movement, which in turn allows for active threat detection and monitoring and provides scalable, shared computing resources for researchers as needed. AnVIL currently provides harmonized access to more than 300,000 genomes from several key NHGRI projects, such as as the CCDG (Centers for Common Disease Genomics), CMG (Centers for Mendelian Genomics), eMERGE (Electronic Medical Records and Genomics), and GTEx (Genotype-Tissue Expression Project), with many more on the near horizon.

The platform is built on a set of established components that have been used in a number of flagship scientific projects. The Terra platform provides a compute environment with secure data and analysis sharing capabilities. Dockstore provides standards based sharing of containerized tools and workflows. Jupyter, R/Bioconductor and Galaxy provide analysis environments for users at all skill levels to interactively explore and understand data with thousands of tools available. The Gen3 data commons framework provides data and metadata ingest, querying, and organization. Together, AnVIL provides a collaborative environment for creating, analyzing, and sharing data and analysis workflows for even the largest projects.

Long-term, the AnVIL will provide a unified platform for ingestion and organization for a multitude of current and future genomic and genome-related datasets. Importantly, it will ease the process of acquiring access to protected datasets for investigators and drastically reduce the burden of performing large- scale integrated analyses across many datasets to fully realize the potential of ongoing data production efforts.

Omics Technologies Posters - Wednesday

PB2896. A validated clinical whole genome sequencing system for the detection of germline variants.

Authors:

C. Glidewell-Kenney, V. Onuchic, M. Bekritsky, M. Mehan, F. Fiocco, N. Khan, J. Kleinert, K. Eluwa, R. Akula Suresh Babu, Y. Liu, D. Naidu, K. Magdirila, J. Su, S. Iyer, C. Ludman, A. Leelin, C. Abaya, L. Rivas Yepes, T. Nguyen, J. Vo, D. Andrews, K. Sabourov, J. Lee, V. Corey, N. Dukler, S. Kim, F. Long, A. Crawford, R. Taft; Illumina Inc, San Diego, CA

Abstract Body:

Clinical whole-genome sequencing (WGS) has been shown to increase diagnostic yield and decrease time to diagnosis over usual care in patients with undiagnosed genetic disease. Despite the positive potential impact on patients with a suspected genetic condition, and historical cost reductions, WGS adoption by laboratories has been slowed by the complexity of bioinformatic analysis and the burden of analytical validation.

To address these needs, Illumina has developed the TruSight Whole Genome system, a fully validated clinical WGS solution for the detection of germline variants, including single-nucleotide variants (SNVs), insertions and deletions (Indels), copy number variants (CNVs), runs of homozygosity (ROH), short tandem repeat (STR) expansions, and mitochondrial SNVs. The assay starts from genomic DNA extracted from peripheral whole blood and includes library preparation reagents, sequencing on NovaSeq 6000Dx system, and variant calling with DRAGEN 3.9 software. The output of TruSight Whole Genome system includes a sample QC report and high confidence variant calls. These outputs can be utilized in a number of downstream clinical applications such as genetic disease testing.

Development included in-depth characterization and optimization of library preparation, sequencing and bioinformatic analysis. Modeling of empirical data was used to set the reportable ranges for each reportable variant type and the stratification of variants into different genomic context categories (high, intermediate, and low confidence). Filtering by these reportable ranges and genomic context categories in the analysis pipeline improved variant calling accuracy (assessed with clinical reference samples, Genome in a Bottle (GIAB) and Platinum Genome truth sets) and reproducibility (assessed by the Jaccard Similarity Index with NA12878 replicates). A set quality control metrics (coverage, bias of coverage, base quality, and contamination) for the routine monitoring of sample data were selected and shown to be effective using a dataset of over 5000 genome sequencing libraries. Validation showed the assay to be robust, with a low incidence of sample failure, no interference from endogenous or exogenous substances, and accurate and reproducible across all variant types. Taken together, these data demonstrate that the TruSight Whole Genome system can produce high quality genomes with substantially reduced laboratory validation or operating overhead and may serve as a platform for molecular genetic testing across multiple clinical indications.

Omics Technologies Posters - Thursday

PB2897. Age-dependent accumulation of somatic mutation and cell fusion in human cardiomyocytes revealed by single-cell whole-genome sequencing

Authors:

Y. Huang^{1,2}, S. Choudhury^{1,2}, J. Kim^{1,2}, Z. Zhou^{1,2}, K. Morillo¹, E. A. Maury^{1,2}, J. W. Tsai^{1,2}, M. B. Miller^{1,2,3}, M. A. Lodato⁴, S. Araten¹, N. Hilal^{1,2}, E. Lee^{1,2}, M-H. Chen¹, C. A. Walsh^{1,2,5}; ¹Boston Children's Hosp., Boston, MA, ²Harvard Med. Sch., Boston, MA, ³Brigham and Women's Hosp., Boston, MA, ⁴Univ. of Massachusetts Med. Sch., Worcester, MA, ⁵Howard Hughes Med. Inst., Boston, MA

Abstract Body:

The accumulation of somatic DNA mutations is a hallmark of aging in many dividing cells and has been reported to play critical roles in tumorigenesis and an increasing number of non-cancer diseases. Here we survey the landscape of somatic single-nucleotide variants (sSNVs) in heart muscle cells (cardiomyocytes) which normally do not proliferate but often become polyploid with age. Using single-cell whole-genome sequencing, we analyzed sSNVs of 56 single cardiomyocytes obtained from 12 postmortem donors with 0.4 - 82 years of age. Cardiomyocyte sSNVs increased strikingly with age, at rates faster than reported in non-dividing neurons. In addition to an age-related “clock-like” mutational signature commonly present in many cell types, cardiomyocytes showed a distinct mutational signature that is absent in other non-cancer cells, implicating failed nucleotide excision repair (NER) and base excision repair (BER) of oxidative DNA damage, and defective mismatch repair (MMR). A lineage tree of ~500 single cardiomyocytes, constructed using clonal sSNVs, revealed that a significant portion of cardiomyocytes with tetraploidy or higher ploidy derive from distinct clonal origins, implying cell fusion as a mechanism contributing to the polyploidization of human cardiomyocytes. Since age-accumulated sSNVs create dozens of damaging exonic mutations, cell fusion to form multiploid cardiomyocytes may represent an evolutionary mechanism of cellular genetic compensation that minimizes the complete knockout of essential genes during aging. The rates and patterns of accumulation of cardiac mutations provide a paradigm to understand the impact of genomic aging on age-related heart diseases.

Omics Technologies Posters - Wednesday

PB2898. Algorithmic and Assay-based Simplification of Multiallelic Variants for Genotyping

Authors:

B. Main, D. Oliver, V. Missirian, J. Brodsky, B. Wong, J. Gollub; Thermo Fisher Scientific, Santa Clara, CA

Abstract Body:

The Applied Biosystems™ Axiom™ genotyping microarray platform uses a two-color, hybridization- and ligation-based assay. Genotyping interrogates a SNP or indel, biallelic or multiallelic. Multiallelic genotyping uses a set of probes that is specific for each allele. Calling multiallelic variants is more complex and can be less successful than calling biallelic variants, as it uses more probe sequences and a more complex calling algorithm with more degrees of freedom. In many applications, such as research in inherited diseases, only a few pathogenic alleles are of interest, while benign alleles do not need to be distinguished from the reference or wild-type allele. We present two methods to simplify the multiallelic genotyping problem by reducing the number of alleles to distinguish, both at the algorithmic and the assay levels. Algorithmic simplification consists of summarizing a set of allele-specific signals into a single value, representing an abstract, “uninteresting” allele. Probe signals for the interesting alleles are preserved individually, or possibly could be combined as well. The number of alleles that the calling algorithm must distinguish is reduced, simplifying the task, including reducing it down to a biallelic problem. We evaluated this approach using signals from the Coriell 1000 Genomes Project sample collection on several hundred samples, assayed at 40759 3-allele loci using probes designed to be specific to each allele. Signals for the two most common alleles were computationally combined before being evaluated by the AxiomGT1 calling algorithm. 74.1% of these 3-allele loci were successfully collapsed to 2 alleles, reducing the complexity of the calling problem, resulting in equivalent or better accuracy than was achieved using multiallelic calling. Moreover, loci with improved accuracy were distinguished by standard QC measures. Assay simplification consists of intentionally designing probes that ambiguously match two or more alleles that do not need to be distinguished. We assessed the resulting probe design algorithm on a set of ~48k multiallelic variants from ClinVar, of which ~11k had two or more benign alleles. Nonspecific probe design was able to reduce the number of probes required for ~9k variants, of which ~2k (20%) were reduced to a biallelic calling problem. A test array with both allele-specific and ambiguous probes will be used to confirm equivalent sensitivity of this approach. Together, these methods of simplifying multiallelic genotype calling serve to increase the accuracy and accessibility of many variants of interest.

Omics Technologies Posters - Thursday

PB2899*. An Algorithm for Sequence Location Approximation using Nuclear Families (ASLAN) Validates Regions of the Telomere-to-Telomere Assembly and Identifies New Hotspots for Genetic Diversity

Authors:

B. Chrisman¹, K. Paskov¹, N. Stockham¹, D. Wall², J-Y. Jung¹, C. He¹, P. Washington¹; ¹Stanford Univ., Stanford, CA, ²Stanford, Stanford, CA

Abstract Body:

Although it is heavily relied on to study genetic contributors to health and disease, the current human reference genome (GRCh38) is incomplete in two major ways: it is missing large sections of heterochromatic sequence, and as a singular, linear reference genome it does not represent the full spectrum of genetic diversity that exists in the human species. In order to better understand and characterize gaps in GRCh38 and genetic diversity, we developed a method - ASLAN, an Algorithm for Sequence Location Approximation using Nuclear families - that identifies the region of origin of short reads that do not align to the GRCh38. Using unmapped reads and variant calls from whole genome sequencing (WGS) data from nuclear families, ASLAN relies on a maximum likelihood model to identify the most likely region of the genome that a subsequence belongs to, given the phasing information of family and the distribution of the subsequence in the unmapped reads. Validating ASLAN on a synthetically generated dataset, and on true reads originating from the alternative haplotypes in the decoy genome, we show that ASLAN can localize more than 90% of 100-basepair sequences with above 92% accuracy. We then run ASLAN on 100-mers from unmapped reads from WGS from over 700 families, and compare ASLAN localizations to alignment of the 100-mers to the T2T-CHM13 assembly, recently released by Telomere-to-telomere (T2T) consortia. We find that many unmapped reads in GRCh38 actually originate from telomeres and centromeres that are gaps in GRCh38 reference. We also confirm that ASLAN localizations are in high concordance with T2T-CHM13 alignments, except in the centromeres of the acrocentric chromosomes. Comparing ASLAN localizations and T2T-CHM13 alignments, we identify sequences missing from T2T-CHM13 or sequences with high divergence from their aligned region in T2T-CHM13, thus highlighting new hotspots for genetic diversity.

Omics Technologies Posters - Wednesday

PB2900*. An algorithmic framework for isoform-specific functional analysis

Authors:

P. Robinson¹, G. Karlebach¹, L. Carmody¹, S. C. Jagadish¹, E. Casiraghi², P. Hansen¹, J. Reese³, C. Mungall⁴, G. Valentini⁵; ¹The Jackson Lab., Farmington, CT, ²Università degli Studi di Milano, Italy, Milan, Italy, ³Lawrence Berkeley Natl. Lab., Berkeley, CA, ⁴Lawrence Berkeley Lab., Berkeley, CA, ⁵Università degli Studi di Milano, Milan, Italy

Abstract Body:

Gene Ontology (GO) overrepresentation analysis, which characterizes the biological mechanisms common to sets of differentially expressed genes, has proven to be one of the most successful approaches to interpreting data from high-throughput omics experiments. So far, GO overrepresentation analysis has mainly been used to evaluate differentially expressed genes, but short- and long-read RNA-seq technologies now allow increasingly accurate identification of differential alternative splicing. The availability of genome-wide isoform-level annotation remains an unsatisfied prerequisite for effectively extending this approach to interpretation of differential alternative splicing. To address this problem, we present isopret (Isoform Interpretation), a new paradigm for isoform function prediction. We observe that an isoform-level annotation is essentially an assignment of GO terms to isoforms that is consistent with gene-level annotations and is maximally consistent with isoform sequence similarity. isopret optimizes this objective using an expectation-maximization framework. It addresses several inherent computational challenges by utilizing prior knowledge and machine learning heuristics, generating an annotation that is more comprehensive and accurate than any curated or computational annotation to date. This annotation enabled us to adapt GO overrepresentation analysis to assess overrepresentation of GO annotations in differentially spliced isoforms. We used this procedure to study the relationship between differential gene expression and differential alternative splicing. We collected and uniformly processed data from 100 RNA-seq studies including investigations of development, cancer, and common disease, and used them to demonstrate that expression and splicing regulate different sets of biological functions. We found that GO terms related to protein-protein interactions (PPI) we most commonly overrepresented in sets of differentially spliced isoforms. This was further supported by enrichment of experimental PPI between the corresponding gene's products, and inclusion of known PPI domains in differential transcripts. In contrast, differential gene expression was most commonly associated with extracellular functions or signals. Isopret also implements a visual summary of analysis results, which we utilized to identify differential splicing events that provide a plausible explanation to several conditions, based on known protein domains and the scientific literature. The software and predictions are freely available at <https://github.com/TheJacksonLaboratory/isopret>.

Omics Technologies Posters - Thursday

PB2901. An empirical comparison of calling pipelines for whole genome sequencing demonstrates the importance of mapping and alignment

Authors:

A. Ziegler¹, R. O. Betschart¹, A. Thiéry¹, R. Twerenbold², D. Aguilera-Garcia³, M. Zoche³, H. Moch³, T. Zeller⁴, S. Blankenberg²; ¹Cardio-CARE, Davos, Switzerland, ²UKE, Hamburg, Germany, ³USZ, Zurich, Switzerland, ⁴Med. Univ. Hamburg-Eppendorf, Hamburg, Hamburg, Germany

Abstract Body:

Rapid advances in high-throughput DNA sequencing technologies have enabled the conduct of whole genome sequencing (WGS) studies, and several bioinformatics pipelines have become available. We conducted a WGS study involving >9000 subjects, sequenced on one Illumina NovaSeq 6000 with an average coverage of 35x using a PCR-free protocol and unique dual indices (UDI). As part of the quality control, we sequenced one genome in a bottle (GIAB) sample 70 times in different runs, and one GIAB trio in triplicate. In this study, we compared the performance of 6 pipelines, involving two mapping and alignment approaches (GATK utilizing BWA-MEM2 2.2.1, and DRAGEN 3.8.4) and three variant calling pipelines (GATK 4.2.4.1, DRAGEN 3.8.4 and DeepVariant 1.1.0). The truth set of the GIABs was used for comparison, and performance was assessed by computation time, F₁ score, precision and recall. In the mapping and alignment step, the DRAGEN pipeline was faster than the GATK with BWA-MEM2 pipeline. DRAGEN showed systematically higher F₁ score, precision, and recall values than GATK for single nucleotide variations (SNVs) and Indels in both simple-to-map and complex-to-map regions. In the variant calling step, DRAGEN was fastest. In terms of accuracy, DRAGEN and DeepVariant performed similarly and both superior to GATK, with slight advantages for DRAGEN for Indels and for DeepVariant for SNVs. The DRAGEN pipeline showed the lowest Mendelian inheritance error fraction for the GIAB trios. Mapping and alignment played a key role in variant calling of WGS, with the DRAGEN substantially outperforming GATK.

Omics Technologies Posters - Wednesday

PB2902*. An evaluation of the Ultima Genomics sequencing platform: Scalable, high-throughput sequencing for low-cost whole genome sequencing

Authors:

M. Coole¹, T. Howd¹, F. Oberstrass¹, M. Sosa², S. Low¹, T. Clark², A. Bernier¹, J. Stohlman¹, M. Shand¹, C. Nolet¹, Y. Farjoun³, N. Lennon⁴, S. Gabriel⁵; ¹The Broad Inst. of MIT and Harvard, Cambridge, MA, ²Ultima Genomics, Newark, CA, ³Lady Davis Inst., JGH, Montreal, QC, Canada, ⁴Broad Inst. Genomic Services, Cambridge, MA, ⁵Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract Body:

The cost for Human Whole Genome Sequencing (WGS) has stabilized at a little less than \$1,000 over the past few years, but the path to significant reduction is not clear. Lower WGS costs offer the chance to dramatically increase the production of population-scale studies and ultimately drive discovery and demonstration of utility of such data in the healthcare system. Ultima Genomics (UG) is a platform designed to enable continual cost decreases from a starting point of \$1/Gb and operate at an industrial scale. Here we introduce and evaluate the performance of the UG novel DNA sequencer in our lab at the Broad Institute and demonstrate an automated, production-ready library construction workflow.

In order to operationalize the platform, we leveraged an off the shelf library preparation kit, coupled with Ultima Genomics indexed adapters, to generate PCR-Free libraries. After library quantification, normalization, and pooling, samples undergo a proprietary process to anneal library pools to sequencing beads. We evolved a highly manual sample preparation process into a robust and automated workflow which results in a final bead-bound sequencing library.

Sequencing was performed on UG sequencers, which utilize mostly-natural bases in a single-read, sequencing by synthesis process. To achieve a target read length greater than 300bp, sequencing runs were 464 flows (116 cycles across the four nucleotides). Our early version of the UG sequencer has an overall runtime (including instrument setup) of 24 hours and consistently generates enough data to drive 12 samples to > 30x mean genome coverage (average = 40.1x) per run. We will present on the benchmarking performance of a large cohort of HapMap samples using early versions of the chemistry and instrumentation. Utilizing a new sequencing analysis pipeline developed to analyze the single-read, flow-based nature of the UG sequencing systems, we observe high quality SNP and short INDEL calling over the High Complexity Regions of the genome, with lower quality for long homopolymers and in Low Complexity Regions. In addition to developing an end-to-end workflow for generating PCR Free WGS data on the Ultima platform in a scalable manner that is suitable for large scale population studies, we have also explored utilizing the tool for an array of other applications including perturbation sequencing.

Omics Technologies Posters - Thursday

PB2903. Analysis of alternative polyadenylation from long-read or short-read RNA-seq with LAPA

Authors:

M. Celik, A. Mortazavi; UC Irvine, Irvine, CA

Abstract Body:

Alternative polyadenylation (APA) is a major mechanism that increases transcriptional diversity and regulates mRNA abundance. Existing computational tools to analyze APA have low precision because these tools are designed for short-read RNA-seq, which is a sub-optimal data source to study APA. Long-read RNA-seq (LR-RNA-seq) accurately detects complete transcript isoforms with polyA-tails, providing an ideal data source to study APA. However, current computational tools are incompatible with LR-RNA-seq.

Here, we introduce LAPA, a computational toolkit to study alternative polyadenylation (APA) from diverse data sources such as LR-RNA-seq and 3' sequencing (3'-seq). LAPA counts, clusters reads with polyA tails, and performs peak-calling to detect poly(A)-site in a data source agnostic manner. The resulting peaks are annotated based on genomics features and regulatory sequence elements such as polyA-signal. Finally, LAPA can perform robust statistical testing and multiple testing corrections to detect differential alternative polyadenylation. We analyzed ENCODE LR-RNA-seq data from human WTC11, mouse C2C12, and F129/Castaneus ES cells using LAPA. Benchmark of LR-RNA-seq from multiple platforms and libraries against 3'-seq with LAPA shows that LR-RNA-seq detects poly(A)-sites with high precision. Moreover, LAPA consistently improves TES site support by at least 20 % over baseline transcript annotation generated by TALON, independent of protocol or platform. Differential APA analysis detects 528 statistically significant genes with unique polyadenylation signatures between undifferentiated and differentiated C2C12 cells. Among these genes, 3' UTR elongation is significantly associated with higher expression, while shortening is linked with lower expression. This analysis reveals a link between cell state/identity and APA. Overall, our results show that LR-RNA-seq is a reliable data source to study post-transcriptional regulation by providing precise information about alternative polyadenylation.

LAPA is publicly available at github.com/mortazavilab/lapa and PyPI.

Omics Technologies Posters - Wednesday

PB2904. Analysis of pathogenic tandem repeat variation in congenital disorders.

Authors:

G. Altman, W. Lee, B. Jadhav, A. Martin Trujillo, A. J. Sharp; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Tandem repeat expansions (TREs) are known to underlie over 60 different human diseases. Pathogenic expansions of coding or intronic repeats occur in nine different genes that cause different congenital disorders. However, despite evidence TREs can act as the causative mutation in multiple congenital disorders, there have been no systematic screens for novel TREs in cohorts of patients with congenital disorders. We hypothesized that some cases of congenital disorders are caused by rare, highly penetrant pathogenic variations of short tandem repeat (STRs). Here we report results from a screen of novel TREs in 1,591 parent/offspring trios across six different congenital disorder cohorts, using whole-genome sequencing (WGS) data from the Gabriella Miller Kids First Pediatric Research Program. This represents the largest genome-wide screen of pathogenic TREs in congenital disorders to date. We genotyped STR loci genome-wide using three bioinformatic tools, optimized to identify a full range of STR variants. We used *HipSTR* to genotype STRs enclosed within single reads (less than 150 bp), *STRetch* to genotype expanded STRs (greater than 150 bp), and *ExpansionHunter Denovo (EHdn)* to genotype expanded STRs not previously cataloged. After genotyping, we searched for rare, de novo TREs present in probands but not their unaffected parents.

In the European ancestry orofacial cleft (OFC) cohort (n=330 trios), each proband had an average of 42 rare, de novo TR variants identified by *HipSTR*, 51% of which occurred in exons, introns, or UTRs. We identified multiple examples of de novo TREs in exonic, intronic, and regulatory regions of genes functionally related to OFCs, such as genes previously associated with OFCs, craniofacial super-enhancers, and genes important for early embryonic development. We performed outlier calling on the genotyping results from *STRetch*, identifying an average of 3 rare TREs per proband, including several in genes that have previously been implicated as causal for OFCs. Using the results of *EHdn*, we filtered to retain only rare, de novo TREs in probands, supported by multiple independent reads, identifying an average of 1.1 expansions per proband. Genes associated with this set of rare de novo TREs were functionally enriched for involvement in cell morphogenesis ($p=8 \times 10^{-5}$), biological adhesion ($p=9 \times 10^{-4}$), and anchoring junction ($p=0.008$), indicating functional relevance for OFCs. Our study indicates that TREs likely contribute to the genetic basis of congenital disorders, which have been missed with traditional variant analysis pipelines.

Omics Technologies Posters - Thursday

PB2905. Application of a novel scRNA-seq method based on Pre-templated Instant Partitions (PIPseq) to evaluate therapeutic induced intestinal epithelium regeneration

Authors:

P. Hettige¹, A. May-Zhang², L. Chen³, O. Pellon-Cardenas⁴, A. Dupre⁴, X. Qiu⁴, M. Verzi³; ¹Fluent BioSci.s, Watertown, MA, ²Fluent BioSci., Watertown, MA, ³Rutgers Univ., New Brunswick, NJ, ⁴Rutgers Univ., Piscataway, NJ

Abstract Body:

The intestinal epithelium is a complex, fragile, and highly labile tissue that undergoes continuous regeneration and remodeling throughout an organism's lifetime. In order to understand transcriptomic changes in the cells over time, single cell techniques are preferred over bulk RNA-seq, as it allows studying individual cellular responses in heterogeneous cell mixtures that occur in intestinal crypts or organoid cultures. Single-cell transcriptomics enables important research focused on mechanisms of intestinal homeostasis and epithelium regeneration over time. Here, we have used a novel scRNA-seq method based on Pre-templated Instant Partitions (PIPseq) to evaluate therapeutic-induced intestinal epithelium regeneration after irradiation of primary murine intestinal organoid cultures. In response to therapeutic intervention, we observe emergence of regenerative stem cell populations over time. RNA velocity plots have shown that regenerative expression profiles emerge in populations of treated cells within 24 hours after treatment. Overall, the flexibility of the PIPseq workflow, cost efficiency, and convenient stable stopping point after cell capture and lysis are particularly enabling for therapeutic interventions in time course studies as presented here.

Omics Technologies Posters - Wednesday

PB2906. Application of DRAGEN Graph read alignment to challenging medically relevant genes and other difficult regions in GRCh38 and T2T-CHM13 genomes.

Authors:

C. Roddey¹, S. Catreux¹, W-T. Chen¹, C. Colombo², J. Gimbernat², V. Jain¹, Z. Tasev¹, M. Ruehle¹, J. Han¹, R. Mehio¹; ¹Illumina, San Diego, CA, ²Illumina, Cambridge, United Kingdom

Abstract Body:

The recently released Telomere-to-Telomere (T2T) CHM13 human reference genome enables improved read mapping and variant identification for diverse samples, mainly due to the addition of about 200 million bases of sequence and the correction of structural errors present in GRCh38. To better understand the nature of these improvements, the T2T consortium developed small variant benchmarks on GRCh38 and T2T-CHM13 for 269 challenging medically relevant genes (CMRG) in the Ashkenazi son HG002. The benchmarks were used to test the accuracy of several short and long read variant calling (VC) pipelines. The DRAGEN Bio-IT Platform, which placed first in multiple categories in the 2020 PrecisionFDA VC challenge, was not included in that study, so we evaluated its performance versus the CMRG benchmarks and in other difficult regions.

The key feature driving DRAGEN's performance in the PrecisionFDA challenge was its graph read-mapping capability ('DRAGEN graph'). The GRCh38 genome was augmented with more than 800,000 short alternate contigs derived from phased population haplotypes, and 400,000+ population SNPs. These graph contigs and SNPs provide diverse haplotypes for improving read alignment in difficult regions of the genome. For the PrecisionFDA challenge, using DRAGEN graph, we observed an almost 50% reduction in DRAGEN's genome-wide VC error count (false positives + false negatives) compared to using the conventional DRAGEN read mapper ('DRAGEN non-graph').

For the current study, we created two graph references (for GRCh38 and T2T-CHM13) which included population haplotypes and SNPs in difficult-to-map and CMRG regions. We then assessed the SNP and indel VC accuracies for the DRAGEN graph and non-graph references, using Illumina read data from the recent T2T study. For GRCh38, the DRAGEN graph CMRG VC error count was 30% lower than that of DRAGEN non-graph, almost 40% lower than that of the short read VC pipeline in the T2T study, and on par with the best performing long read pipeline in that study. For T2T-CHM13, DRAGEN graph's CMRG error count was 40% lower than the non-graph count, and 55% lower than the T2T short read pipeline error count. Over GRCh38 difficult-to-map regions (GRCh38_allwomapandsegdupregions.bed), DRAGEN graph's VC error count was 50% lower than that of DRAGEN non-graph, and 70% lower than that of the T2T short read VC pipeline.

In conclusion, DRAGEN graph delivers dramatic gains in variant calling accuracy in CMRG genes with either GRCh38 or T2T-CHM13, and in GRCh38 difficult-to-map regions.

Omics Technologies Posters - Thursday

PB2907. ASCoT - Application for Sanger COncfirmatory Testing - A graphical software interface for management of confirmatory sequencing in genomic medicine

Authors:

G. Smith, E. Gilliland, K. Crooks, C. Gignoux; Univ. of Colorados Med. Campus at Anschutz, Aurora, CO

Abstract Body:

As genomic medicine programs and institutional biobanks grow, there is an increasing need for software to manage and track biospecimens needed for downstream analyses such as in confirmatory sequencing. Here, we have developed a standalone and containerized cross-platform software package for sample management particularly tailored for user-friendliness. ASCoT is developed using Python and javascript relying on the flask libraries and encodes changes in workflows via text in JSON format, allowing for high reproducibility and downstream parsing by other software packages. The software takes candidate pathogenic variants per patient as input and allows the user to sort and assign individuals to plates for simplicity. It intersects the user-submitted data with a modifiable database of amplicons and primer pairs for targeted efforts such as Sanger sequencing. It has participant, variant, and plating views allowing for ease of aliquoting and sample management to improve recording and sample workflow, with controlled signing in the user log. The program does not require an internet connection for use thus making it suitable for use in secure environments with PHI/PII as is common in clinical settings. Overall, ASCoT provides a useful tool suitable for the academic genomic medicine environment that is designed for purpose while maintaining recording and reproducibility necessary in biobanks and clinical labs

Omics Technologies Posters - Wednesday

PB2908. Assessing the impact of bioinformatics tools on genomic reproducibility: opportunities and pitfalls.

Authors:

F. Liu¹, M. Kim¹, P. I. Baykal^{2,3}, N. Beerenwinkel^{2,3}, S. Mangul⁴; ¹Dept. of Quantitative and Computational Biology, Dornsife Coll. of Letters, Arts, and Sci., Univ. of Southern California, Los Angeles, CA, ²ETH Zurich, Basel, Switzerland, ³SIB Swiss Inst. of Bioinformatics, Basel, Switzerland, ⁴Dept. of Clinical Pharmacy, USC Sch. of Pharmacy, Univ. of Southern California, Los Angeles, CA

Abstract Body:

Introduction: The reproducibility of results obtained using bioinformatics analysis is crucial, especially if such results are used in clinical settings and in the development of genomic based medicine. To assess the reproducibility of bioinformatics tools, one would repeat the same experiment multiple times across multiple laboratory sites and compare the bioinformatics results. However, such an approach carries an additional cost and is often infeasible. In this study we expand the existing approach with new *synthetic replicates* alternatives by altering the properties of sequencing reads. In this study we evaluated the reproducibility of read alignment tools and structural variant (SV) callers. We found that some bioinformatics tools failed to maintain consistent results across synthetic replicates.

Methods: We have developed a scalable and robust method to assess the reproducibility of bioinformatics tools. We generate two types of synthetic replicates, one obtained by randomly shuffling the order of reads (type 1), and the other obtained by taking the reverse complement of reads (type 2). Using these two types of synthetic replicates we assess reproducibility of selected read alignment tools and SV callers using DNA-seq paired end data. We only consider unambiguous reads, which are read reported as primary alignment only. We ran each alignment tool with original data, synthetic replicate type 1, and synthetic replicate type 2. We compared BAM files obtained from synthetic replicates to the BAM file obtained from original data.

Results: Overall, *Bowtie2*, *HISAT2*, *Minimap2*, *NextGenMap*, *SMALT*, *SNAP* and *Subread* produce consistent alignments with more than 99% of identical reads across original and technical replicate type 1 data while *BWA-mem* has the lowest percentage of ~96.5%. On the other hand, with technical replicate type 2 data we observe that apart from *NextGenMap* all other tools produce inconsistent alignments. The percentage of identical reads varies from ~99% (*NextGenMap*) to ~79% (*Bowtie2*) across the tools. We also observe that deletion coordinates reported from original data and synthetic replicate type 1 data differ for the majority of the SV caller ranging from 0% to more than 60%.

Conclusion: Our results confirm that the majority of the tools fail to maintain results across all types of technical replicates limiting applicability of bioinformatics tools in the clinical settings. We believe our method would be a useful and scalable solution for the biomedical community to assess reproducibility of newly developed bioinformatics tools at no cost and help choosing the most appropriate tool for bioinformatics analysis.

Omics Technologies Posters - Thursday

PB2909. Assessing the utility of genomic deep learning models for disease-relevant variant effect prediction

Authors:

P. Kathail¹, R. Shuai¹, R. Chung¹, C. Ye², G. Loeb², N. Ioannidis¹; ¹Univ. of California, Berkeley, Berkeley, CA, ²Univ. of California, San Francisco, San Francisco, CA

Abstract Body:

Over the past several years, highly accurate deep learning models have been developed to predict epigenetic features such as chromatin accessibility directly from DNA sequence. These models have the potential to assign function to disease-relevant non-coding variation, but their utility for variant effect prediction remains limited. It was recently noted that predictions from current deep learning models contain limited unique information about complex disease heritability, when conditioned on a broad set of coding, conservation, and regulatory annotations. **Here, we identify two features of these models that may limit their utility for disease variant functionalization.** Using two ATAC-seq datasets from kidney and immune cells, we train deep learning models, similar to the Basset architecture, to predict chromatin accessibility in multiple cell types from DNA sequence. We base our analysis of our trained models on two key ideas. First, disease-relevant non-coding variation is not uniformly distributed throughout the genome. Cell-type specific *cis* regulatory elements (CREs) are known to harbor a large fraction of the heritability of complex diseases. **We find that our models have reduced performance in cell-type specific CREs (0.30 avg. Pearson R for immune cell types; 0.39 avg. Pearson R for kidney cell types), as compared to ubiquitously accessible CREs (0.68 avg. Pearson R for immune cell types; 0.69 avg. Pearson R for kidney cell types).** Second, the accuracy metric—which we refer to as reference accuracy—used to evaluate models measures correlation between experimental measurements and model predictions for the reference genome, but does not directly measure a model’s ability to predict variant effects. Using allele-specific accessibility data to measure variant effects, we compare the reference and variant effect accuracy of several different model architectures and training procedures. We find that reference and variant effect accuracy are not correlated across these different models, **highlighting the importance of directly assessing variant effect accuracy** for the task of non-coding variant functionalization. Our results highlight the importance of evaluating models within disease-relevant regulatory regions, and show how allele-specific accessibility data and variant effect accuracy can be used as a tool for model selection.

Omics Technologies Posters - Wednesday

PB2910. Assessment of Whole Genome Sequencing Quality Metrics from a Large Cohort of Saliva-derived DNA Samples.

Authors:

G. Smith¹, D. Hupalo¹, G. Sukumar¹, C. Alba¹, J. Martin¹, C. L. Dalgard²; ¹The Henry M. Jackson Fndn., Bethesda, MD, ²Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD

Abstract Body:

Whole genome sequencing (WGS) is essential for profiling genetic variation. DNA samples for WGS can be derived from variable sources. Saliva offers the advantage of being non-invasive and more stable at room temperatures which widens the cohort possibilities. However, these techniques present a greater risk of contamination by extra species DNA content. Previously, DNA from paired blood and saliva samples showed high quality WGS, low microbial contamination, and comparable detection of SNVs in both cases. Another larger scale study with non-paired samples recommended correction for the impact of non-human reads on alignment and variant identification using decoy references that include known bacterial genres. Saliva library preparation, sequencing dropout rates and quality metrics have not been previously reported in detail and may provide insight for future study design when phlebotomy is not available. Saliva samples were collected using the OraGene OG-500 kit, DNA extracted using the user guide recommended kit, and quantification performed using a fluorescence-based assay. Sequencing libraries were prepared with a semi-automated PCR-free library preparation. Fragment analysis was used to determine bp size and qPCR was used to determine the yield. Sequencing was performed on Illumina Novaseq 6000 using a plexity of 24 or 26 for saliva and blood derived samples respectively. Raw sequencing data was demultiplexed and analysis was conducted on the HAS2.2 pipeline for alignment and variant calling.

For this comparative analysis of quality metrics, we conducted whole genome sequencing for 1755 saliva-derived and compared to 1718 blood-derived DNA samples. Library QA resulted in 0.57% of samples failing library preparation compared to 0.2% for samples extracted from blood. Median alignment percentage was determined at 91.62% and 98.23% for saliva and blood derived libraries, respectively. For single depth sequencing targeting 30x coverage, 84% of the saliva-derived libraries reached the coverage threshold while blood-derived cohorts passed at 95%. Low coverage libraries made up 16% of the total sequenced samples, and requeueing saliva-derived libraries for increased depth was possible for 12.4% of the total. Comparative analysis of SNV count, het/hom ratio, and within sample contamination (via VBID) between high (>33%) and low (<33%) extraspecies DNA content displayed no observable cohort-wide differences. Variant calling metrics were not associated with extra species DNA content as a function of percentage of aligned reads. Large cohorts with stabilized saliva using OG-500 kits should have confidence in sequencing metrics produced by WGS.

Omics Technologies Posters - Thursday

PB2911. Associations between genetically predicted levels of blood metabolites and pancreatic cancer risk

Authors:

H. Zhong, S. Liu, J. Zhu, L. Wu; Cancer Epidemiology Div., Population Sci. in the Pacific Program, Univ. of Hawaii Cancer Ctr., Univ. of Hawaii at Manoa, Honolulu, HI

Abstract Body:

Background: Pancreatic ductal adenocarcinoma (PDAC) is one of the most aggressive solid malignancies, which is featured by systematic metabolism. A growing number of conventional studies report associations between serum levels of several metabolites and PDAC risk. However, these studies could be influenced by incoherent limitations of traditional epidemiological study design. The design of genetic instruments may decrease many of these limitations. Furthermore, compared with the design of merely using metabolite associated quantitative trait loci as instruments, a design of leveraging comprehensive metabolite genetic prediction models could capture higher proportions of genetically regulated components of metabolites levels in blood, thus providing an improved power to detect the metabolite-PDAC associations. Specific microbiome features may also play a role in the metabolite-PDAC associations. **Methods:** In this study, we performed a large metabolome-wide association study (MWAS) to systematically explore associations between genetically predicted metabolite levels in blood and PDAC risk. Using data from 881 subjects of European descent in the TwinsUK Project, comprehensive genetic models were built to predict serum metabolite levels. These prediction models were applied to the genetic data of 8,280 cases and 6,728 controls included in the PanScan (I, II, and III) and PanC4 consortia. Integrated with gut microbial information, two-sample Mendelian randomization (MR) analyses were further performed to investigate the relationship among serum metabolites, gut microbiome features, and PDAC. **Results:** After assessing the metabolite-PDAC risk associations by a slightly modified TWAS/FUSION framework, we identified five metabolites (including two dipeptides) showing significant associations with PDAC risk at false discovery rate (FDR) < 0.05. Of them, three metabolites (alpha-glutamylglycine, glycylglycine, and X - 21849) showed negative associations while other two (X - 21735 and X - 24309) showed positive associations with PDAC risk. Of them, metabolite X - 21849 was found to be associated with flavonoid-degrading bacteria *Flavonifractor sp90199495*, which was also shown to be associated with PDAC risk. **Conclusion:** In summary, our research provides information of novel metabolite biomarkers for PDAC risk that warrants further investigation, which could lead to new insights into the etiology of PDAC and improved treatment options.

Omics Technologies Posters - Thursday

PB2912. Autoencoder-based Model for Genotype Imputation

Authors:

M. Song¹, J. Greenbaum², J. Luttrell¹, W. Zhou³, C. Wu⁴, Z. Luo², C. qiu², L. Zhao², K-J. Su², Q. Tian², H. Shen², P. Gong⁵, X. Shi⁶, C. Zhang¹, H-W. Deng²; ¹Univ. of Southern Mississippi, Hattiesburg, MS, ²Tulane Univ., New Orleans, LA, ³Michigan Technological Univ., Houghton, MI, ⁴Florida State Univ., Tallahassee, FL, ⁵U.S. Army Engineer Res. and Dev. Ctr., Vicksburg, MS, ⁶Temple Univ., Philadelphia, PA

Abstract Body:

Background: Genotype imputation has a wide range of applications in genome-wide association studies (GWAS), such as fine-mapping, meta-analysis, and boosting the statistical power of association tests. In recent years, a deep learning (DL)-based approach, named as sparse convolutional denoising autoencoder (SCDA), has been proposed for genotype imputation. However, it remains a challenging task on how to achieve higher imputation accuracy by optimizing the learning process in DL-based methods. **Methods:** We proposed a convolutional autoencoder (AE) model for genotype imputation. Instead of using sequential or functional methods to define the neural network architectures, we utilized the model subclassing method to build the AE model since it can be easily extended to other omics data (e.g., gene expression). The encoder of the AE model includes two 1D convolutional layers and an embedding layer. The decoder of it has an inverted symmetry structure with the encoder, including two 1D convolutional layers and an output layer with “Softmax” as its activation function. For the loss function of the AE model, we utilize categorical cross entropy (CCE) to train it as genotype data are discrete values. The main difference between our AE model and SCDA model is that we implemented a customized training loop and modified the training process by using a single batch loss rather than the average loss over batches used by SCDA. **Results:** We developed the convolutional AE model for genotype imputation with tolerance to high missing ratios. This model was evaluated using three different genotype datasets from yeast, the human leukocyte antigen (HLA) region in the 1000 Genomes Project, and the Louisiana Osteoporosis Study (LOS). Taking the HLA region imputation results as an example, the AE model achieved an average accuracy of 1.0, 0.952 and 0.950 at missing ratios of 0%, 10%, and 20%, respectively. In addition, it yielded comparable accuracy of 0.932, 0.917, and 0.906 at higher missing ratios of 50%, 70%, and 90%, respectively. Our modified AE imputation model systematically outperformed the existing SCDA model in terms of several evaluation metrics including accuracy, squared Pearson correlation coefficient, Imputation Quality Score (IQS), Hellinger score, and Scaled Euclidean Norm (SEN) score. **Conclusions:** We developed a convolutional AE imputation model with an improved learning process and achieved better imputation performance compared with SCDA model. The AE imputation model with improved accuracy has the potential to increase the power of GWAS for the LOS data generated in our ongoing research projects and other GWAS data beyond.

Omics Technologies Posters - Wednesday

PB2913. Automated Walk away NGS Sample Preparation Using a Flexible Library Prep System with Enhanced Error Correction

Authors:

S. Bigdeli¹, J. Godoski¹, D. Rhodes¹, B. Vaughan¹, R. Aguilar¹, B. Buehler¹, B. Arezi¹, J. Zhu², R. Allen¹; ¹Agilent Technologies, La Jolla, CA, ²Agilent Technologies, Santa Clara, CA

Abstract Body:

Next Generation Sequencing (NGS) target enrichment protocols for preparation of Illumina sequencing-ready libraries from samples of varying quality and input such as tissue, blood, FFPE specimens and circulating tumor DNA are in high demand. Automation of these complex protocols eliminates the need for highly trained specialists while considerably improving the NGS library prep turnaround time, quality, and reproducibility. We have developed a fully automated walk-away protocol on the Magnis NGS Prep System that is optimized for a wide range of DNA input (10-200 ng), various sample types (intact, FFPE, or ct DNA samples), different shearing methods (mechanical or enzymatic within the same automated protocol) and equipped with 192 Unique Dual sample Indices (UDI) to minimize the effects of possible index hopping. Furthermore, we incorporate inline duplex molecular barcodes at the ligation step to filter out PCR or sequencing errors by making consensus calls using MBC information. The automated workflow on Magnis is easily set up from pre-aliquoted single-use reagents to minimize contamination risk, delivers up to eight target-enriched NGS Illumina sequencing-ready libraries per run in ~ 9-10 hours without the need of any user intervention. Both catalog and custom probes can be utilized, and PCR cycles can be easily adjusted from the touchscreen to optimize the library yields. For Research Use Only. Not for use in Diagnostic procedures.

Omics Technologies Posters - Thursday

PB2914. Automation of single-cell NGS library preparation for Tapestri platform to accelerate oncology discoveries.

Authors:

A. Barner¹, K. Patel², E. Cervantes¹, E. Carvajal¹, S. Mandali¹, D. Mendoza², T. Libby², A. Li², B. Schroeder²; ¹Miroculus, San Francisco, CA, ²Mission Bio, South San Francisco, CA

Abstract Body:

In the oncology field, from discovery to diagnostics, next-generation sequencing (NGS) has played a critical role over the last decade. However, bulk NGS sequencing has limitations in truly measuring the complexity of tumor heterogeneity at a cellular level. The development of single-cell sequencing technologies allowed researchers to elucidate the intricacy of tumor clonal heterogeneity, which is pertinent for studying underlying mechanisms in cancer progression; as well as clinical applications where identifying minimal residual disease comprised of low prevalent cells carrying mutations could guide treatment regimens. Tapestri®, a high-throughput single-cell sequencing platform developed by Mission Bio, Inc., enables single-cell targeted DNA sequencing to best interrogate tumors' genomic landscape. It provides high-quality single-cell variant calling of SNV and indels, as well as clonal level CNV analysis. The entire workflow, which includes sample preparation, single-cell encapsulation, cell-specific barcoding PCR, and NGS library preparation, takes two days. The automation of NGS library preparation utilizing the digital micro fluidic system 'Miro Canvas' is demonstrated here. All of the phases in the NGS library processing were successfully translated onto Miro Canvas, including enriched library purification, enzymatic digestion, library amplification, and bead-based purification. The automation procedure reduces hands-on time significantly by 5-fold, from 2.5 hours to 30 minutes. The library quality in terms of yields, size distribution, and sequencing results from Miro Canvas is comparable to that of manual preparation. Combining Tapestri® and Miro Canvas in a stream-lined workflow greatly enhances the efficiency and sample turnaround time to accelerate discoveries in cancer biology.

Omics Technologies Posters - Wednesday

PB2915*. AUTOSurv: Interpretable autoencoder aided deep learning framework for breast cancer survival analysis incorporating multi-omics data

Authors:

L. Jiang¹, C. Xu², Y. Bai³, A. Liu¹, Y. Gong¹, Y-P. Wang³, H-W. Deng¹; ¹Tulane Univ. Sch. of Med., New Orleans, LA, ²The Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK, ³Tulane Univ. Dept. of BioMed. Engineering, New Orleans, LA

Abstract Body:

Accurate prognosis for cancer patients can provide critical information for optimizing treatment plans and thereby improving their life quality. Furthermore, combining omics data (e.g., gene expression data, miRNA expression data) and demographic/clinical information (e.g., age, ethnicity, disease stage) can give a more comprehensive view of cancer prognosis and reveal the underlying disease mechanisms at molecular level. One of the biggest challenges in utilizing omics data, however, is the high-dimension-low-sample-size issue. This may cause severe overfitting problem, undermine the generalizability of the model, and reduce the reliability of prediction results. In this study, we developed a specially designed Variational Autoencoder (VAE) to extract lower-dimension latent features from high-dimensional gene expression and miRNA expression data simultaneously. A downstream Multi-Layer Perceptron Neural Network (MLPNN) then combines the extracted latent features with some clinical variables and conducts the prognosis prediction. The VAE incorporates pathway information in its structure for better interpretability and reduced risk of overfitting. Moreover, the results suggest that a KL-annealing training scheme made our VAE more efficient in multi-omics integration. To tackle the “black-box” nature of deep neural networks, we applied an activation-based interpretation approach (i.e., DeepSHAP) on the trained model. We identified features (i.e., genes, miRNAs, clinical variables) with biggest contributions to distinguishing predicted high- and low-risk patients. In addition, by virtue of our VAE’s interpretation-friendly design, we obtained importance scores for the pathways directly and linked the highlighted pathway to the important genes. We collected demographic/clinical data, survival data, gene expression data and miRNA expression data for 1058 female patients with stage I to stage IV breast cancer from the GDC TCGA Breast Cancer cohort (BRCA) on the UCSC Xena online portal. When comparing the prediction performance of our model with conventional Cox Proportional Hazard model (Cox-PH) and other existing deep learning approaches, our model achieved significantly better prognosis prediction in various settings.

Omics Technologies Posters - Thursday

PB2916*. Bayesian Causal Inference Method applied to CRISPR perturbations Estimates the Causal Gene Regulatory Network of CD4+ T cells

Authors:

J. Weinstock^{1,2}, J. W. Freimer^{1,3}, M. Arce^{1,3}, A. Marson³, A. Battle², J. K. Pritchard^{1,4}; ¹Dept. of Genetics, Stanford Univ., Stanford, CA, ²Dept. of BioMed. Engineering, Johns Hopkins Univ., Baltimore, MD, ³Gladstone-UCSF Inst. of Genomic Immunology, UCSF, San Francisco, CA, ⁴Dept. of Biology, Stanford Univ., Stanford, CA

Abstract Body:

INTRODUCTION: Complex traits are influenced predominately through the contribution of regulatory genetic variation. The bulk (60-90%) of expression heritability is mediated through trans-regulatory variation. Despite the importance of trans-regulatory variation, mapping trans-regulators through natural genetic variation, including eQTL mapping, has been challenging. Experimental perturbations facilitate the identification of trans-regulation, thus enabling deeper understanding of the etiology of complex traits in general, and mechanistic understanding of trait specific gene regulatory networks. However, to estimate direct regulatory effects, novel computational approaches are needed to distinguish direct regulatory pathways from indirect effects to construct full causal networks. **OBJECTIVES:** We recently utilized CRISPR knockouts of 24 autoimmune associated disease genes in primary CD4+ T cells to experimentally perturb a key autoimmune network, including IL2RA, STAT5A, and JAK3 (Freimer et al., 2021). To distinguish direct regulatory effects from indirect effects, we developed a novel Bayesian structure learning method to estimate the causal regulatory network from RNA-seq, enabling a probabilistic causal estimate of the regulatory network. **METHODS:** In contrast to most causal inference procedures, which assume an acyclic directed graph, we utilized the perturbations to enable identification of a causal cyclic graph, as preliminary inspection indicated that feedback loops are prevalent among these genes. We construct a linear estimator of the perturbed RNA-seq data that estimates directed regulatory relationships that are constrained by the total regulatory effects. We include a batch effect correction and infer the posterior using Markov chain Monte Carlo. We demonstrate the feasibility of our approach using simulations. **RESULTS:** We identified a highly interconnected causal network, observing 83 directed edges among the 24 genes (posterior inclusion probability > 50%). MED12 has a direct effect on 17 of the 24 genes, suggesting that it may be a master regulator. We observe that indirect effects between genes contribute 19.5% of the variance in total regulatory effects, and we explore how the 24 perturbed genes affect non-perturbed genes. **CONCLUSION:** We use a Bayesian structure learning approach to estimate the causal directed regulatory network from perturbations in primary human immune cells. We anticipate that our approach applying bespoke causal inference methods to experimental perturbations will provide a roadmap for mapping causal regulatory networks in other cell types and contexts.

Omics Technologies Posters - Wednesday

PB2917. Benchmarking algorithms for joint integration of unpaired and paired single-cell RNA-seq and ATAC-seq data

Authors:

M. Lee^{1,2}, K. K. Kaestner¹, M. Li³; ¹Dept. of Genetics, Univ. of Pennsylvania, Philadelphia, PA, ²Graduate Group in Genomics and Computational Biology, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, ³Dept. of Biostatistics, Epidemiology and Informatics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

With the development of single-cell multi-omics technologies, we can now simultaneously measure two modalities such as mRNA and chromatin accessibility (enabled by ATAC-seq) in the same cells. These technologies allow researchers to decipher cell types/states, and their functions in health and disease with unprecedented resolution. While useful, single-cell multi-omics are expensive to generate. On the other hand, many single-modality datasets that measure either RNA or ATAC have been generated. It is desirable to integrate single-modality and multi-modality datasets to maximize the number of cells for cell type identification and facilitate consistent annotations of all cells. However, it is unclear to what extent multi-modality data such as 10x multiome aid cell type identification and the understanding of transcriptional regulation. There is also no consensus on what the best way is to integrate these three data types. One key advantage of the exclusive to multiome samples is that the paired presence of mRNA transcripts and open chromatin region indicates a likely regulatory relationship. It is worth investigating if integration methods could help recover these gene-peak associations in unpaired single-modality data. To answer these questions, we benchmarked 7 integration methods that fall into 2 categories: (1) *unpaired* integration methods that ignore the paired information provided by multiome samples and integrate them as RNA-only/ATAC-only cells; (2) *unpaired + paired* methods that integrate three data types together by treating the multiome as the reference, or that embed the three datasets into one joint space in an unsupervised manner. Using 2 publicly available multiome datasets, we simulated the 3 data types by splitting multiome cells into RNA-only, ATAC-only, and RNA+ATAC datasets. We evaluated how cell number, sequencing depth, and presence of batch structure influences the performance. For each scenario, we evaluated performance on cell type identification, batch mixing, and the recovery of gene-peak pairs (for methods that can impute RNA expression for ATAC-only cells). Preliminary results show that annotation accuracy plateaued after a certain number of multiome cells. Thus, when identifying 6 or 7 cell types, no further improvement is gained after 2000 multiome cells. Among the methods, Seurat 4, Seurat 3, and MultiVI consistently performed best at cell type identification. Seurat 3 and FigR were better at recovering gene-peak pair. However, even the best performing method only recovered 60% of gene-peak pairs. Thus, this benchmarking project serves as a guide for integrating single-cell unpaired and paired multi-omics data.

Omics Technologies Posters - Thursday

PB2918. Benchmarking NGS Integration Site Analysis Methods in Support of Long-Term Safety Monitoring of Cell and Gene Therapy Products

Authors:

J. Otto¹, Y. Fan¹, G. Bao¹, D. Corney², Q. Yu¹, B. Khan¹, L. Turner¹, C. Mozdierz¹, H. Latif¹, G. Zhou¹; ¹Azenta Life Sci., South Plainfield, NJ, ²Azenta, South Plainfield, NJ

Abstract Body:

The FDA Guidance to Industry on Long Term Follow-Up (LTFU) After Administration of Human Gene Therapy Products states the importance of longitudinal testing of gene products introduced into human subjects. Depending on the delivery mechanism, the therapeutic gene product may or may not integrate into the genome. Of particular interest are gene-product integrations near proto-oncogenes which might lead to malignancies. The FDA LTFU guidance states that recipients of an integrating gene therapy modality should be tracked for 15 years, while those receiving a non-integrating therapy modality should be tracked for 5 years. Therefore, advanced analytical methods are needed to identify, quantify, and track integration events across the genome. Here, we provide a comprehensive evaluation of methods leveraging next-generation sequencing approaches for genome-wide analysis of lentiviral integration events. Our analysis employed well-characterized standards consisting of varying copy number and known integration sites.

The approaches we characterized can be bucketed into two major groups: PCR amplification approaches and target capture-based approaches. All methods detected true positives with strong correlation to theoretical integration site dosage levels down to 1% allele frequency. Comparatively, PCR amplification-based approaches have lower data requirement per sample suggesting higher sensitivity, greater molecular capture, and lower limit of detection compared with target enrichment-based approaches. Target enrichment-based approaches can afford the flexibility to capture the integrated vector, which is of interest for characterizing partial integration events. While all methodologies performed well in our study, the choice of assay (or assays) for testing will depend on numerous factors including but not limited to the viral vector system and construct and starting material availability.

Omics Technologies Posters - Wednesday

PB2919*. Benchmarking of computational methods for ancestry prediction from RNA-seq data.

Authors:

A. Yadav¹, Y. Patel¹, W. Wen², S. Mangul¹; ¹Univ. of Southern California, Sch. of Pharmacy, Los Angeles, CA, ²Univ. of Southern California, Keck Sch. of Med., Los Angeles, CA

Abstract Body:

Advances in omics technologies have reshaped the landscape of current research providing a rich set of tools and methods. Of late, much interest has also been placed on identifying the role of genetic architecture on differential risk patterns based on admixture or one's ancestry. Outside of the gold-standard traditional methods such as WGS (whole genome sequencing), WES (whole exome sequencing) or genotype data from the 1000 Genomes project, RNA-sequencing (RNA-seq) is now popularly used to call genomic variants providing inferred estimates for single nucleotide variants (SNVs) via the GATK pipeline. RNA-seq is a prominent technology for transcriptome profiling and provides accurate measurement of the level of transcripts and genes, and is highly favorable for its low usage costs. The aim of our study is to utilize the SNVs inferred from RNA-seq data to accurately predict both global and local genetic ancestry based on individual genetic variants inferred from RNA-seq data to capture the proportion of ancestral estimates across the genome and locus specific allelic ancestral effects respectively. We then compared the ancestry inferred from RNA-seq to the gold standard methods (WGS, WES or genotype data) inferred from genomics data and identified differences and measurements for percentage of accuracy between the methods. These comparisons served as a benchmark method and in total, we have benchmarked 33 existing tools which were able to infer local and global ancestry from RNA-seq genetic variants and compared the RNA-seq based ancestry estimates with the ground truth. We also explore the effect of sequencing parameters such as the length of the RNA-seq reads, throughput and error rate. Finally, we compared the results from different tools' analyses and identified the tools with the best performance across various human populations. Our study highlights the use of RNA-seq data for ancestry estimation and will inform the genomics community about the best computational tools for identifying ancestries from RNA-seq data. We can leverage our benchmarking results that will allow the biomedical field to effectively annotate the much larger cohort of 80,000 public RNA-Seq samples to evaluate the complex population substructure by expression quantitative trait loci (eQTL) analysis within and across ancestries in a diverse cohort.

Omics Technologies Posters - Thursday

PB2920. Benchmarking splice variant predictive algorithms with saturation screens.

Authors:

C. Smith, J. O. Kitzman; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Messenger RNA splicing plays key roles in gene regulation and proteome diversification, and its disruption is implicated across a variety of human disease genes. Variants beyond essential splice dinucleotides can disrupt splicing but are challenging to identify due to the degenerate and redundant underlying sequence determinants. The over-representation of canonical splice mutations in clinical databases makes it difficult to establish a representative set of splice-disruptive variants to evaluate computational splice effect predictors. Massively parallel splicing assays (MPSAs) avoid this ascertainment bias by measuring the impact of all possible point mutations to nominate candidate splice disruptive variants. We compared high-throughput functional splicing measurements from four recent MPSAs (targeting exons of *BRCA1*, *POU1F1*, *RON*, and *FAS*) with predictions from seven contemporary variant prediction algorithms. We also benchmarked informatic predictors across a set of clinically observed variants in *MLH1* curated from the literature ($n=297$). As expected, a median of 94.5% of essential splice site mutations were correctly identified as disruptive across the tools, however at positions beyond splice sites, less than half (median across predictors=45.6%) of the splice-altering variants identified by MPSA were predicted as such informatically. Across tools and datasets, exonic variants were more challenging to predict than intronic ones. In 23 of 24 database/predictor combinations, the area under the balanced precision-recall curve (prAUBC) values - a metric of classifier performance - were lower for exonic vs intronic variants, a trend that holds even after removing canonical sites from the intronic sets. This result indicates these algorithms have room for improvement within exons. Pangolin and SpliceAI - two deep learning-based tools trained on gene annotations - corresponded best with MPSA-measured splicing outcomes, both at intronic and exonic positions (median prAUBC = 0.925 and 0.912 respectively). We then use the results of each MPSA screen and *MLH1* patient variants to propose a clinical threshold for each tool that best identifies splice disruptive variants. As a measure of genome-wide specificity, we then scored randomly selected genic variants with each tool, and found that at optimal cutoffs, the tools call 4.9-43.0% (median = 17.7%) of coding and proximal intronic SNVs as pathogenic, indicating that at acceptably sensitive empirically derived thresholds, splicing effect predictors may call a large number of false positives.

Omics Technologies Posters - Wednesday

PB2921. Beyond the "one gene, one disease" paradigm with DIVAs, an Explainable AI tool for phenotype-driven digenic variants interpretation.

Authors:

F. De Paoli¹, G. Nicora¹, I. Limongelli¹, **E. Rizzo**¹, P. Magni², R. Bellazzi², S. Zucca¹; ¹enGenome S.r.l., Pavia, Italy, ²Dept. of Electrical, Computer and BioMed. Engineering, Univ. of Pavia, Pavia, Italy

Abstract Body:

Introduction

Extending genetic analysis beyond the "one gene, one disease" paradigm can significantly uplift the diagnostic yield for rare diseases. To support scientists in the investigation of complex inheritance patterns, enGenome developed DIVAs, a phenotype-driven Artificial Intelligence (AI) tool for digenic variant interpretation that prioritizes the causative digenic combination among thousands. DIVAs can identify Dual Molecular Diagnoses (DM), wherein two disease-causing primary variants in two different genes lead to two independent clinical diagnoses, True Digenic combinations (TD), where the mutations co-occurrence in the two genes is required to trigger the disease and Composite combinations (CO), with a driver-modifier mechanism.

Materials and Methods

DIVAs is an ensemble of AI classifiers trained on almost 400 pathogenic digenic combinations extracted both from DIDA (<http://dida.ibsquare.be>), from an internal curated dataset and on a large set of benign combinations. Starting from the patient's variants, our tool computes all possible digenic combinations and exploits gene-gene interaction, gene-phenotypes association and variants impact to evaluate each gene pair. Unlike other recent tools (ORVAL, DiGePred), DIVAs also considers the patient's traits, expressed through Human Phenotype Ontology (HPO) terms, to identify the causative digenic combination. Family analysis mode is also available. Once a causative digenic combination is identified, DIVAs applies on it an additional layer of AI, based on the SHAP Explainability (XAI) methodology. The combination can be further subclassified as DM or TD/CO, allowing the user to further investigate the digenic mechanism.

Results

On 500 repeated hold-out, DIVAs showed promising results (94% of sensitivity, 97% of precision and 99% of specificity). Among 500 samples analyzed so far, here we focus on 18 clinical samples with heterogeneous phenotypes (e.g. skeletal dysplasia, hearing impairment), whose genetic diagnosis includes two concurrently mutated genes. DIVAs was able to predict as pathogenic the causative digenic pairs in 77% of them (14 out of 18). On 8 predicted pathogenic gene pairs with known digenic mechanisms, DIVAs correctly subclassified 7 of them (3/4 DM and 4/4 TD/CO). Benchmark analysis with currently available automatic digenic interpreters, such as ORVAL and DiGePred, is ongoing.

Conclusion

DIVAs is a reliable tool to identify digenic combinations. Applied to undiagnosed cases, it can uncover missing heritability of genetic disorders and uplift the diagnostic yield.

Omics Technologies Posters - Thursday

PB2922. Biallelic *CORO1A* Mutation in a Patient with Late-onset Cutaneous Tuberculosis, Epidermodysplasia Verruciformis, and Staphylococcus Aureus Infection

Authors:

H. Vahidnezhad¹, L. Youssefian², F. Palizban³, A. Saeidian⁴, A. Tavasoli¹, J. Park⁵, M. Mesdaghi⁶, H. Mahmoudi⁷, S. Zeinali⁸, V. Béziat⁹, J-L. Casanova¹⁰, J. Uitto¹; ¹Thomas Jefferson Univ., Philadelphia, PA, ²Dept. of Pathology and Lab. Med., David Geffen Sch. of Med., Univ. of California Los Angeles, Los Angeles, CA, ³Lab. of Complex Biological Systems and Bioinformatics, Inst. of Biochemistry and Biophysics, Univ. of Tehran, Tehran, Iran, Islamic Republic of, ⁴Children's Hosp. of Philadelphia, Philadelphia, PA, ⁵Geisinger Commonwealth Sch. of Med., Scranton, PA, ⁶Immunology and Allergy Dept., Mofid Children's Hosp., Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, ⁷Dept. of Dermatology, Razi Hosp., Tehran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, ⁸Dept. of Molecular Med., Biotech Resea, Tehrn, Iran, Islamic Republic of, ⁹Imagine Inst., Paris Univ., Paris, France, ¹⁰The Rockefeller Univ., New York, NY

Abstract Body:

Biallelic mutations in the *CORO1A* gene causes severe combined immunodeficiency (SCID) with susceptibilities to multiple infections including granulomatous tuberculoid leprosy, epidermodysplasia verruciformis, HHV-4, molluscum contagiosum, mucosal HHV-1, HHV-3, HHV-8 and visceral leishmaniasis. We describe a novel association of a homozygous *CORO1A* with late-onset cutaneous tuberculosis (CTB) and Staphylococcus aureus infection. Clinical and immunological evaluations, including fluorescence-activated cell-sorting and lymphocyte-transformation test, provided evidence for immunophenotype of SCID. Combined whole-exome sequencing and genome-wide homozygosity mapping were utilized to disclose candidate sequence variants. With a novel bioinformatic pipeline we utilized whole-transcriptome sequencing (RNA-Seq) to concomitantly detect the viral and bacterial agents and the consequences of detected sequence variants in the host. The proband, a 17-year-old male, was found to be homozygous for the *CORO1A*:NM_007074:exon3:c.248_249del: p.P83Rfs*10, this variant was located inside the 26 Mb of runs of homozygosity on 16p11. RNA-Seq confirmed the deleteriousness of the *CORO1A* variant in four skin biopsies revealing significant downregulation of *CORO1A*. Reads unaligned to the human genome were applied to a catalog of 926 different viruses, and pathogenic bacteria, β -HPV5, Mycobacterium tuberculosis (TB) H37Rv, and Staphylococcus aureus were detected. Collectively, we describe a previously unrecognized inborn error of immunity to CTB, indicating that autosomal recessive *CORO1A* deficiency can underlie EV and CTB. In addition, we innovatively harnessed the RNASeq data to detect abnormal microbiota underlying recalcitrant skin lesions.

Omics Technologies Posters - Wednesday

PB2923*. Blended Genome Exome (BGE) Sequencing as an Alternative to SNP Arrays for Cost Effective Imputation of Variants in Globally Diverse Cohorts.

Authors:

M. DeFelice, J. Grimsby, M. Lee, D. Zdeb, S. Anderson, B. Blumenstiel, G. Nakashian, S. DeLuca, M. Giles, M. Cipicchio, S. Chapman, C. Stevens, H. Huang, K. Yuan, D. Howrigan, A. Martin, B. Neale; The Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract Body:

Due to the cost of whole genome sequencing it is not yet economical to sequence high-coverage genomes for large cohorts of samples. As a result, researchers often combine single nucleotide polymorphism (SNP) array data with whole exome sequencing to capture both rare and common variants for the study of associations between genetic variants and disease. However, SNP arrays have been developed primarily based on data from European populations impacting their utility for population genetic estimates and analyses in more diverse global populations. To overcome the bias introduced by the selection of specific probes on an array, we developed a combined, “blended genome exome” (BGE) in which we construct PCR-free whole genomes, take an aliquot for PCR amplification, select the exome from the amplified genome through hybridization and capture, blend the exome libraries (33%) back with the PCR-free whole genome libraries (67%) (same index ligation event) into a single tube, and sequence them together (Illumina Nova S4, 2x151 bp reads, 64 samples/lane). This creates a single BAM file with a low-coverage whole genome combined with a higher coverage exome that can be used for imputing variants throughout the genome. Using high-throughput processing (384 sample batch sizes), reduced-reaction volumes and automated liquid-handling, we produced 836 BGEs from blood and saliva-derived DNA (including 64 South African, 320 Ethiopian, and 384 Chinese) samples yielding, on average, 2.6x estimated whole genome coverage and 34.1x mean exome coverage per sample. Prior development experiments showed >0.99 r^2 concordance between 30X whole genome sequencing data and imputed data derived from BGEs using Genotype Likelihoods Imputation and Phasing Method (GLIMPSE). Use in non-European association studies was the original driver for creating this new BGE sequencing product, but it will likely prove useful in many areas of genomics (i.e. polygenic and monogenic risk scores) due to its low cost, ease of lab-processing, and the quality of the data that it provides for accurate imputation of genome wide genetic variants. The process also enables the flexibility to utilize specific portions of the workflow for other non-BGE methods (i.e., exome only, non-exome targeted panels, PCR-free whole genome).

Omics Technologies Posters - Thursday

PB2924. Brain transcriptomic profiling of Parkinson's disease patients reveals disease stage specific gene expression patterns.

Authors:

C. Cappelletti¹, S. P. Henriksen², H. Geut³, A. J. M. Rozemüller³, W. D. J. van de Berg³, L. Pihlstrøm², M. Toft²; ¹OsloMet, Oslo, Norway, ²Oslo Univ. Hosp., Oslo, Norway, ³Vrije Univ.it, Amsterdam, Netherlands

Abstract Body:

Parkinson's disease (PD) is a neurodegenerative disorder mainly characterized by the appearance of motor symptoms and, in approximately 30-60% of the patients, by the development of dementia in later stages of the disease (PDD). Protein aggregates, called Lewy bodies (LBs), can be found in neurons and neuronal processes of all patients. LBs appear in different brain regions with a specific pattern described by the Braak staging system. LBs can also be found during routine postmortem examination in the brains of clinically healthy individuals that are regarded as pre-symptomatic PD individuals, termed incidental Lewy body disease (iLBD). The etiology of PD and PDD has not been elucidated yet. Genome-wide transcriptomic studies have implicated a variety of molecular mechanism in PD and PDD pathogenesis. However, these studies focused on differences between clinically defined groups and had some limitations: a small sample size, low quality samples and exclusion of long non-coding RNAs. In this study we meticulously collected the frontal cortex from a total of 84 donors including neurologically healthy controls, iLBD, PD and PDD individuals and we sequenced both the coding and non-coding RNAs. We divided our samples in four groups based on their Braak stage aiming to recapitulate disease progression. Using this setup and carefully selecting the covariates to include in our analysis, we found no differences in cell composition between the neuropathological groups. Moreover, we discovered major transcriptional changes, involving also lncRNAs, in the frontal cortex when LBs first appear in this brain region. We provided insights into the pathological mechanisms driving PD and PDD development showing functional enrichment of pathways related to mitochondrial dysfunction throughout all Braak stages. We found enrichment of pathways related to the immune response up-regulated in early stages and down-regulated in late stages of the disease (Braak stage 1-4 and 6, respectively), whereas Braak stage 5 was characterized by the functional enrichment of brain specific pathways. Additionally, we reported an association between a PD risk variant and the expression of its nearest gene.

Omics Technologies Posters - Wednesday

PB2925. Breast Cancer Specific Enhancer-Interactome Analysis Identifies Candidate Target Genes Enriched in Common Variant Risk Regions

Authors:

B. Davis¹, F. Segato Dezem², K. Ayaluri², S. Chen³, S. Gayther², J. Plummer²; ¹Cedars Sinai, LOS ANGELES, CA, ²Cedars Sinai Med. Ctr., Los Angeles, CA, ³Cedars-Sinai Med. Ctr., Los Angeles, CA

Abstract Body:

Chromatin interaction data from genome-wide interactome methods like HiChIP can provide context-specific resolution to validate candidate genes associated with common variant risk regions identified in genome wide association studies (GWAS) that are predicted to connect with functionally significant regulatory elements (REs). We used this principle to identify novel candidate genes in breast cancer risk regions where no gene associations have been found and to validate previously identified candidate susceptibility genes using transcription wide association studies (TWAS). We established enhancer-interactomes for normal breast cancer precursor cells (MCF10a) and ER+ breast cancer cells (MCF7) and then integrated this disease-specific looping data with 150 fine mapped breast cancer risk regions and 132 candidate susceptibility genes. HiChIP identified strong evidence for multiple interactions between risk alleles and gene promoters in breast cancer risk regions. We identified 3,627 genomic interactions in MCF7 cells and 374 genomic interactions in MCF10a cells linking the most likely target gene to risk alleles, gene promoter and one or more enhancers. These data show that genetic risk at chromatin loops is relevant for understanding breast cancer genetics. We interrogated global contact propensity and cell type specificity of looping for our candidate genes and detected 15,899 differential loop regions (DLR) between MCF7 and MCF10a, representing 7.8% and 20.4% of the interactions respectively. These quantitative differences in contact frequency were associated with larger changes in gene expression, assessed through RNA sequencing. We found 15 eQTL associated genes specific to MCF7 in DLRs enriched in MCF7 cells in connection with an enhancer, and 6 eQTL genes in DLRs in MCF10a cells connected predominantly with enhancers. HiChIP offers high resolution and validation of putative target genes because looping is relatively static, helping inform on the context-specific nature of disease pathogenesis. In summary, this study describes a functional framework to provide a greater understanding of the biological significance of risk alleles and candidate gene targets at breast cancer susceptibility loci identified by GWAS.

Omics Technologies Posters - Thursday

PB2926. Building a transcriptomic atlas of adult human ovarian tissue to gain insight into folliculogenesis.

Authors:

D. Hannum¹, A. S. K. Jones¹, S. Hammoud¹, J. Li², A. Shikanov¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Univ Michigan, Ann Arbor, MI

Abstract Body:

Background: Folliculogenesis in the ovary is the process by which female germ cells develop into a mature fertilizable egg. As part of the Human Cell Atlas Seed Networks, we used the emerging technologies of single-cell RNA sequencing (scRNAseq) and spatial transcriptomics to generate a comprehensive map of the adult human ovary. One of our goals is to gain deeper insights into the steroidogenic cells, which play a crucial role in folliculogenesis. **Methods:** We used the NanoString GeoMx platform to generate spatial transcriptomic data for 74 Regions of Interest (ROIs), selected by staining patterns of DAPI and several histological markers. RNA samples from individual ROIs were profiled for ~18,000 coding genes. Separately, we used the 10X Genomics platform to collect scRNAseq data for >21,000 dissociated cells from the ovaries of three healthy donors. **Results:** After stringent QC, cluster analysis of our scRNAseq data revealed four major cell types: stroma, immune, pericyte, and endothelial cells. The immune cells can be further separated into T cells, natural killer cells, macrophages, and mast cells. Refinement of the stroma cluster proved to be more difficult, as these cells occupy a continuum of transcriptome states. Iterative re-clustering, velocity analysis, marker gene-based scoring, and comparison to NanoString data helped to define a provisional set of 14 stroma sub-clusters, two of which were annotated as steroidogenic cells. By leveraging differentially expressed genes between the granulosa-enriched and theca-enriched ROIs from GeoMx data we were able to distinguish subtypes of steroidogenic cells in the scRNAseq data. Other GeoMx ROIs included a set of 12 consecutive cortical layers to study the gradual change of transcriptomic profile from the surface of the cortex towards medulla, and multiple concentric rings surrounding the oocyte in antral follicles. **Conclusion:** By using both scRNAseq and spatial transcriptomics we were able to create an atlas of cell types and spatial gene expression patterns in healthy ovarian tissue. The two data types reinforce each other, especially for subtypes of stromal cells that are difficult to distinguish by using one data type alone. These data resources contribute to the Human Cell Atlas' goal of building a reference catalog of cell types, molecular features, and spatial organization of healthy human tissues.

Omics Technologies Posters - Wednesday

PB2927. bulk and single-cell gene co-expression network analysis of AD using PyWGCNA

Authors:

N. Rezaie, A. Mortazavi; Univ. of California, Irvine, Irvine, CA

Abstract Body:

Mouse models of human diseases such as Alzheimer's disease (AD) are one of the most important tools for studying pathogenic mechanisms and testing interventions and therapeutics. While much of the characterization of gene expression changes to date has been done with “bulk” RNA-seq in tissues such as the hippocampus and cortex, the advent of single-cell and single nucleus functional genomics assays allows us to analyze those changes at the level of individual cell subtypes. However, we would like to compare the relatively sparse and noisy single-cell assays to bulk RNA-seq to deconvolve which cells are driving the changes observed in the previously collected bulk datasets.

Here we develop a python package called PyWGCNA that is designed to identify and compare gene co-expression modules in bulk RNA-seq data as well as compare them to scRNA-seq data and snRNA-seq clusters. As a proof of concept, we apply this to bulk RNA-seq of the cortex and hippocampus from 5xFAD and 3xTgAD as well as matching control mice from MODEL-AD to identify gene modules associated with the genotype and phenotypes. Comparisons of these modules to single-nuclei and modules associated with AD in humans from AMP-AD reveal the specific cell types driving specific modules in a region-specific manner. This deconvolution approach allows us to interrogate existing bulk datasets using single-cell data to look for the signature of cell-type-specific changes in gene expression.

Omics Technologies Posters - Thursday

PB2928. Calling structural variants in extended rat pedigree using PacBio HiFi sequencing.

Authors:

D. Chen¹, K. Nguyen¹, K. Cohen¹, O. Poleskaya¹, J. Sebat^{1,2,3,4}, A. Palmer^{1,2}; ¹Dept. of Psychiatry, Univ. of California San Diego, La Jolla, CA, ²Inst. for Genomic Med., Univ. of California San Diego, La Jolla, CA, ³Beyster Ctr. for Psychiatric Genomics, Univ. of California San Diego, La Jolla, CA, ⁴Dept. of Cellular and Molecular Med., Univ. of California San Diego, La Jolla, CA

Abstract Body:

Genome-wide association studies (GWAS) in mammalian model organisms, such as rats, are valuable compliments to GWAS in humans, and present the opportunity to confirm associations with direct experimental manipulations, which can elucidate novel biological mechanisms. However, determining the causal genes within an associated locus remains challenging for a variety of reasons. The causal alleles may not be single nucleotide polymorphisms (SNPs), and SNPs may not always adequately tag the causal variants, such as structural variants (SVs) and short tandem repeats (STRs). To address this limitation, we are using Pacific Biosciences (PacBio) high fidelity (HiFi) sequencing to discover structural variants and perform GWAS using SVs for gene expression and complex behavioral traits in rats. All analyses are performed on our in-house SVs calling using ~10x coverage HiFi reads. SVs discovery is being performed in the 8 inbred founders of the outbred heterogeneous stock (HS) population, which we are using for GWAS. By imputing SVs into the outbred population, we will be able to identify associations between SVs and a variety of traits in more than 10,000 outbred HS rats that have been extensively genotyped and phenotyped with the goal of revealing novel genetic mechanisms underlying drug abuse and other human diseases.

Omics Technologies Posters - Wednesday

PB2929. Cas9 targeted enrichment and adaptive sampling for characterizing repetitive elements.

Authors:

C. Mumm, K. Van Deynze, W. Zhou, J. Switzenberg, C. Maltby, P. Todd, R. Mills, A. Boyle; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

The characterization of complex and repetitive structural variation in the genome remains a challenge using short read sequencing technologies. Examples of such variation include mobile element insertions (MEIs), short tandem repeat (STR) expansions, and Human Papilloma Virus (HPV) integrations. Transposable elements and viruses have the potential to cause disease through insertional mutagenesis as well as through effects on local gene expression and chromatin architecture, and accurate characterization of these elements is important in understanding about disease risk and clinical diagnosis. In addition, STR expansions are associated with over 50 neuropathological disorders and the mechanism by which these expansions bring about disease is not yet fully understood. Here, we combine two Oxford Nanopore Technology (ONT) long-read sequencing enrichment strategies to characterize repetitive elements, nanopore Cas9-targeted sequencing (nCAsTs) and readfish adaptive sampling. CRISPR-Cas9 enrichment uses sgRNA to selectively ligate adapters to target sequences and we use a combination of internal and flanking sgRNAs to capture our regions of interest. Computational selection during sequencing using readfish further allows us to enrich for our targets. Using these reads we can also investigate the CpG methylation at these loci and surrounding regions to further explore how they can contribute to disease. Additionally, we can use these techniques to identify events on a per-cell basis. We apply these techniques to selectively capture our targets and characterize complex structural variation in GM12878, as well as patient derived cell lines and tissues.

Omics Technologies Posters - Thursday

PB2930. Causal relationship analysis of DNA openness and RNA expression at the single cell level using scRNA-seq and scATAC-seq data.

Authors:

J. Park, B. Han; Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Abstract Body:

Single cell RNA sequencing (scRNA-seq) is widely used to analyze the expression environment of individual cells and classify the cell types. Recently, Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) was also implemented at the single cell level. While RNA-seq can figure out the expression level of the gene, ATAC-seq can recognize whether the chromatin region is opened. In general, chromatin needs to be opened so that the general expression process involving the access of transcription factor (TF) into DNA can progress. Therefore, there is a causal relationship from chromatin openness to expression, and we can expect that this relationship can be inferred from the RNA-seq and ATAC-seq data. In this study, we tried to analyze whether ATAC-seq and RNA-data show any causal relationship, especially at the single cell level. If the causality is clear, the expression level of the gene will be high in the vicinity of the open region, and vice versa. We analyzed this relationship by employing the PBMC 10K public data provided by 10x genomics, in which scRNA-seq and scATAC-seq were performed simultaneously for the same cell. For analysis, we used Signac which is an R package for analyzing single-cell chromatin data (Stuart et al., 2021). We were able to obtain a notable correlation between gene expressions and open chromatin regions at the single cell level. For example, the Pearson correlation coefficient was 0.466 for a most correlated peak of *MS4AI*, and was 0.196 for *LYZ*. However, in the region where the expression count exists, a lot of cells with an open count of 0 were found (i.e. 889 cells out of 1366 *MS4AI* expressed cells, 6879 cells out of 7569 *LYZ* expressed cells). This phenomenon (expressions at not-detected-as-open region) was consistent with recent reports that the sensitivity of ATAC-seq was as low as 5-15% (Preissl et al., 2018). Because of this low sensitivity, the causal relationship between scATAC-seq and scRNA-seq could not be directly inferred from the current data. Until the sensitivity of ATAC-seq phenomenally increases in the future, it is unclear whether the development of a new methodology for modeling the causal relationship between openness and expressions will be necessary.

Omics Technologies Posters - Wednesday

PB2931. CCC: An efficient not-only-linear correlation coefficient based on machine learning.

Authors:

M. Pividori¹, M. D. Ritchie¹, D. H. Milone², C. S. Greene³; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Univ. Nac. del Litoral, Consejo Nac. de Investigaciones Científicas y Técnicas (CONICET), Santa Fe, Argentina, ³Univ. of Colorado Sch. of Med., Aurora, CO

Abstract Body:

Background. New technologies have vastly improved data collection, generating a deluge of information across different disciplines that provides unique opportunities to address unanswered scientific questions. Correlation analysis is an essential statistical technique for discovering relationships between variables, and it has been used in exploratory data mining methods such as clustering or community detection.

These techniques compute a similarity value between a pair of variables of interest, such as genes or disease-relevant lifestyle factors. In transcriptomics, many genomic analyses start with estimating the correlation between genes. More sophisticated approaches built on correlation analysis can suggest gene function, capture important interactions in a living organism that can uncover molecular processes in other species, and reveal complex transcriptional mechanisms underlying human diseases. The Pearson correlation coefficient is widely used because it reveals intuitive relationships and can be computed quickly. However, it can only capture linear patterns and may miss complex yet critical relationships. Thus, improvements in these techniques could have enormous consequences in industry and research.

Methods. We introduce the Clustermatch Correlation Coefficient (CCC), an efficient not-only-linear coefficient based on machine learning models. CCC assumes that if there is a relationship between two variables measured on a set of objects, then the clusterings of those objects using each variable should match. CCC has a single parameter that limits the maximum complexity of relationships found and computation time. To assess its performance, we applied Pearson, Spearman, Maximal Information Coefficient (MIC) and CCC to gene expression data from GTEx and used the Genome-wide Analysis of gene Networks in Tissues (GIANT, without GTEx samples) for replication.

Results. CCC captured both strong linear relationships and novel nonlinear patterns, which were entirely missed by standard coefficients. For example, *UTY* and *KDM6A* (paralogs) showed a consistent nonlinear relationship across all tissues, which is explained by sex differences in expression. Gene pairs detected by CCC had significantly higher interaction probabilities in tissue-specific gene networks from GIANT. We also found that CCC behaves similarly to MIC, although it is up to two orders of magnitude faster to compute. Furthermore, the CCC's ability to efficiently handle diverse data types (including numerical and categorical features) reduces complex preprocessing steps and makes it appealing to analyze large and heterogeneous repositories.

Omics Technologies Posters - Thursday

PB2932. Cell manufacturing genome integrity analysis by optical genome mapping.

Authors:

A. Pang, A. Reyes, A. Chaubey, A. Hastie; Bionano Genomics, San Diego, CA

Abstract Body:

There are several types of cell lines and manufactured cell-based products used in research and therapeutics, such as transformed cell lines for constitutional and oncological research, stem cell lines like iPSC, CAR-T cells, and various types of engineered cells. Cells in culture can lose genome integrity often seen as structural and numerical abnormalities. While quality control (QC) by karyotyping is effective in detecting large DNA rearrangements, it is very low resolution, laborious, and highly specialized. PCR and targeted sequencing can characterize small variants at specific loci, but they are not able to identify rearrangements beyond the targets. Finally, while whole genome sequencing can identify small variants genome-wide, it has limited sensitivity in finding structural variants (SVs), especially at low allele fraction. Optical genome mapping (OGM) is an emerging technique for SV analysis in cell manufacturing applications. Here we present the performance of three workflows for genome integrity analysis. To detect clonal variants, 400 Gbp of data is collected from the pre-edited and edited samples. Homozygous or heterozygous SVs are called for each sample and then compared with a built-in comparison tool to identify unique SVs in the edited samples and not in the pre-edited controls. Alternatively, to discover somatic SVs down to 5% or 1% variant allele fraction (VAF), one would generate 1.5 or 5.0 Tbp of data with a longer machine run (24-65 hour runs). We designed a series of simulations to examine the limit of detection of OGM. We targeted a total coverage of 5 Tbp, and ran the SV analysis workflow to detect variants from 100% to <1% VAF. The results indicated that OGM has a sensitivity of detecting small insertions and deletions (> 5kbp) at about 0.25% to 2.5% with large megabase duplications, deletions, and translocations down to ~1%. We have applied clonal and subclonal workflows to verify integrity of genomes after cell line immortalization, induce pluripotency, transgene-integration, and gene-editing. With easy, cost effective, robust, and sensitive workflows, we envision that the OGM technique to be a valuable solution for genome modification and integrity QC.

Omics Technologies Posters - Wednesday

PB2933. Characterizing the genetic etiology of patients with neuromuscular disease through the integration of omics data

Authors:

V. Triassi^{1,2}, É. Bareke¹, S. Audet^{1,3}, C. Brunel-Guitton⁴, M. Tétreault^{1,3}; ¹CRCHUM, Montréal, QC, Canada, ²Dept. of BioInformatics, Université de Montréal, Montréal, QC, Canada, ³Dept. of NeuroSci.s, Université de Montréal, Montréal, QC, Canada, ⁴BC Children's Hosp., Vancouver, BC, Canada

Abstract Body:

Advances in sequencing technologies have played an important role in the molecular diagnosis of rare diseases, such as myopathies and muscular dystrophies. However, several patients with these neuromuscular diseases remain undiagnosed. This is due to the great clinical and genetic heterogeneity as well as the highly polymorphic nature of the genes associated with myopathies and muscular dystrophies. The interpretation of genetic data is a great challenge and genetic testing often results in the identification of variants of unknown significance. Several of these variants may disrupt normal RNA splicing or affect gene expression.

To determine if variants have a functional impact, our pipeline focuses on alternative splicing as well as the integration of exome and transcriptome data. Our aim is to put in place a pipeline identifying and characterizing variants of interest in a pathological context. We hypothesize that a portion of patients without a diagnosis for their neuromuscular disorder is explained by intronic variants having a regulatory role or affecting mRNA splicing and abundance. This multi-omics approach will make it possible to determine whether the variants have a functional impact.

To do so, we performed exome and RNA sequencing using muscle biopsies from 4 patients. Data was aligned with STAR (RNA) or BWA (exome) and annotated with ANNOVAR as well as custom scripts. To identify splicing events, we used two tools: SpliceAI, an effective tool to predict the impact of variants on alternative splicing, and rMATS, which best predicts whether an alternative splicing event occurs from exon-intron junctions. A combination of these two tools further increases detection efficiency. Since we do not have replicate for our data, an exploratory differential expression analysis was performed with LPEseq. Lastly, several custom python scripts filter and integrate transcriptomic and genomic data. For now, potential pathogenic variants are found for two of our patients, but further investigations are needed. The variants sought are rare and therefore require a coherent and efficient pipeline to facilitate the detection of these events. In addition, the integration of omics data provides a more interesting and prioritized gene research avenue. This project will have a significant benefit for patients by allowing them to obtain a diagnosis and thus have access to better clinical follow-up. Ultimately, we want to identify pathogenic variants for our patients with our pipeline, recruit additional patients without molecular diagnosis, optimize and automate the pipeline with the intention that it can be used by other researchers working on rare diseases.

Omics Technologies Posters - Thursday

PB2934. Characterizing the immune dysfunction in X-CGD using single-cell transcriptomics.

Authors:

S. Muzumdar¹, S. Ballouz², F. Lam³, M. DeGrange⁴, S. Kreuzburg⁴, H. Chong⁵, C. Zerbe⁴, A. Jongco³, J. Gillis⁶; ¹Cold Spring Harbor Lab., Cold Spring Harbor, NY, ²GARVAN Inst. OF Med. RESEARCH, Darlinghurst, Australia, ³Northwell Hlth.Long Island Jewish Med. Ctr., Great Neck, NY, ⁴NIAID/NIH, Bethesda, MD, ⁵Univ. of Pittsburgh Med. Ctr., Pittsburgh, PA, ⁶Donnelly Ctr. for Cellular + Biomolecular Res., Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

X-linked chronic granulomatous disease (X-CGD) is a rare primary immunodeficiency affecting about 1 in 200,000 live births associated with specific bacterial and fungal infections owing to the inability of phagocytes to generate reactive oxygen species (esp. superoxides) due to inactivating mutations in cytochrome b-245, beta chain (*CYBB*), a key component of the NADPH oxidase complex. Patients and carriers are typically also subject to a higher burden of autoimmune morbidity indicating potential systemic consequences of the loss of *CYBB*. Constituting the largest sequenced cohort in X-CGD, we recruited 14 probands, 10 carriers, and 15 age- and sex-matched controls (onto IRB-approved protocol NCT00404560), and subjected total PBMCs and isolated monocytes to scRNA sequencing. Cluster-specific differential expression analysis revealed a predominant type I interferon-driven signal across all identified cell-type clusters. This was complemented by an enrichment of genes involved in antigen processing and presentation through the MHC class I pathway, important for presenting intracellular antigens to immunocytes, and crucial in maintaining self-tolerance as well as immunity from viruses and several intracellular bacteria. Interferon signaling and dysregulated antigen presentation have been previously implicated in several autoimmune processes, and might play an important role in driving the predisposition to autoimmunity seen in X-CGD probands and carriers. We were also able to define proband-derived, cluster-specific disease signatures which could differentiate carriers from controls in a train-test setting (average AUROC=0.79). Disease penetrance is linked with X-inactivation skew in carrier females, however, expression of these disease signatures appeared to be uncorrelated with residual oxidative burst in female carriers, pointing towards a non-cell autonomous origin. Expression of these signatures also appeared to be highly linked in control human data, as well as across several chordate species using 1-1 orthologs, suggesting the disruption of a deeply conserved mechanism crucial in immune homeostasis in X-CGD. Genes upstream of a canonical type I interferon response such as *IRF5* and *MNDA* were also observed to be highly co-expressed with *CYBB*, and might constitute novel drug targets in this disease.

Omics Technologies Posters - Wednesday

PB2935. Circulating proteomic signatures of pulmonary function

Authors:

M. Lee, S. J. London; NIH/NIEHS, Research Triangle Park, NC

Abstract Body:

Pulmonary function is the basis of COPD diagnosis and respiratory health. Many genes have been associated with pulmonary function from genomic and epigenomic studies; proteomic profiles have been less examined. We analyzed plasma proteomic profiles in relation to pulmonary function in 1828 participants in the Agricultural Lung Health Study, a case-control study of asthma nested within a farming cohort. Forced expiratory volume in one second (FEV₁) and forced vital capacity (FVC) were measured using spirometry, and 4979 human protein analytes were measured with the Slow Off-rate Modified Aptamer (SOMA)Scan platform. We evaluated associations between plasma protein abundances and pulmonary function traits: FEV₁, FVC, and FEV₁/FVC. Using robust linear regression, we modeled pulmonary function as the predictor of natural log-transformed analyte levels and adjusted for potential confounding factors. Seventy-five analytes were differentially abundant in relation to one or more pulmonary function traits after Bonferroni multiple-testing correction ($P < 1.0e-05$). A network constructed using the significant analytes using the STRING database had more protein-protein interactions than expected for a random set from all analytes examined ($P = 1.8e-08$). Follow-up analyses will include replication in an independent population and functional interpretation. To our knowledge, this is the first large-scale study of plasma proteins in relation to pulmonary function. Findings from this study could aid development of novel biomarkers of respiratory health.

Omics Technologies Posters - Thursday

PB2936*. Clinical genome and RNA sequencing with a novel, cheaper sequencing-by-synthesis technology

Authors:

P. Liu¹, J. Sinson¹, C. Yeh¹, K. Schulze¹, C. Eng¹, K. Worley¹, S. Pollock², I. Soifer², D. Lipson², Undiagnosed Diseases Network, B. Lee¹; ¹Baylor Coll. of Med., Houston, TX, ²Ultima Genomics, Newark, CA

Abstract Body:

Clinical genome sequencing has been shown to be effective in revealing molecular diagnoses for patients with Mendelian disorders. However, its clinical uptake is still lagging due to the high sequencing cost of most current technologies.

We have evaluated the analytical validity of a novel sequencing technology (Ultima Genomics) employing a mostly-natural sequencing-by-synthesis chemistry. The Ultima platform significantly reduces the cost of the sequencing consumable, and thus has the potential to overcome the high sequencing cost barrier.

Single nucleotide and insertion/deletion variants (SNVs, indels) were benchmarked in the sample NA12878. High-quality variant calling performance was observed across high-confidence non-repetitive genomic regions with accuracy for SNVs > 99% and error rate for indels < 0.5%. We found that the limitation of the technology is in calling indels in homopolymers >10nt, but anticipated this caveat to have minimal impact on the clinical diagnostic yield, as regions of long homopolymers tend to be intronic or non-genic.

We obtained high concordance results for >20 clinically confirmed pathogenic SNVs or indels both in the nuclear and the mitochondrial genome, including challenging regions where mapping is complicated by pseudogenes. We tested >25 clinically validated pathogenic copy number variants (1.6 kb-4.5 Mb in span), and 2 balanced structural variants. Each event was detected by a read-depth and/or a split-read algorithm.

Furthermore, we piloted a study in a different diagnostic context, bulk RNA-sequencing to clarify abnormal splicing consequence of disease-causing variants. By increasing the sequencing yield 20-30 fold (approximately >1 billion), i.e., DeepRNAseq over the standard 30-50 million read data output, challenging illegitimate splicing events in genes with low or promiscuous expression from clinically accessible tissues can be readily detected. We then performed RNA-sequencing at >100 million reads on multiple patient specimens with previously identified diagnostic splicing or expression abnormalities. The expected molecular findings were identified with similar or improved transcriptome analysis workflows that we previously implemented on Illumina data. Importantly, DeepRNAseq has the potential of increasing diagnostic yield in combination with WGS in undiagnosed disease cases by providing function correlation of genomic variants from accessible clinical tissues.

These data highlight the potential impact of low-cost sequencing technology on improving clinical diagnostic yields, increasing research discovery, and potentially, reducing disparity to genomic testing.

Omics Technologies Posters - Wednesday

PB2937*. Clinical use of the amniotic fluid cells transcriptome in deciphering Mendelian disease

Authors:

H-Y. Chung¹, M. Lee¹, A. S. Y. Kwong¹, M. M. C. Chui¹, J. F. T. Chau¹, C. C. Y. Mak¹, S. L. K. Au¹, H. Lo¹, K. Y. K. Chan¹, V. A. Yepez², J. Gagneur², A. S. Y. Kan¹; ¹The Univ. of Hong Kong, Hong Kong, China, ²Technical Univ. of Munich, Garching bei München, Germany

Abstract Body:

RNA sequencing (RNA-seq) is an emerging technology that is becoming increasingly useful to facilitate the detection and diagnosis of genetic disorders because it provides functional validation to support the interpretation of variants of uncertain significance (VUS). Blood, skin fibroblasts, and muscle are the most commonly used clinically accessible tissues for RNA-seq (Kremer et al., 2017; Fresard et al., 2019 and Vicente et al., 2022); however, the use of amniotic fluid (AF) cells, a prenatally clinically accessible tissue, has yet to be explored in RNA-seq.

Here, we examined the expression of clinically relevant genes in AF cells compared with blood and skin fibroblasts by collecting AF cells from 52 pregnant women to perform RNA-seq. Furthermore, we evaluated an RNA-seq analytical pipeline, DROP, in three prenatal cases with splice-region variants in *CHD7*, *MYRF*, and *COL1A2*.

We showed that AF cells have a similar expression profile to skin fibroblasts as they have a similar embryonic origin. The number of well-expressed genes in AF cells was comparable to that in skin fibroblasts and much higher than that in whole blood for 2,020 clinically relevant genes. Similar patterns were observed when analyzing their expression profiles in 11 disease categories.

Using AF cells RNA-seq with DROP, we successfully detected the disease-causing genes *CHD7*, *MYRF*, and *COL1A2* as splicing outliers. We demonstrated that AF cells RNA-seq is feasible and could be beneficial in prenatal diagnosis as transcriptomic data helped elucidate the molecular mechanisms of the candidate variants and upgraded the pathogenicity of VUS in *COL1A2*. It could also be a logical choice for postnatal patients with advantages over fibroblasts and whole blood as it avoids further invasive procedures. Moreover, the use of AF cells rather than blood in RNA-seq maximizes the number of testable genes. This study may lead a new era of non-invasive diagnostics for all patients whose AF is readily available.

Omics Technologies Posters - Thursday

PB2938*. Closing the gap: Solving complex medically relevant regions of the human Genome

Authors:

F. Sedlazeck¹, M. Mahmoud¹, H. Corbitt², J. Harting³, X. Chen³, S. Jhangiani¹, H. Doddapaneni¹, Q. Meng¹, T. Hon³, I. McLaughlin³, Y-C. Tsai³, T. Han², G. Metcalf¹, R. Gibbs¹, M. Eberle³, S. Kingan³, C. Morrison⁴, G. Henno³, D. Muzny¹; ¹Baylor Coll. of Med., Houston, TX, ²Twist BioSci., South San Francisco, CA, ³Pacific BioSci.s, Menlo Park, CA, ⁴Twist BioSci., Oceanside, CA

Abstract Body:

This year marks the release of an improved reference (T2T-CHM13), but despite this effort, many medically relevant regions within the human genome are still lacking detailed characterization with standard short-read methods due to their repetitiveness (*SMN1 and SMN2*), interaction with pseudogenes (*GBA*), and complex nature (*LPA*). Furthermore, several genes show complex mutational patterns that are challenging, but important to resolve. To overcome this at scale and cost-efficiently, we partnered with TWIST Bioscience and Pacific Biosciences to design and implement a long-read panel to capture the entire length of the 396 genes. Furthermore, we developed new computational approaches to optimize analytics to fully characterize the 396 genes across the human genome. These genes are reported to impact a range of diseases including neuropathies (~15%) to immunodeficiencies (~7%) down to vision-related diseases (1%) and also include cancer driver genes (e.g. *PTEN*). Over the past months, we have optimized this approach using Laboratory optimizations to improve the on-target rate (60%) and also evaluated PCR free methodologies to reduce GC bias. We will present results showing the ability to multiplex samples in a single HiFi SMRTcell, optimizing the cost and efficiency. In addition to the lab optimizations, we designed novel analytical methodologies to identify all types of variants. Most of these genes can be characterized and provide single nucleotide variants, small insertions and deletions, complex structural variation, and copy number alterations together with phasing information per gene, enabling the most comprehensive insight into the gene diversity. Additionally, our pipeline utilizes assembly approaches and specialized per gene methods (e.g. graph genomes) for certain complex genes such as *Cyp2d6*, *LPA*, and *SMN1&2* to detangle its complex repetitive nature and provide key insights into the e.g. mutations across the KIV-2 repeat unit. We demonstrate the utility of this panel across HG002, five control samples and have started to apply the panel across several research cohorts including cardiovascular and neurological phenotypes. Using these samples, we demonstrate highly accurate variant identification. Together using laboratory optimization and analytical development we present a cost-efficient method to further investigate medically relevant genes missing from existing short read datasets. Our approach is easily scalable, allowing it to be run with every short-read whole-genome data to improve key insights in the so-far dark regions of the human genome that play a critical role across a variety of diseases.

Omics Technologies Posters - Wednesday

PB2939. Combination of annotation enrichment approaches for characterization of blood-based RNA sequencing biomarkers.

Authors:

S. Listopad, T. M. Norden-Krichmar; Univ. of California, Irvine, Irvine, CA

Abstract Body:

Background: Gene set enrichment analysis is commonly performed on RNA sequencing (RNA-seq) biomarkers in order to glean biological insight. However, all annotation enrichment tools are inherently biased by the contents of their knowledgebases and statistical methods used. Our goal in this study was to apply four different enrichment analysis tools to our data in an effort to extract the maximum amount of information from our biomarker gene set. **Methods:** This study was conducted on RNA sequencing data from peripheral blood mononuclear cells (PBMCs) from participants, consented for an IRB-approved study in the Southern California Alcoholic Hepatitis Consortium (SCAHC), with alcohol-associated hepatitis (n=38), alcohol-associated cirrhosis (n=40), non-alcohol-associated fatty liver disease (n=20), chronic hepatitis C viral infection (n=19), and healthy controls (n=20). The RNA-seq data was aligned to the human genome hg38 using the STAR software and ENSEMBL gene annotation. Distribution normalized counts were generated using Cuffnorm, and differential expression analysis was performed using Cuffdiff. We identified a set of 75 genes that was highly effective at discriminating between the four liver diseases and controls. Annotation enrichment was performed on the 75 genes using four different software tools: Enrichr, Ingenuity Pathway Analysis (IPA), Gene Set Enrichment Analysis (GSEAPreranked), and BloodGen3Module. **Results:** Application of all four tools resulted in many common pathways, such as various immune system processes, complement cascade, inflammatory response, and oxidative stress. Additionally, the top enriched tissue gene sets across the four tools were peripheral blood, granulocytes, and neutrophils. However, there were also some differences in the results. For example, chemical and iron homeostasis were among the top hits returned by GSEAPreranked and IPA, but were absent from the Enrichr and BloodGen3Module top hits. GSEAPreranked returned very few statistically significant hits, while IPA, Enrichr, and BloodGen3Module resulted in up to hundreds of significant hits. **Conclusion:** Each of the tools tested had some advantages and disadvantages that primarily had to do with ease of use, availability of in-built visualizations, and quality of knowledgebase. However, when used in combination, we found that these tools provided confirmation of similar enrichment annotations, as well as complementary information and visualizations that added novel dimensions to the biological interpretation of our data. From these results, we encourage the use of multiple annotation enrichment tools for gene set analysis.

Omics Technologies Posters - Thursday

PB2940. Combination of low-coverage whole genome and deep-coverage exome sequencing is the cost-effective way to drive large-scale genetics studies forward.

Authors:

M. Yu^{1,2}, M. DeFelice¹, M. Daly^{2,1,3}, B. Neale^{2,1}, R. Xavier^{1,4,5}, D. Howrigan^{2,1}, A. Martin^{2,1}, K. Yuan^{2,1}, H. Huang^{2,1}; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, ³Inst. for Molecular Med. Finland, Helsinki Inst. of Life Sci., Univ. of Helsinki, Helsinki, Finland, ⁴Dept. of Molecular Biology, Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, ⁵Ctr. for Computational and Integrative Biology, Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA

Abstract Body:

Over the past decade, large-scale genetic studies have been very successful in finding genetic loci associated with human complex traits and disorders. Given the high cost of deep whole genome sequencing, a commonly used strategy for large-scale genetic studies is to use genotyping array to capture genome-wide common variants and deep whole exome sequencing for rare coding variants. This strategy, however, complicates studies as the two technologies involves two independent pipelines with different equipment, experimental protocols, reagents, materials and technicians, increasing the cost of data generation. In addition, there has not been a widely accepted solution in harmonizing the two sets of genomic data generated using this strategy.

Here we introduce a new sequencing method that combines low-coverage whole genome sequencing with deep whole exome sequencing. Using PCR free library spiked into TWIST exome capture, sequencing of the genome-wide content and the deep exome content can be done in one experiment and the reads will be processed in a harmonized manner, reducing cost and batch effect.

We tested this new technology on 68 subjects diagnosed with the inflammatory bowel diseases (IBD), an autoimmune disorder implicated by both common and rare variants. We performed genotype imputation on low-coverage sequencing data using GLIMPSE, a recently published method tailored to this utility, and haplotype reference panel from Haplotype Reference Consortium. With 4x coverage, 99.64% variants in the reference panel with minor allele frequency(MAF) > 5% can be imputed with quality score > 0.8, and 98.69% for MAF > 1% variants. With 30x coverage in the exome, we captured 29919 coding variants that have MAF between 1% and 0.01% in the gnomAD callset, 11.56% of the total variants in the same MAF range captured in gnomAD callset.

We searched for the 45 established IBD causal variants (Huang et al., Nature 2017) in the sequencing data. These variants cover the full range of allele frequency (0.2% to 49.2%), located in both coding and non-coding genome, and include indels and splicing variants that can be challenging to capture by exome sequencing. Despite these challenges, our sequencing data covered all the sites for these variants.

In conclusion, we showed using a sequencing pilot on 68 IBD cases that our new method combining low-coverage whole genome and deep-coverage exome sequencing is a cost-effective approach to drive large-scale genetic studies for human complex disorders and traits across allele frequency spectrum.

Omics Technologies Posters - Wednesday

PB2941. Combined multiple structural variant algorithms achieved the highest recall rate in GIAB large deletion callsets

Authors:

P-X. Chen¹, Y-H. Tseng¹, P-M. Chien¹, S-k. lai², T-K. Hung¹, P-L. Chen^{1,2,3,4}, C-Y. Chen^{2,5}, J-J. Hsu¹; ¹Graduate Inst. of Med. Genomics and Proteomics, Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan, ²Genome and Systems Biology Degree Program, Natl. Taiwan Univ. and Academia Sinica, Taipei, Taiwan, ³Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, ⁴Graduate Inst. of Clinical Med., Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan, ⁵Dept. of Biomechatronics Engineering, Natl. Taiwan Univ., Taipei, Taiwan

Abstract Body:

Structural variants (SVs) define as genomic changes > 50 bp, and SV calling tools with different algorithms have been proposed. Genome in A Bottle (GIAB) consortium has released a benchmarking workflow and large insertion/deletion variants (NIST_SV0.6) as the benchmark truth set. We adopted SV benchmarking pipeline and analyzed the truth set composition, including sequencing technologies, calling tools and variant sizes. We utilized HG002 35x WGS data to test different SV callers, including Dragen, Manta, Lumpy, Delly, GATKHC and the combination of their output. We evaluated caller performance by different SV types and variant sizes in each subset. We found Dragen had the highest recall rate (74%) in deletion. Delly and Lumpy were appropriate for detecting large size of deletions. In addition, we observed the combination of multiple callers' output could increase the overall recall rate. For large insertion, we found poor performance for short read data, with a higher false-negative rate (40%) than deletion. Even combination strategy cannot overcome it due to the lack of SV caller suitable for covering large size of insertions. In conclusion, we utilized the GIAB benchmarking pipeline to evaluate SV caller performance, especially in short-read data. Moreover, it could also designate the best combination of SV callers for clinical testing and population database.

Omics Technologies Posters - Thursday

PB2942. Community detection analysis in multilayer COVID-19 patient similarity networks

Authors:

P. Sliwa, COMBAT Consortium, H. A. Harrington, G. Reinert, J. C. Knight; Univ. of Oxford, Oxford, United Kingdom

Abstract Body:

Advancement of biological assays has enabled generation of datasets containing multiple modalities, such as single cell RNA-seq data and proteomics data, per patient for large cohorts. Integrating the information from different modalities is an open research challenge; a popular method is Similarity Network Fusion. Here we propose an alternative network-science based approach, where we construct patient similarity networks in a principled fashion, one network per modality, and then combine these similarity networks into a multilayer network. On this network, communities are detected using the Leiden algorithm. We apply this approach to a recent comprehensive multimodal dataset, the Covid-19 Multi-omics Blood Atlas (COMBAT), containing among others information on COVID-19 and sepsis patients and healthy volunteers across modalities spanning proteomics, bulk and single cell transcriptomics and mass cytometry, and explore the resulting communities for correlations with important clinical variables and activity of drug-relevant and immune system related molecular pathways. The analysis reveals a signal which differentiates between healthy, COVID-19, and sepsis patients. We show that using multilayer networks we can refine classifications made using one modality. Our pipeline can also be applied to other similar datasets and may help to better understand multimodal molecular health data.

Omics Technologies Posters - Wednesday

PB2943. Comparison of exome performance in an inherited cancer context

Authors:

C. Johnson, A. Ramesh, K. Shortt, K. Eyring, S. Raghunath; Intermountain Hlth.care, St George, UT

Abstract Body:

Background: Whole-exome sequencing has excellent utility in clinical genomics due to the breadth of coverage. In this setting however it is important to evaluate if the exome can detect clinically significant variants with high accuracy. In this study we compared 3 exomes (Agilent WESv7, Twist WES, and IDT Research Exome) and 2 panels (IDT Inherited Disease Panel and Agilent BrExome 1.3pM Panel) against cell lines with NIST high-confidence datasets and clinically relevant variants. These were evaluated for sensitivity (PPA) and positive predictive value (PPV), as well as coverage and other quality control (QC) metrics.

Methods: We sequenced 4 cells lines with NIST high confidence variants datasets (NA12878, NA24143, NA24631, and NA24694) to evaluate broad PPA and PPV of each panel. We also sequenced one engineered control with 22 SNVs and Indels relevant in inherited cancer, and 3 CNVs across 3 cell lines (NA02718, NA14626 and NA10401) to evaluate how the assays would perform on clinically relevant variants. BED files of regions common to all exomes or panels were generated for this analysis. QC was performed using Fastp and Picard. SNV and Indel calling was done using GATK. CNV calling was done using CNV Kit, ExomeDepth, and DECON.

Results: The two exomes which use UMIs (Agilent and IDT) had a similar mean depth of coverage, but the IDT exome had a higher % target bases covered at 50x than Agilent (80.1% vs 69.2%). This was true for the panel designs as well (89.4% vs 73.2%). For NIST cell lines, all 3 exomes had similar sensitivity, with IDT having higher mean PPA (99.7%), followed by Twist (99.5%) and Agilent (99.3%). Twist, with no UMIs, had a lower PPV than the other exomes. The Twist exome and Agilent panel had the highest PPA for tested clinically significant SNVs and indels (95.5%), followed by the IDT exome (90.9%). The Agilent exome had higher depth of coverage than IDT for these variants. All tested CNVs were detected using Decon and Exome Depth, and 2/3 using CNVKit.

Conclusions: This study suggests that the IDT exome had better coverage uniformity and higher overall PPA for exomes, while the Agilent exome had reads concentrated in clinically significant regions and higher PPA in panels. While the Twist exome had high PPA, it also had the lowest PPV.

Omics Technologies Posters - Thursday

PB2944. Comparison of genotyping arrays and low-pass sequencing for predictive genomics research.

Authors:

J. Gollub, A. Mittal, K. Schweighofer, J. Schmidt; Thermo Fisher Scientific, Santa Clara, CA

Abstract Body:

Predictive genomics is a powerful capability to help predict disease risk and understand drug response. Key considerations include high genome-wide coverage, high accuracy of genotyping calls and low cost. Genotyping by array has proven to be an excellent technology to support predictive genomics research, due to its overall high accuracy and favorable cost structure. Recently, low-pass whole genome sequencing (lpWGS) at a target coverage of 1x or less, combined with imputation, has emerged as a possible alternative technology. We compare the relative performance of lpWGS with respect to genotyping arrays. We focused on evaluating genome-wide imputation performance as well as the performance of calling specific variants of interest which had not previously been extensively investigated. We evaluated genome-wide imputation accuracy on 60 samples of European and African ancestries from the 1000 Genomes project. For microarray data we assessed imputation performance using the pan-ethnic Applied Biosystems™ Axiom™ PMD array. For lpWGS we used publicly available data at 0.5x and 1x coverage from Gencove [<https://gencove-sbir.s3.amazonaws.com/index.html>]. For the specific variant performance analysis, we used 44 samples of diverse ancestries with known pathogenic variants associated with severe Mendelian conditions or pharmacogenomic relevance. The samples were sequenced at a target coverage of 0.5x and 1x lpWGS and genotyped on CarrierScan 1S and PMD array. The PMD array showed an imputation accuracy comparable to 1x lpWGS for the European cohort, and between 0.5x and 1x lpWGS for the African cohort. For the 38 pathogenic variants associated with severe Mendelian conditions, the array provided 100% non-reference allele concordance (NRC) across all MAF ranges in 144 carriers and 25 affected individuals. In contrast, the NRC using lpWGS ranged between 95.1% (0.5x) and 97.6% (1x) for the common variants, between 87.5% (0.5x) and 93.8% (1x) for low frequency variants and between 73.3% (0.5x) and 96.7% (1x) for the rare variants. For the pharmacogenomic variants (N=3191), PMD array showed an overall NRC of 99.6±4.9 and lpWGS showed NRC of 96.1±12.8 and 97.3±10.4 for 0.5x and 1x respectively. In conclusion, our results show that low and rare frequency variants and the variants in complex regions cannot be imputed with high accuracy using lpWGS at a depth of 1x or less. On the other hand, these variants can be directly genotyped on the array, making microarrays well suited for predictive genomics research applications that rely on high genome wide imputation accuracy as well as highly accurate genotyping of specific variants of interest.

Omics Technologies Posters - Wednesday

PB2945. Competing sampling biases balance concordance of single-cell and nucleus RNA-seq

Authors:

J. Chamberlin¹, G. Marth², A. Quinlan³; ¹Univ. of Utah, Salt Lake City, UT, ²Univ. OF UTAH, SALT LAKE CITY, UT, ³Univ of Utah, Salt Lake City, UT

Abstract Body:

Single-nucleus RNA-seq (snucRNA-seq) is typically used as an alternative to single-cell RNA-seq (scRNA-seq) when dissociation of whole cells is problematic. Nuclei contain less RNA than cells, are depleted of mRNA (inferred from exon-mapped reads), and are enriched for pre-mRNA (inferred from intron-mapped reads). mRNA and pre-mRNA are subject to distinct sampling biases; however, the standard pre-processing workflow does not distinguish the origin of measured transcripts, which obfuscates these biases and their variable impact across cell types. In particular, mRNAs exhibit gene-specific localization within the cell, while pre-mRNAs are typically captured by a gene length-associated process called internal priming. Because pre-mRNA contributes more of the data in nuclei, overall gene abundance estimates show stronger length bias than in cells. Here, we re-analyze public scRNA-seq and snucRNA-seq data from mouse cortex in order to dissect the relative role of these biases across cell types, with an emphasis on a recently-published normalization method intended to increase comparability of the two assays by accounting for discrepant gene length (10.1101/gr.275509.121).

We show that pre-mRNA abundances are consistently more similar than mRNA abundances across all cell types. This is unsurprising given that all pre-mRNAs are inside the cell, but not all mRNAs are inside the nucleus. However, we also find that the proportion of pre-mRNA captured in nuclei is not predictive of the proportion in cells of the same type, such that the magnitude of inter-assay gene length bias is variable and unpredictable. These properties are important to the proposed normalization scheme, which attempts to remove gene length bias by scaling all pre-mRNA derived transcript counts by a factor of gene length, indirectly increasing the relative contribution of mRNA sampling. We observe adverse performance in nearly every cell type, and that performance is worst where inter-assay gene length bias is smallest.

These results are not inconsistent with those reported in the original study system, where the relative difference in pre-mRNA content between cells and nuclei was considerably higher than any cortex cell type. However, they suggest that normalization of raw data is only appropriate when pre-mRNA bias is stronger than mRNA bias, inference of which is only possible when matched samples are available. We argue that data should be pre-processed in a manner conducive to assessing these sampling biases and suggest that post hoc normalization is more appropriate in typical experiments.

Omics Technologies Posters - Thursday

PB2946. Comprehensive characterization of structural variants in human brain genomes using single flow cell nanopore sequencing protocol

Authors:

M. Kolmogorov¹, K. Billingsley², M. Mastoras³, M. Meredith³, R. Lorig-Roach³, P. Jerez¹, L. Malik¹, K. Shafin⁴, T. Pesout³, P. Carnevali⁵, A. Rhie⁶, M. Jain⁷, S. Scholz⁶, B. Traynor⁸, W. Timp⁹, A. Phillippy¹⁰, F. Sedlazeck¹¹, C. Blauwendraat⁶, B. Paten³, R. Karra¹²; ¹NIH, Bethesda, MD, ²NIH, Bethesda, MD, ³Univ. of California, Santa Cruz, Santa Cruz, CA, ⁴Google LLC, Mountain View, CA, ⁵Chan Zuckerberg Initiative, Redwood City, CA, ⁶NIH, Bethesda, MD, ⁷Northeastern Univ., Boston, MA, ⁸Natl. Inst Aging, Bethesda, MD, ⁹Johns Hopkins Univ., Baltimore, MD, ¹⁰NIH/NHGRI, Bethesda, MD, ¹¹Baylor Coll. Med., Houston, TX, ¹²Lab. of Neurogenetics, NIA, NIH, Silver Spring, MD

Abstract Body:

Many recent studies highlighted the improved capability of long-read sequencing to detect structural variation in the human genome. Long read sequencing was also recently used to produce the first complete assembly of the human genome by the Telomere-to-Telomere consortium. Further, Human Pangenome Reference Consortium has recently released 30 nearly-complete haplotype-resolved human genomes from diverse backgrounds sequenced using multiple long-read technologies.

Despite the recent advances, the cost of generating long-read sequencing remains a barrier for large-scale association studies. In this work, we developed a specialized DNA preparation protocol for Oxford Nanopore PromethION to optimize the yield of a single flow cell. We combined it with a novel computational pipeline to produce haplotype-resolved de novo assemblies and characterize small and structural genomic variants. The developed protocol is currently being used to sequence 4000+ human brain samples at the Center of Alzheimer and Related Dementias at the NIH.

Using our DNA preparation protocol we produced Oxford Nanopore sequencing data for three cell lines (HG002, HG02723, HG00733) and 15 postmortem brain tissue samples from the National Institute of Aging collection. Each genome was sequenced using one flow cell, on average yielding 142 Gb of reads with average read N50 30 kb.

We performed reference-based small variant calls using PEPPER-Margin-DeepVariant, and compared the performance against Illumina and HiFi-based variant calls. On the HG002 genome within the GIAB confidence intervals, our SNP calls outperformed Illumina calls (F1-score 0.9976 vs 0.9965). HiFi calls slightly improved over ONT-based calls (F1-score 0.9986 vs 0.9976). Two other cell line datasets showed similar small variant detection performance. In difficult-to-map regions, the improvement of ONT-based SNP calls over Illumina becomes more noticeable: MHC F1-score 0.9946 vs 0.9732, SegDup F1-score 0.9799 vs 0.9512.

Next we used Shasta to assemble each genome into haploid contigs. To produce haplotype-resolved assemblies, we developed a new pipeline called Hapdup that converts a haploid assembly into a pair of locally phased haplotypes. The resulting assemblies had mean NG50 ~20 Mb with base-level quality ~QV35.

Using the generated assemblies, we produced structural variant (SV) calls via reference alignment. On HG002 genome, Shasta/Hapdup SV calls had slightly better concordance with GIAB HG002 callset (F1 score = 0.9555), compared to Sniffles2 (F1-score = 0.9523). Hifiasm assemblies produced using HiFi and parental Illumina data only slightly improved on both approaches (0.9628).

Omics Technologies Posters - Wednesday

PB2947. Comprehensive detection of trait-associated structural variations using short read sequencing data.

Authors:

C. Terao¹, S. Kosugi¹, Y. Kamatani², K. Tomizuka¹, Y. Momozawa¹, T. Morisaki², The Biobank Japan Project; ¹RIKEN IMS, Yokohama, Japan, ²The Univ. of Tokyo, Minato-ku, Tokyo, Japan

Abstract Body:

Genomic structural variation (SV) affects genetic and phenotypic characteristics in diverse organisms, but the lack of reliable methods to detect SV has hindered genetic analysis. We developed a computational algorithm (MOPline) that iteratively merges optimized overlapping calls from multiple algorithms in each SV category to increase precision and genotypes SV sites including reference alleles to increase recall using short-read whole genome sequencing (WGS) data. The genotyping method, called SMC, restores missing variants and increases high confidence variants by ~43%. Evaluation with multiple real datasets demonstrated that the SV detection and genotyping accuracy of MOPline outperforms existing SV detection and genotyping algorithms. Validation using NA12878 PacBio CCS long-read alignment data showed 95-98% precision for any SV types except for INVs. Using 414 high coverage 1000 Genomes WGS data and 3,258 BioBank Japan (BBJ) WGS data, MOPline stably detected ~16,000 SVs per individual, which is over 2-fold higher than previous large-scale projects while exhibiting a comparable level of the statistical quality metrics. Principal component analysis and allele frequency comparison using the detected SVs and SNVs showed that the genetic architecture in multiple populations observed for the SVs was consistent with that observed for the SNVs. The overall AF concordance of common SVs matched between the BBJ-SVs and gnomAD-EAS-SVs was 0.94 (Pearson correlation coefficient). BBJ-SVs included many rare coding SVs that disrupt known and potential novel disease risk genes and common SVs in high linkage disequilibrium with published disease genome-wide association study (GWAS) variants. Using the BBJ-SVs as a reference panel, we imputed SVs from 181,622 Japanese individuals for 42 diseases and 60 quantitative traits. SVs had lower imputation quality than SNVs in the lower AF range, but almost 90% of SVs were imputed with > 85% overall precision. GWAS with the imputed SVs revealed 41 top-ranked or nearly top-ranked genome-wide significant SVs, including 8 exonic SVs with 5 novel associations (height-associated 1.9 Kb DEL in *MUC22*, HbA1c- and MCHC-associated 119 Kb DUP in *GYPB/GYPB*, ALP-associated 24 Kb DEL in *FUT2*, and A/G-associated 23 Kb DEL in *RP11-219A15.2*). The trait-associated DELs and INVs were enriched for DNase I transcription factor footprints and mobile element insertions, respectively. These results demonstrate that short-read WGS data can be used to identify rare and common SVs associated with a variety of traits. The methodology and the SV datasets created in this study provide a valuable resource for SV analyses in diverse research areas.

Omics Technologies Posters - Thursday

PB2948. Comprehensive discovery of CRISPR-targeted sequences in the human gut metagenome

Authors:

R. Sugimoto, L. Nishimura, I. Inoue; Natl. Inst. of Genetics, Mishima, Japan

Abstract Body:

Bacteriophages, or phages, are viruses that infect bacteria. There are about 10^9 viruses per gram of human stool. Through lytic or lysogenic interaction, these viruses involve the bacteria composition and diversity in the human intestine. The intestinal bacteria play important roles in human health. On the other hand, the implication of bacteriophages to human health remains unknown due to the little knowledge about them. To understand viral evolution and its interactions with the hosts, viral genomes provide essential information. Metagenomics, a method that sequences all DNA extracted from a given sample, became a popular approach to collecting unculturable viral genomic sequences. However, there are several challenges to extracting viral genomes from metagenomic sequences. Firstly, there is no marker gene for viruses such as the 16S rRNA which is commonly used for bacteria studies. Secondly, the majority of viral genes show little or no homology to known genes. Therefore, we need a method to detect viral sequences from the metagenome without relying on reference genes. For this, we developed a virus detection method that utilizes CRISPR. CRISPR is a prokaryotic adaptive immunological memory that encodes the fragments of viral genomes previously infected the cell. Using these viral genome fragments, we successfully discovered about 11 thousand complete viral genomes from human gut metagenome datasets. The discovered genomes include 4 novel Jumbo-phages and many small novel phages. Furthermore, we resolved the hosts of nearly 70% of discovered viruses. This host information could be particularly useful for some clinical applications such as phage therapy. We continuously apply our method to further metagenome sequences from a variety of environments and will provide a comprehensive viral genome database which will be basal information to resolve the implication of phages to human health.

Omics Technologies Posters - Wednesday

PB2949. Comprehensive Structural Variant Detection: From Population to Mosaic level

Authors:

L. Paulin¹, M. Smolka¹, C. M. Grochowski², M. Mahmoud^{1,2}, S. Behera¹, M. Gandhi³, K. Hong⁴, D. Pehlivan^{2,5}, S. Scholz^{6,7}, C. M. B. Carvalho^{3,2}, C. Proukakis^{8,9}, F. Sedlazeck^{1,10}; ¹Baylor Coll. of Med. Human Genome Sequencing Ctr., Houston, TX, ²Dept. of Molecular and Human Genetics Baylor Coll. of Med., Houston, TX, ³Pacific Northwest Res. Inst., Seattle, WA, ⁴Bionano Genomics, San Diego, CA, ⁵Div. of Neurology and Dev.al NeuroSci., Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, ⁶NIH, Bethesda, MD, ⁷Johns Hopkins Univ. Med. Ctr., Baltimore, MD, ⁸Univ. Coll. London, London, England, United Kingdom, ⁹Aligning Sci. Across Parkinson's (ASAP) Collaborative Res. Network, Chevy Chase, MD, ¹⁰Dept. of Computer Sci., Rice Univ., Houston, TX

Abstract Body:

Long read Structural Variation calling remains a challenging, but a highly accurate way to identify simple and complex genomic alterations. To address this challenge, we developed Sniffles2, a successor to SV-detection method Sniffles. Sniffles2 increases both accuracy (e.g., genotype and insertion sequence), speed (higher parallelism), and has more advanced functionality to further promote new insights into the organism or diseases. Sniffles2 outperforms its closest competitors (e.g., SVIM and CuteSV) in the well-established GIAB benchmark, both in terms of accuracy and speed. Additionally, Sniffles2 also outperforms its competitors in the more challenging SV data set across 386 medically relevant challenging genes, where Sniffles2 is 12% more accurate than the second-fastest SV caller and almost 10 times faster than the second most accurate. Furthermore, Sniffles2 solves the problem of population level SV calling by producing fully genotyped VCF files orders of magnitude faster than current methods. Sniffles2 population mode introduces a gVCF file concept to then, efficiently and accurately identify SV across a collection of samples from familial trios up to large-scale cohorts of hundreds to thousands of individuals. We used Sniffles2 population mode coupled with copy number data to determine the breakpoints of complex SVs in individuals affected with a Mendelian disease that often is caused by extreme complex alleles impacting the *MECP2* gene at the Xq28 locus. Sniffles2 is able to detect a complex allele which consists of a duplication-normal-duplication (DUP-NML-INV/DUP) with breakpoints spanning segmental duplications. Moreover, we also used Sniffles2 population mode to improve tumor-normal comparison and detected multiple candidates including deletion of the two exons of the *PTEN* gene in the COLO892 cancer cell line. Sniffles2 further enables the detection of somatic SV across bulk long-read data. We used Sniffles2 somatic mode on the genomic data from the brain of a patient with multiple system atrophy, a rare sporadic neurodegenerative condition similar to Parkinson's disease, with 55x WGS ONT coverage. We were able to identify multiple somatic SVs for this brain region. For example, a repeat contraction in *KCNIP4* an interactor of neuronal voltage gated potassium channels. We are now applying Sniffles2 across 4000 brain samples across multiple neurological diseases. Thus, overall demonstrating the utility and versatility of Sniffles2 to identify SV from mosaic to population levels and pushing the field of SV calling further.

Omics Technologies Posters - Thursday

PB2950. CRISPR Streamline: Creating a Pipeline That Evaluates Phenotype Information From Off-Target Sequence Identification

Authors:

P. Anderson¹, J. Gold¹, K. Pinnipati¹, C. Viridi¹, J. Davidson²; ¹California Polytechnic State Univ., San Luis Obispo, CA, ²Cal Poly, San Luis Obispo, San Luis Obispo, CA

Abstract Body:

CRISPR/Cas9 is a tool which is undergoing rapid growth in the fields of genetic engineering due to its ability to target edits to highly specific genomic sequences in vivo. However, suboptimal gRNA design can lead to frequent off-target splicing events at unintended genome locations which can greatly decrease CRISPR/Cas9 efficiency. Numerous bioinformatics programs exist which attempt to predict potential off-target sites within genomes, however, no such programs exist to further predict potential phenotypic effects from those off-target splices despite the value such information holds to scientists. Here we present, Crispr Streamline: a command-line program which builds on the previously established off-target prediction tool FlashFry by passing output genome locations to the ClinVar database as a means of predicting consequential phenotypic effects. This program accumulates in a web-based tool which is freely accessible for those performing CRISPR/Cas9 experiments or gRNA validation. Optimal design of such gRNA's is crucial to genome editing success and is an active area of study in bioinformatics. Previous observations reveal that CRISPR/Cas9 gRNA sequences frequently introduce double strand breaks (DSB's) unintentionally at sites outside a target gene commonly referred to as off-target sites, due to mismatched base pairing between gRNA and genomic sequences. Introduction of DSB's at off-target's compromise overall sequence specificity and genome editing efficiency for any CRISPR/Cas9 system, hence it is of great importance to be able to predict the locations and effects of off-target's associated with any gRNA sequence for optimization before in vivo experiments. Crispr Streamline is a program that utilizes a contemporary off-target identifier and genomic databases containing phenotypic annotation data to predict both the locations and phenotypic consequences of potential off-target sites in the GRCh38 human reference genome from target sequence input. Website: streamline.birg.dev

Omics Technologies Posters - Wednesday

PB2951. Cryptic splice variants contribute to the genetic etiology of congenital heart disease

Authors:

R. Lesurf¹, J. Bouwmeester¹, G. Persad¹, R. Yao¹, T. Papaz¹, G. Blue², C. Bezzina³, D. Winlaw⁴, S. Mital¹; ¹The Hosp. for Sick Children, Toronto, ON, Canada, ²The Children's Hosp. at Westmead, Westmead, Australia, ³Amsterdam Univ. Med., Amsterdam, Netherlands, ⁴Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH

Abstract Body:

Congenital heart disease (CHD) affects 1 in 100 live births and is a leading cause of neonatal deaths. Pathogenic variants in protein-coding regions and in canonical splice sites have been associated with CHD. Yet 80% CHD cases remain gene-elusive using conventional genetic testing. Recent developments into deep-learning classifiers (e.g. SpliceAI) have extended our ability to accurately identify cryptic splice variants outside of canonical splice sites, which in turn have established a role for cryptic splice variants in human diseases such as cardiomyopathy. Our goal was to investigate the role of cryptic splice variants in CHD. We performed whole genome sequencing (WGS) in 1264 CHD probands and their family members, including 839 probands with tetralogy of Fallot (TOF) and 229 probands with D-transposition of the great arteries (TGA). 1,002 patients were gene-elusive i.e., did not harbor pathogenic coding or canonical splice site variants. Within these gene-elusive probands, we searched for rare (gnomAD popmax allele frequency < 0.0001), cryptic splice variants (SpliceAI delta score > 0.65), in 158 Tier 1-3 known and candidate disease-causing CHD genes that have an autosomal dominant mode of inheritance. This identified 32 rare cryptic splice variants in 3% of TOF probands and 2% of TGA probands. Variants were observed in seven genes associated with isolated CHD (Tier 1), nine associated with syndromic CHD (Tier 2), and five candidate CHD genes (Tier 3). The effect of the putative cryptic splice variants was validated in RNA-sequencing and proteomics profiles generated from myocardium available in 101 TOF probands, further supporting their pathogenicity. These findings demonstrate that cryptic splice variants contribute to the genetic etiology of CHD, and underscore the improvement in genetic yield with WGS compared to standard clinical and exome tests.

Omics Technologies Posters - Thursday

PB2952. Custom morphology markers allow better tissue stratification to study tumor heterogeneity using NanoString® GeoMx® Cancer Transcriptome Atlas

Authors:

J. Runyon, V. Baichwal, C. Nievera; Canopy BioSci.s, St. Louis, MO

Abstract Body:

Understanding tissue heterogeneity is a key goal in oncology research and new spatial biology technologies are now available to address this. These platforms provide spatial context about cells and their interactions within a tissue or tumor sample while also producing high-plex gene expression data. The NanoString® GeoMx® Digital Spatial Profiler platform combines spatial and molecular profiling technologies and uses fluorescently labeled morphology markers to guide selection of regions of interest on tissue samples. These markers broadly target tumor and immune cells within tissue, while the probes from these regions of interest are collected and analyzed to generate gene expression profiles for each sample.

To evaluate the capability of this technology to selectively enrich specific cell types, we stained tonsil sections with CD3 and CD20 and compared transcriptional profiles and determined whether they reflected those of T and B lymphocytes. In other experiments regions of interest in tonsil sections were segmented into CD4 and CD8 positive sub-populations. Analysis of the transcriptional profiles with cell deconvolution algorithms confirmed enrichment of these T-cell sub-types. Using similar approaches, segmentation of adenosquamous lung carcinoma samples with markers specific for squamous cell carcinomas (p40) and adenocarcinoma (TTF-1) revealed significant differences in transcriptional maps of these two tumor cell types.

For all experiments, transcriptional profiles were assessed with the Cancer Transcriptome Atlas (CTA), a sequencing panel for the GeoMx platform that covers 1,800 key oncology gene targets to profile the tumor, tumor microenvironment and immune response. Analysis of this data set was performed using the GeoMx DSP Analysis Suite software. The findings, including the inferred presence or absence of specific cells based on gene expression profiles, support the notion that custom morphology markers enable better tissue stratification providing more meaningful gene expression analysis data. Studies with additional tumor types and different morphology markers are ongoing to further explore utility of this technology for analyzing gene expression in specific cells within different tumor types.

Omics Technologies Posters - Wednesday

PB2953. D-CoMEx: Differential co-expressed module extraction to identify phenotype-specific miRNA biomarkers related to diseases.

Authors:

T. Kakati^{1,2}, D. Bhattacharyya², J. Kalita³, T. Norden-Krichmar¹; ¹Univ. of California, Irvine, Irvine, CA, ²Tezpur Univ., Tezpur, India, ³Univ. of Colorado, Colorado Springs, Colorado Springs, CO

Abstract Body:

Introduction: Investigating mRNA-miRNA associations helps reveal how they regulate biological pathways and contribute to diseases. In this study, we report a novel method, D-CoMEx (Differential Co-expressed Module Extraction) to extract differentially co-expressed (DCE) modules from mRNA-miRNA expression data to ascertain interactions between miRNA and dysregulated biological pathways.

Methods: The proposed approach integrates a co-expression network (CEN) module extraction technique, THD-Module Extractor, with a differential expression analysis technique. Using THD-Module Extractor, the co-expressed modules were extracted from control and disease states of both mRNA and miRNA expression profiles of Parkinson's disease (PD) and Breast Cancer (BC) datasets. Differential co-expression analysis was performed on the mRNAs and miRNAs of the co-expressed modules extracted from the disease datasets. The differentially co-expressed mRNAs and miRNAs which were found to be dysregulated across control and disease states were then mapped to pathways using the ToppGene Suite, Panther, and MiRSystem. From the dysregulated pathways, we filtered the miRNAs using a reverse search technique in DIANA (mirPath v.3).

Results: D-CoMEx identified biologically and statistically significant dysregulated pathways from the DCE mRNA and miRNA modules of both datasets. For example, Apoptosis signaling and Dopamine receptor mediated signaling pathways were found to be significant for PD, and pathways such as T-Cell receptor and WNT signaling pathways were identified for BC. The corresponding miRNAs associated with these pathways may be potential biomarkers regulating progression of the two diseases. Several examples of the miRNA biomarkers mapped from dysregulated pathways related to PD were hsa-miR-195 and hsa-miR-153, while miRNA biomarkers mapped from pathways related to BC were hsa-miR-205 and hsa-miR-155. The proposed approach, D-CoMEx outperformed the existing methods, namely, DCGL, MODA and DiffCoEx, in terms of statistical and biological significance of the dysregulated pathways.

Conclusion: D-CoMEx identified miRNA biomarkers in dysregulated pathways mapped from differentially co-expressed modules of mRNA-miRNA expression profiles which were related to the etiology of the PD and BC diseases. This approach and analysis of dysregulated pathways focused on pathway-miRNA interactions, which may aid in further understanding these diseases.

Omics Technologies Posters - Thursday

PB2954. *De novo* variant detection with HiFi reads

Authors:

J. A. Lake, W. J. Rowell, M. A. Eberle; Pacific BioSci.s, Menlo Park, CA

Abstract Body:

Unencumbered by the process of selection, *de novo* mutations (DNMs) are more likely to be deleterious than inherited variants and account for much of the variation driving rare disease. Spontaneous germline mutations are implicated in many severe early-onset genetic diseases such as congenital malformation syndromes and neurodevelopmental disorders and postzygotic (somatic) mutations are the primary cause of cancer. Advances in genomic technologies, especially the development of whole-genome sequencing (WGS), have enabled detection of DNMs, thereby increasing discovery rates of causal variants in rare undiagnosed genetic disease studies. However, many rare disease cases with a strong presumption of a genetic cause remain unexplained despite comprehensive short-read WGS.

Compared to short reads, high-quality long reads provide a more complete assessment of global variation, improving detection of both small variants and structural variants (SVs) especially in difficult-to-map regions of the genome that are not accessible with short-reads. While detection and identification of true *de novo* variants can be important for identifying the causes of rare diseases, a high false-positive (FP) rate can also increase the burden of manually curating call sets. To appropriately leverage the strengths of HiFi sequencing in detection of DNMs, we must first establish a baseline.

To quantify the number of *de novo* single nucleotide variants (SNVs), small indels (<50 bp), and SVs, we analyzed trios using sequence data derived from either cell lines or blood where each sample was sequenced with PacBio HiFi reads to ~30x depth. Additionally, because parents are often sequenced to lower depth, we quantified the *de novo* variation identified from a high-depth (~30x) proband with two low-depth (~10x) parents.

In three trios where blood-derived DNA was sequenced, the total number of *de novo* variants detected in each child ranged from 501 to 742, including 25 to 41 SVs, 114 to 197 SNVs and 501 to 742 small indels. In two trios sequenced from cell lines, the total number of *de novo* variants detected in each child ranged from 555 to 2086, including 48 to 69 SVs, 176 to 1350 SNVs, and 310 to 688 small indels. Reducing the parental sequencing coverage from ~30x to ~10x in these two samples increased the number of variants identified as *de novo* variants by 49% and 264%. We will develop and present a more extensive analysis using larger pedigrees to further assess the FP *de novo* rate by variant type, and characterize reference regions where these putative *de novo* variants are found. These results support development of strategies for reducing FPs, especially in trios with low-depth parental sequencing.

Omics Technologies Posters - Wednesday

PB2956. Deep immune repertoire analysis of SARS-CoV-2 antibodies from a vaccinated human patient via mass-spectrometry and synthetic long reads.

Authors:

N. Castellana¹, T. Lima¹, S. Bonissone¹, A. Patel¹, R. Carson¹, R. Kelley²; ¹Abterra BioSci.s, Inc., San Diego, CA, ²Element BioSci.s, San Diego, CA

Abstract Body:

The immune system of an individual contains the capacity to produce trillions of unique antibody proteins. Identifying the antibodies selected for secretion in serum is key to understanding the human immune response to disease and crucial for future antibody therapeutic development. Mass spectrometry can be used to analyze serum antibodies that are reactive to a target, however, recovering complete antibody sequences requires a comprehensive database of candidate antibodies. Deep antibody repertoire analysis from B cell sequencing can provide this database but is currently expensive and low throughput. The combination of synthetic long read technology and high-throughput short read sequencing is a cost-efficient method to enable deep immune repertoire analysis. In this study, blood was collected from a human donor vaccinated against SARS-CoV-2. One hundred million B cells were isolated from whole blood and sorted based on B-cell receptor (BCR) reactivity to the SARS-CoV-2 receptor binding domain (RBD), with collection of both reactive and non-reactive B cells. LoopSeq synthetic long reads and the AVITI sequencing platform were used to sequence antibody variable regions of both samples. Serum antibodies from the same donor were enriched for reactivity to RBD and spike trimer, analyzed by mass spectrometry, and matched to the NGS-derived antibody repertoire using the Alicanto platform to identify antibodies critical to the SARS-CoV-2 immune response.

Omics Technologies Posters - Thursday

PB2957. Deep learning approaches for predicting virus integration sites in the host genomes

Authors:

Z. Zhao^{1,2}, H. Xu¹, J. Jia^{1,2}; ¹Univ Texas HSC Houston, Houston, TX, ²MD Anderson UTHlth. Graduate Sch. of BioMed. Sci., Houston, TX

Abstract Body:

Background: Virus infection is commonly observed in nature. Recently, SARS-CoV-2 has caused a global pandemic with over 530 million infected cases (as of June 9, 2022). Currently, the main challenges include effective detection of viruses in the host cells and predicting how they interrupt gene regulation in the host genomes. **Methods:** To predict virus integration sites (VISs) in the host genomes, we first collected and curated highly accurate VISs in the literature. Such VISs are typically discovered from next generation sequencing data followed by experimental validation. Then, we developed two deep learning methods based on a convolutional neural network (CNN) architecture with attention mechanism: DeepVISP for VIS prediction and motif discovery of oncogenic DNA viruses and DeepHTLV for VIS prediction of retroviruses. **Results:** Using our benchmark VIS datasets (77,632 VISs from 108 studies covering 15,064 target genes of five DNA viruses and four retroviruses), DeepVISP achieves high accuracy and robust performance for DNA viruses through automatically learning informative features and essential genomic positions from the DNA sequences. Our comparison with classical machine learning methods indicated an enhancement of 8.43-34.33% area under curve (AUC) values. Many promising *cis*-regulatory factors involved in virus integration were identified, such as HOXB7, IKZF1, and LHX6. DeepHTLV has similar VIS prediction when applied to human T-lymphotrophic virus 1 (HTLV-1), which has the largest VISs in the retrovirus category. Decoding the informative features captured by DeepHTLV resulted in eight representative clusters with consensus motifs for potential HTLV-1 integration. DeepHTLV revealed many interesting *cis*-regulatory elements in regulation of VISs. Literature evidence demonstrated nearly half (34) of the predicted transcription factors were involved in HTLV-1 associated diseases. **Conclusion:** We introduce robust deep learning models, DeepVISP and DeepHTLV, for rapid and accurate VIS prediction from primary sequence information. DeepVISP and DeepHTLV can elucidate the *cis*-regulatory landscapes of VISs in both oncogenic DNA viruses and retroviruses, respectively.

Omics Technologies Posters - Wednesday

PB2958. Deep plasma protein profiling of Alzheimer's subjects with a novel unbiased and scalable proteogenomics approach

Authors:

A. Siddiqui, M. Zamanighomi, H. Guturu, J. Wang, T. Huang, R. Benz, S. Batzoglu; Seer, Redwood City, CA

Abstract Body:

Biofluids are a rich source of protein biomarkers for early detection of diseases, but the large dynamic range of protein concentrations in some biofluids like plasma necessitates complex workflow and trade-offs between throughput, coverage, and precision. Here we use an unbiased, deep, and rapid proteome interrogation approach, which leverages multiple physicochemically distinct nanoparticles to provide broad coverage of the plasma proteome at scale. In this study, we aim to identify protein biomarkers that distinguish early and late-stage Alzheimer's disease (AD) from blood plasma. Plasma samples from 200 subjects, comprising 100 AD and 100 controls, were interrogated using the Proteograph™ Product Suite (Seer Inc.). Proteins were quantified by data-dependent acquisition (DDA) and data-independent acquisition (DIA) liquid-chromatography mass-spectrometry (LC-MS) analysis. Across the samples, DIA detected 39,699 peptides, and 5,060 plasma proteins. DDA yielded similar results, detecting 36,496 peptides, and 4,706 proteins. Additionally, for DDA data, we performed a semi-tryptic peptides and phosphorylated peptides search, which increased the peptide identification by 41%, and the protein group identification rate by 2%, resulting in higher sequence coverage per protein. Differential expression analysis using DDA/DIA peptide intensities from the cohort showed 788, 96/608, 83 significantly over-expressed peptides and 132, 93/85, 74 significantly under-expressed peptides for age and sex, respectively. Furthermore, 126/129 protein groups were identified that are enriched in the early stage of AD; resulting in a robust classifier with blood plasma samples that can separate healthy controls from any stage of AD with an AUC-ROC of 0.75/0.72. Finally, pQTL analysis was conducted on a subset of the individuals (139) genotyped on the Global Screening Array platform. We used BOLT-LMM to apply a simple linear model with a χ^2 statistic with 1-degree of freedom to establish significance of association between every SNP and Protein:Nanoparticle intensity. We further performed shuffled cohorts for FDR < 0.01 to achieve high confidence in our associations. We were able to identify 73/59 cis-pQTLs associated with 20/20 proteins and 41/38 SNPs, and approximately 1,527/3,288 trans-pQTLs from DDA/DIA. These analyses identified a combination of known and potential new candidate plasma protein markers, demonstrating the Proteograph platform's ability to perform unbiased, deep, and rapid interrogation of the plasma proteome, enabling large-scale studies to detect novel insights with clinically relevant potential.

Omics Technologies Posters - Thursday

PB2959. DeepCell, A novel cell type specific marker detection machine learning tool for single cell transcriptome

Authors:

A. Ali¹, N. Nassir¹, A. Ahmed¹, A. Nazir², M. Uddin^{1,3}; ¹Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates, ²Zayed Univ., Abu Dhabi, United Arab Emirates, ³GenomeArc Inc., Cellular Intelligence Lab, Toronto, ON, Canada

Abstract Body:

We can gain critical insight into the development and differentiation of multicellular organisms at the cytological level using single cell transcriptomics. However, due to a lack of robust cell marker genes, cell type annotations pose a challenge. In this study, we try to develop a method for detecting markers in single cell clusters. DeepCell is a tool that uses machine learning algorithms, characteristically, artificial neural networks and support vector machines to annotate cell types from single cell transcriptome data. DeepCell was applied to Single nuclei sequencing data from brain tissues of patients with progressive supranuclear palsy (PSP), in addition to publicly available data such as single cell transcriptome data from the developmental single cell atlas of gene regulation and expression. This facilitated detecting robust gene markers relevant for different cell types. We were able to correctly discern cell types using pretrained DeepCell with high classification accuracy (greater than 90%) and minimal running time. The analysis identified known canonical gene markers of sub clusters such as *GFAP* in astrocytes from the PSP data, in addition to, *NPHS2* and *SLC12A1/NKCC2* in podocytes and thick ascending limb of the loop of Henle respectively using metanephric data from the fetal single cell transcriptome atlas. We were also able to find less well-known gene markers such as, *OBII-ASI* in astrocytes and *KIRREL3* in podocytes, which clustered well within their respective cell types ($P < 1.00 \times 10^{-3}$). To identify other novel markers, this analysis was repeated for multiple cell types in different tissues. DeepCell is an efficient method that uses multiple machine learning algorithms to identify gene markers with good performance on real Single cell genomics data in terms of sub cluster annotation and novel marker identification.

Omics Technologies Posters - Wednesday

PB2960. DeepConsensus v0.3: Gap-aware sequence transformers for sequence correction.

Authors:

A. Belyaeva¹, G. Baid¹, D. E. Cook¹, K. Shafin¹, T. Yun¹, F. Llinares-López¹, Q. Berthet¹, A. Töpfer², A. M. Wenger², W. J. Rowell², H. Yang¹, A. Kolesnikov¹, W. Ammar¹, J-P. Vert¹, A. Vaswani¹, C. Y. McLean¹, M. Nattestad¹, P-C. Chang¹, A. Carroll¹; ¹Google LLC, Mountain View, CA, ²Pacific BioSci.s, Menlo Park, CA

Abstract Body:

Pacific Biosciences (PacBio) circular consensus sequencing (CCS) generates long (10-25 kb), accurate "HiFi" reads by combining serial observations of a DNA molecule into a consensus sequence. The standard approach to consensus generation uses a hidden Markov model (pbccs). We have developed DeepConsensus, a deep-learning based approach for correcting errors in HiFi sequence data. DeepConsensus uses a novel alignment-based loss to train a gap-aware transformer-encoder (GATE) for sequence correction. Here, we introduce DeepConsensus v0.3, which includes significant speed and accuracy improvements.

DeepConsensus v0.3 reduces errors in PacBio HiFi reads and improves Q30 yield by 62% compared to pbccs in human sequence data (an improvement of 12% from DeepConsensus v0.2). Additionally, DeepConsensus v0.3 reduces model runtime by 4.5x compared to v0.2. These improvements are achieved via changes to the training data (CHM13), filtering (at read and window level), and model architecture. For downstream applications, DeepConsensus has been shown to improve the contiguity, completeness, and correctness of genome assembly when compared to assemblies generated using pbccs reads. DeepConsensus also demonstrates improved accuracy of variant calling when using DeepConsensus reads in comparison to pbccs reads. Instructions for downloading and using DeepConsensus v0.3 are available at <https://www.github.com/google/deepconsensus>.

Omics Technologies Posters - Thursday

PB2961. *DeepLoop* robustly maps chromatin interactions from sparse allele-resolved or single-cell Hi-C data at kilobase resolution

Authors:

S. Zhang, D. Plummer; Case Western Reserve Univ., Cleveland, OH

Abstract Body:

Hi-C has transformed our understanding of mammalian genome organization and can reliably identify high-order 3D genome features such as compartments and topological associated domains (TADs). However, the Hi-C contact heatmaps quickly become noisy due to the increasingly complex bias structure and severe data sparsity. To date, genome-wide mapping of chromatin loops, especially the enhancer-promoter (E-P) interactions within TADs (sub-TAD), remains a major challenge in Hi-C analyses. Here we present *DeepLoop*, which performs rigorous bias-correction followed by deep-learning-based signal-enhancement for robust chromatin interaction mapping from low-depth Hi-C data. The lower limit of read depth is ~10M mid-range cis contacts, typically can be obtained from 50~100M total reads. Nearly all published Hi-C datasets have adequate reads for *DeepLoop* reanalysis. Existing single-cell Hi-C technologies can yield enough reads from a few dozen cells. *DeepLoop* also achieves a cross-platform convergence between different Hi-C protocols and micro-C. *DeepLoop* allowed us to map the genetic and epigenetic determinants of allele-specific (AS) chromatin interactions in human genome. We nominate new loci with AS-interactions governed by imprinting or allelic DNA methylation. We also discovered that in the inactivated X chromosome (Xi), local loops at the DXZ4 “megadomain” boundary escape X-inactivation, but the FIRRE “superloop” locus does not escape. Importantly, *DeepLoop* can pinpoint heterozygous SNPs and large structure variants (SVs) that cause allelic chromatin loops, many of which rewire enhancers with transcription consequences. Taken together, *DeepLoop* expands the use of Hi-C to provide loop-resolution insights into the genetics of 3D genome.

Omics Technologies Posters - Wednesday

PB2962. DeepMod2: A deep learning framework for DNA and RNA modification detection using Oxford Nanopore sequencing.

Authors:

M. Ahsan¹, A. Goulu¹, K. Wang^{1,2}; ¹Raymond G. Perelman Ctr. for Cellular and Molecular Therapeutics, Children's Hosp. of Philadelphia, Philadelphia, PA, ²Dept. of Pathology and Lab. Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

DNA and RNA modifications, such as 5-methylcytosine (5mC) and N6-methyladenosine (m6A), play important regulatory roles in gene transcription, translation and other biological processes. Additionally, base analogs with synthetically introduced DNA or RNA modifications can serve as markers for DNA replication speed and transcription velocity in a variety of biological contexts.

DeepMod2 is a deep learning framework for DNA and RNA modification detection using Nanopore sequencing. It extracts features from ionic current signals of Nanopore sequencing and uses deep neural networks to identify modified bases at single-base and single-molecule resolution. DeepMod2 expands upon our previous work, DeepMod, by enabling modification detection from FAST5 files generated by Guppy basecalling or Tombo resquiggle, with a significant improvement in runtime and accuracy. For genome-wide methylation detection, given per-read predictions, DeepMod2 incorporates haplotype phasing with context-aware features such as CpG density and proximity to promoters and enhancers to produce accurate per-site predictions. Similarly, for RNA modification detection, DeepMod2 considers 3' UTR localization and short sequence motifs to accurately predict m6A methylation.

On the 72X HG002 ONT dataset, DeepMod2 can carry out whole genome 5mC detection in 45hrs using 16 CPUs, 1GPU, and 180GB of memory, achieving an F1-score of 99% for per-site modification detection. In addition to naturally occurring modifications, DeepMod2 also provides models for synthetic modifications such as IdU and BrdU. In summary, DeepMod2 allows accurate detection of base modifications from DNA or direct mRNA sequencing data on the Oxford Nanopore platform.

Omics Technologies Posters - Thursday

PB2963. Delivering Comprehensive, Single-molecule Proteomics Using Protein Identification by Short-epitope Mapping (PrISM)

Authors:

G. Kapp, J. D. Egertson, C. Stawicki, T. Gillies, J. Joly, N. Steiner, K. Cardenas, J. Leano, K. Chen, S. Watts, H. Boyce, D. DiPasquo, A. Killeen, V. Lobanov, S. Patel, P. Mallick; Nautilus Biotechnology, San Carlos, CA

Abstract Body:

Genomics has been revolutionized by next-generation sequencing technologies that made comprehensive genome analysis widely accessible. The field of proteomics is poised for a similar revolution to enable comprehensive analysis of all the proteins in a sample with increased sensitivity, reproducibility, and ease of use. The ability to detect and quantify a wider array of proteins, including low expressers, should complement genomic and transcriptomic data to enhance our understanding of mechanisms linking genotype to phenotype in complex biological systems. Here we present a novel approach for protein detection and quantification, Protein Identification by Short-epitope Mapping (PrISM). PrISM is a single-molecule protein analysis method where intact proteins are immobilized and analyzed massively in parallel by using non-traditional fluorescently-labeled affinity reagents to create a pattern of binding that allows for the identification of each individual protein molecule at scale. PrISM uses reagents with high affinity and low specificity that bind to short epitopes in multiple proteins. Simulations using these reagents show that the accumulated information from multiple rounds of detection of short, 2-4 amino acid epitopes enables identification of more than 95% of the human proteome with just 300 reagents. We have developed reagents that bind to short, three amino acid epitopes with low picomolar affinity. PrISM leverages a novel algorithm to process single-molecule binding data to determine protein identity. With single-molecule detection, sensitivity across a wide dynamic range is directly related to the number of molecules analyzed. We demonstrate a prototype chip that can immobilize 10 billion individual proteins to provide up to 9 orders of magnitude dynamic range in plasma. We measured millions of single protein molecules and successfully identified a set of model proteins with a smaller number of affinity reagents, exemplifying the PrISM approach. By combining single-molecule analysis, intact (non-digested) proteins, and iterative probing, PrISM provides a new tool for the discovery and quantitation of proteins. Paired with genomics, PrISM will enable insight into mechanisms linking genotype to phenotype to uncover novel biology, identify more effective biomarkers, and better understand the molecular mechanisms of disease.

Omics Technologies Posters - Thursday

PB2964. Depth normalization for single-cell genomics count data

Authors:

S. Boeshaghi, I. Hallgrimsdottir, Á. Gálvez-Merchán, L. Pachter; California Inst. of Technology, Pasadena, CA

Abstract Body:

Single-cell genomics analysis requires normalization of feature counts that stabilizes variance while accounting for variable cell sequencing depth. We discuss some of the trade-offs present with current widely used methods, and analyze their performance on 526 single-cell RNA-seq datasets. The results lead us to recommend proportional fitting prior to log transformation followed by an additional proportional fitting.

Omics Technologies Posters - Wednesday

PB2965. Detecting and assembling non-reference LINE-1 insertions using clustered long reads.

Authors:

M. Blacksmith, J. V. Moran, J. M. Kidd; Univ. of Michigan Med. Sch., Ann Arbor, MI

Abstract Body:

Accurately detecting and mapping genetic variation is one key to understanding the evolutionary history of a species. This process can be relatively straightforward when investigating single nucleotide variants or short insertions and deletions because Illumina sequencing reads typically span the length of the variant. However, for variants that cannot be spanned by a single Illumina read, such as dimorphic Long Interspersed Element 1 (L1) retrotransposon insertions, other approaches need to be used to accurately define variation. L1s replicate via a copy and paste mechanism termed retrotransposition. Mutations derived from the activity of L1 contribute to multiple canine phenotypes. Analyses using long-read DNA sequence data indicate that dimorphic L1s are eight-fold more common in canines than humans. However, non-reference L1 sequence reads often align to reference L1s, while other segments of the non-reference L1 reads align to alternative integration sites, leading to the generation of supplementary alignments that often are ignored by existing tools. To overcome these mapping challenges, I developed a computational pipeline, Alternative long-Read Alignment Clustering to Identify non-reference L1 insertions (ARACIL) to assemble non-reference L1 insertion locations from Pacific Biosciences (PacBio) long-read sequence data. ARACIL identifies the alternative alignment locations of putative non-reference L1 sequence reads. After extracting the alternative alignments, nearby alignments are clustered to generate candidate non-reference L1 insertion loci. Reads overlapping candidate loci are assembled using various assemblers (*i.e.*, Flye, Canu, and Wtdbg2) and the resultant assembled contigs are remapped to the reference genome to determine the location of the non-reference L1s. Non-reference L1 insertions are then assessed for L1 structural hallmarks, including the presence of a 3' poly(A) tail, flanking target site duplications, and 3' sequence transductions. Of 704 non-reference L1s discovered in a recently published Great Dane genome assembly, ARACIL confirmed 657/704 (93.3%) L1 insertions at identical genomic coordinates and revealed that 30/704 (4.3%) were located within 50 bp of previous annotations. In addition, I identified 3,030 L1s that were greater than 50 bp away from known non-reference L1 insertions. I currently am using ARACIL to systematically identify non-reference LINE-1 insertions in PacBio reads from six additional samples, which include four breed dogs, one wolf, and one dingo. These analyses will help elucidate how L1s have generated and may continue to generate genetic diversity in the *Canis* lineage.

Omics Technologies Posters - Thursday

PB2966. Detection of 5-hydroxymethylcytosine using a modified EM-seq protocol.

Authors:

D. Evanich, V. Ponnaluri, V. Panchapakesa, A. Erijman, M. Campbell, N. Dai, B. Langhorst, R. Vaisvila, T. Davis, L. Williams; New England Biolabs, Ipswich, MA

Abstract Body:

DNA methylation is an epigenetic regulator of gene expression with important functions in development and diseases such as cancer. The modified cytosines, 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), are routinely detected by sequencing Illumina libraries generated using an enzyme-based workflow called EM-seqTM or bisulfite. However, these methods cannot differentiate between 5mC and 5hmC. There is also increasing interest in identifying 5hmC sites due to its role in regulating gene expression in embryonic stem cells and a variety of neuronal cell types. Methods currently exist to enable discrimination of 5mC and 5hmC (e.g., oxBS-seq and TAB-seq), however both require a bisulfite conversion step leading to increased DNA damage. Here we describe a fully enzymatic method that enables specific detection of 5hmC. 5hmC libraries are generated by ligating adaptors onto sheared DNA. Next, 5hmCs are glucosylated, which protects them from deamination by APOBEC. Unprotected cytosines and 5mCs are deaminated resulting in their conversion to uracil and thymine, respectively. Libraries are amplified and 5hmC is discriminated from cytosine and 5mC by Illumina sequencing. 5hmCs are sequenced as cytosines whereas 5mC and cytosine are sequenced as thymines. Additionally, subtracting 5hmC data from EM-seq data (detects 5mC and 5hmC) enables the precise localization of individual 5mCs and 5hmCs. 5hmC data were generated for 0.5 ng to 200 ng DNA isolated from adult human brain (Biochain). The global level of 5hmC in the CpG context measured by these libraries was approximately 20% and was highly consistent across inputs. In addition, 5hmC levels were profiled during mouse E14 cell differentiation over a period of 10 days. 5hmC levels dropped from 3% to 0.7% over the time course as observed by both LC-MS quantification and Illumina sequencing. The 5hmC libraries had similar characteristics to EM-seq libraries with no detectable DNA damage, and as a result the libraries have expected insert sizes, low duplication rates and minimal GC bias. T4147 phage DNA was used as an internal control (all cytosines are 5hmC), with 97-99% of cytosines correctly identified. This provides a high level of confidence in the detection of 5hmC using this method. The specific profiling of 5hmC along with EM-seq analysis are key tools that will enable future studies of 5mC and 5hmC in the epigenetic control of genes.

Omics Technologies Posters - Wednesday

PB2967. Detection of somatic variation in human neurons using long-read sequencing approaches

Authors:

W. Zhou¹, C. Mumm¹, J. Switzenberg¹, P. Todd¹, M. McConnell², A. Boyle¹, R. Mills¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Lieber Inst. for Brain Dev., Baltimore, MD

Abstract Body:

Somatic mutations leading to genomic mosaicism in neurons can alter the expression of brain genetics and behavior and contribute to the complex regulatory landscape of neuropsychiatric and neurodegenerative diseases. While single-nucleotide variants in the genomes of human neurons have been well characterized, the identification of large-scale deletions and cryptic insertions has been elusive due to technical barriers. Here, we explored somatic mosaicism of these copy number variants (CNVs) in bulk tissue and at the single-cell level. First, we developed a Cas9-based enrichment pipeline for targeted whole-genome nanopore sequencing and demonstrated the robustness of our pipeline in a cell-line mixing experiment. Our pipeline is able to detect mosaic non-reference insertions as low as 3% frequency in one sequencing flow cell. We then applied our approach to the bulk frontal cortex and cerebellum tissue from a representative Alzheimer's disease brain and identified potential somatic insertions that could potentially alter gene expression. Next, we developed a single-cell analysis pipeline to identify CNVs with whole-genome nanopore sequencing of MALBAC-amplified DNA from flow-sorted cortical neurons of a neurotypical individual. Our pipeline can identify germline and somatic heterozygous large deletions and cryptic types of insertions, e.g., transposable elements, in single human neurons and identify some events with breakpoint resolution. We report the genomic distribution of these CNVs and explore their mechanistic origins and potential impact on gene expression. Overall, both complementary approaches increase our understanding of the prevalence and relevance of somatic mosaicism in human brains and aid in studies of neuropsychiatric and neurodegenerative diseases.

Omics Technologies Posters - Thursday

PB2968. Developing local compute capacity - a Ugandan experience Integrated Biorepository of H3Africa Uganda, Makerere University College of Health sciences.

Authors:

D. Kezimbira¹, N. L. Lwanga¹, E. Katagirya²; ¹Makerere Univ. Coll. of Hlth.Sci., Sch. of BioMed. Sci., Kampala, Uganda, ²Coll. of Hlth.Sci., Makerere Univ., Kampala, Uganda

Abstract Body:

Background:Data-intensive sciences, including bioinformatics, are relatively new to most parts of Africa, Uganda inclusive. Close to a decade ago, Wellcome Trust, together with the American National Institutes of Science (NIH), co-founded a consortium of projects dubbed the Human Health and Heredity Africa (H3Africa) with the aim of developing capacity on the African continent both of human resource and infrastructure to bridge the south-north disparity in genetics and genomics research gap including but not limited to developing human resource and setting up computational infrastructure among other things. Makerere University College of Health Sciences was one of the collaborators in one of the consortia projects named the Collaborative African Genomics Network (CAfGEN). CAfGEN comprised five institutions, mainly in Africa, i.e., Botswana, Eswatini and Uganda and one collaborating institution in the United States - the Baylor College of Medicine in Houston, Texas. Under CAfGEN servers were set up at Makerere University with the following specifications with 32TB of storage and 128GB of RAM. After a decade, several other servers have been set up at the Infectious Diseases Institute also at Makerere University as well as at the Uganda Virus Institute (UVRI) in Entebbe, Uganda. These servers have varying computational capacities and have been supported by different funding agencies. **Aim:** This study aimed to assess the capacity, accessibility, challenges, and opportunities that have been experienced locally. **Methods:** We used key informant interviews and anecdotal evidence from conversations with the users and administrators of these servers. **Findings:** We found that there was increased computational capacity at all the sites with capacity ranging from 20TB to 80TB of storage and up to 1PB of computational space. All the servers were freely accessible to local users within the institutions within which they were located, although they were also open to non-local users under specific predefined arrangements. The commonest challenges faced were limited space, lack of adequate administrative support, as well as unreliability arising from power failures and poor internet connectivity. **Conclusions and recommendations:** There is a need for additional computational capacity with improved internet connectivity. There is also a need for local funding from the government and other local organizations, as most of the financing for the servers was funded mainly by foreign funding agencies.

Omics Technologies Posters - Wednesday

PB2969. Development of computational methods for analyzing single-cell spatial transcriptomic data with applications to murine spermatogenesis.

Authors:

A. Vargo, D. Hannum, G. Manske, S. Hammoud, J. Li; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Single-cell RNA sequencing (scRNAseq) methods have transformed our ability to characterize cellular heterogeneity in complex tissues. The next frontier is spatial analysis, adopting new imaging approaches that are high-resolution (for single RNA molecules) and highly multiplexed (for hundreds of different genes at the same time using MERFISH-like combinatorial coding approaches) to survey the spatial layout of mRNA molecules in large tissue sections. Unlike existing transcriptomic data, these spatial data reveal the physical organization of intact tissues, allowing for the classification of spatial domains based on architectural features or distinct patterns of cell type heterogeneity.

In this work, we develop methods for analyzing single-molecule spatial data for hundreds of target genes. Our focus is on examining the spatial layout of the transcriptome in cross sections of testes from multiple adult mice to develop a deeper understanding of mammalian spermatogenesis. Each testis cross section contains hundreds of thousands of cells and millions of observed transcripts, and gleaning insights from these massive data sets involves multiple levels of computational challenges.

An early task in our analysis is to learn cell boundaries based on the distribution patterns of detected transcripts, DAPI, and three cell surface markers. Some testis cell types, e.g., Sertoli cells, consist of cells with irregularly shaped boundaries; this makes cell segmentation difficult. Another task is cell type annotation, which is often accomplished by integrating spatial data with existing scRNAseq data. We can then start to investigate higher-level biological questions that have been difficult to explore with past approaches. For instance, germ cells move radially inward towards the center of the seminiferous tubule as they develop into mature sperm. This radial organization, which shifts along the length of the tubule, has traditionally been defined in terms of cycles of tubule stages. We can now provide precise transcriptomic descriptions of tubule stages using our single-molecule spatial data. More specifically, we can annotate the biological staging of the seminiferous tubules at multiple levels: from intracellular transcript distributions, to cell types, to heterogeneous cell communities. This culminates in the ultimate goal of this study: the construction of a 3D reference atlas of a typical tubule, containing the locations of thousands of cells, each with hundreds of detected mRNA molecules.

Omics Technologies Posters - Thursday

PB2970. Diagnosis of neurodevelopmental disorders by RNA-Seq : From urine derived stem cells to induced neural stem cells, a promising alternative tissue.

Authors:

K. Riquin¹, T. Besnard^{2,1}, E. Trochu², S. Bezieau², B. Cogné²; ¹INSERM, CNRS, UNIV Nantes, l'institut du thorax, Nantes, France, ²CHU de Nantes Hôtel-Dieu, Nantes, France

Abstract Body:

Intellectual disability (ID) is a clinically and genetically heterogeneous group of neurodevelopmental disorders (NDD). It is characterized by deficits in intellectual and adaptive functioning beginning before adulthood, with a lasting effect on development. During the last decade, the implementation of genome sequencing helped to improve molecular diagnosis of ID. Even if this technique has a high diagnostic yield, one third of cases stay without candidate variant. In addition, many variants of interest are located in non-coding regions and are *de facto* difficult to interpret.

RNA-Seq have the potential to highlight pathogenic variants outside coding regions. Most recent studies showed an increased diagnostic rate of up to 17% in WGS negative patients with diverse genetic disorders (Murdock et al., 2021). To date, two tissues are frequently used : whole blood and fibroblast. It has been observed that RNA-Seq on fibroblasts enable a better diagnostic yield compare to whole blood. It is particularly true in the context of ID because few genes associated to neurodevelopment are expressed in blood. We recently performed RNA-Seq on fibroblast in 10 patients negative after genome sequencing. We showed that RNA-Seq on fibroblasts can successfully identify differentially expressed genes, to allow studying variants that were previously undetected or not considered during genome analysis.

Nevertheless, fibroblasts culture requires a skin biopsy, an invasive procedure sometime refused by the patient. We thus studied, Urine derived Stem Cells (USC) as an alternative tissue. USC are a subpopulation of cell collected in urine with self-renewal and multipotency capacities (Bharadwaj et al., 2013). They are easy to obtain by a non-invasive procedure and clonal colonies can be isolated in culture. We showed that 63% of ID-associated genes are expressed with a TPM>10, a threshold that should allow to study aberrant expression, monoallelic expression and aberrant splicing. We thus are collecting urine samples from patients suffering from NDD and their parents to study the diagnostic potential of USCs. Additionally, USC can be transdifferentiated to a neuron-like cell type called induced neural stem cells (iNSC). Depending on the protocol followed, iNSC can express up-to 78% genes associated to ID and are thus an interesting cell type to study the expression of genes only expressed in neurons and conduct functional assays.

We will present how RNA-Seq can be highly valuable in combination with exome or genome sequencing for the diagnosis of neurodevelopmental disorders. We will also discuss advantages and limitations of RNA-Seq using two promising cell types : USC and iNSC.

Omics Technologies Posters - Wednesday

PB2971. Diagnostic metabolomic profiling of hypomyelinating leukodystrophy caused by DEGS1 deficiency

Authors:

C. Gijavanekar¹, N. Liu^{1,2}, R. Logan³, S. Keller^{3,4}, S. H. Elsea^{1,2}, V. R. Sutton^{5,1,2}, Q. Sun^{1,2}; ¹Baylor Coll. of Med., Houston, TX, ²Baylor Genetics, Houston, TX, ³Children's Hlth.care of Atlanta, Atlanta, GA, ⁴Emory Univ. Sch. of Med., Atlanta, GA, ⁵Texas Children's Hosp., Houston, TX

Abstract Body:

DEGS1 (NCBI Gene ID: 8560, chr1q42.11) encodes the sphingolipid delta 4-desaturase 1 (NCBI RefSeq NP_003667.1, NM_003676.3) which catalyzes last step of ceramide de novo biosynthesis in the endoplasmic reticulum (ER) by introducing a double bond between positions 4 and 5 of dihydroceramide. Ceramide is the precursor of all sphingolipids which are essential cellular membrane components. Biallelic pathogenic variants in *DEGS1* have been recently associated with hypomyelinating leukodystrophy 18 [MIM: 618404]. Untargeted metabolomic analysis was performed using the Global Metabolomic Assisted Pathway Screening test (Global MAPS®, Baylor Genetics) to assess metabolic abnormalities associated with ceramide and sphingomyelin pathway in two unrelated individuals with variants in *DEGS1* gene: One with a novel homozygous variant of uncertain significance, c.775C>T (p.His259Tyr) and second with a novel homozygous likely pathogenic variant c.825+4_825+5delinsTT (p.?), predicted to affect normal splicing. Both had global developmental delay, leukodystrophy, epilepsy, microcephaly, and failure to thrive. In both individuals, metabolomic profiling identified elevated levels of dihydroceramide (Z-scores +3.4, +4.0) and its various metabolites dihydrosphingomyelins (Z-scores +6.8, +5.0), sphinganine-1-phosphate (Z-scores +2.4, +3.2) which are upstream of the enzyme defect. In addition, analytes downstream of enzyme defect, including multiple ceramides and their metabolites, such as glycosylceramides, lactosylceramides were not detected. Notably, among more than 2000 plasma metabolomic profiles tested, only these two individuals showed absent ceramides. Numerous sphingomyelins (Z-scores < -4.0 and as low as -19.3), and sphingosine-1-phosphate (Z-scores -5.8, -6.6) were reduced. This metabolomic profile is consistent between both unrelated patients, and unique compared to >2000 patient in our database. Functional validation of this ceramide pathway defect supported the reclassification of the novel *DEGS1* VUS, c.775C>T in case 1 and aided in reclassification as likely pathogenic. In conclusion, clinical biochemical assessment and functional validation of the *DEGS1*-metabolic pathway and VUS were achieved by plasma Global MAPS analysis, for which enzyme assay or targeted biochemical testing is not currently available in clinical laboratories. Global MAPS may be utilized for diagnosis of disorders associated with ceramide and sphingomyelin metabolism. In addition, metabolomic testing may potentially aid biomarker screening in developing and monitoring therapeutic targets.

Omics Technologies Posters - Thursday

PB2972*. Direct haplotype-resolved 5-base HiFi genome sequencing allows for linking rare disease variants to non-coding function

Authors:

T. Pastinen¹, W. Cheung¹, W. Rowell², D. Portik², R. Hall², A. Wenger², E. Grundberg¹; ¹Children's Mercy Kansas City, Kansas City, MO, ²Pacific BioSci.s, Menlo Park, CA

Abstract Body:

Long-read PacBio HiFi genome sequencing (GS) allows for accurate detection of single nucleotide variants (SNV), indels and structural variants (SV). Recent algorithmic development now enables the simultaneous detection of CpG methylation (mCpG). The phasing power of long contiguous reads (12-16kb) uniquely accesses haplotype information on allelic mCpG at 10X increased efficiency as compared to whole genome bisulphite sequencing (WGBS). We previously showed that allelic regulatory element (RE) hyper-mCpG reflects allelic regulatory silencing. Leveraging our large rare disease program "Genomic Answers for Kids" (GA4K), we have generated a comprehensive haplotype-resolved mCpG dataset for investigating unsolved pediatric rare disease. WGBS derived from GA4K patients (n=1085) were screened for hyper-mCpG outliers which were then validated in matching HiFi-GS data (n=77). Mean, standard deviation (SD) and 99th quantile (Q99) was summarized for 27M WGBS CpGs. Extreme hyper-mCpG tiles (200bp) were identified if >1 CpG/tile had mCpG >3 SD than Q99. The average mCpG from HiFi-GS were then calculated for each extreme hyper-mCpG tile. We validated the HiFi-GS hyper-mCpG tiles at known rare disease loci including diagnostic mutations (e.g. *DMPK* repeat expansion) and imprinted genes (e.g. *GNAS*). In unsolved cases we identified 5-20 rare (<0.5%) hyper-mCpG tile/genome. Majority of the hyper-mCpG events were observed only in maternal or paternal allele and predicted to cause LRE. The most common cause of LRE was rare SNV within the RE causing a short-range (~250bp) hyper-mCpG event. The second most common cause was local SV (INS, DUP, DEL) yielding large hyper-mCpG events (>1 kb). Similarly, unstable tandem repeats (STR) accounted for larger hyper-mCpG tiles, including novel GCC repeats with complete promoter hyper-mCpG predicting allelic gene silencing in 8% of validated regions. Less than 5% of hyper-mCpG events were biallelic that may represent either rare homozygous regulatory effects (*cis*) or upstream regulatory perturbation by genetic (*trans*) or environmental mechanisms. Interestingly, 1% of validated hyper-mCpG tiles appeared to originate from transcriptional silencing caused by local loss of splice donor sites. Finally, minority of events remained unsolved and may represent distal *cis* (allelic hyper-mCpG) or tissue variation (non-allelic, variable hyper-mCpG). In conclusion, we have identified patient-specific hypermethylation events with evidence of functional impact of non-coding rare variation. We propose that unsolved rare disease genomes sequenced by HiFi-GS will allow detection of unconventional diseases alleles due to LRE.

Omics Technologies Posters - Thursday

PB2973. Magnified Convolutional Enrichment Representation Model

Authors:

G. Chen, W. Zheng; Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract Body:

Feature representation and the learning process formed the cornerstones of machine learning. Feature representation mathematically characterizes domain entities by encoding input features followed by feature normalization and feature extraction. It may immensely reduce the cost of the learning process and improve performance. In this work, we introduced an advanced model to represent entities with a dynamic mechanism, which was further enhanced by fusing the local and global over-representation on the basis of encyclopedic knowledge bases. The model has been evaluated and demonstrated good fitness for predicting the associations of complex diseases.

Omics Technologies Posters - Thursday

PB2974. Efficiency of mutagenesis promoted by CRISPR-Cas9 in *THRAB* gene in zebrafish.

Authors:

L. Rodrigues; Univ.e do Algarve, Faro, Portugal

Abstract Body:

Background: Gene editing tools allowed the study of isolated genes and complex gene networks and have been an asset in several lines of research, especially the CRISPR-Cas9 method. This technique differs from the others as the mutations produced are transmitted to offspring, conferring a stable expression/phenotype. However, despite this efficient tool for DNA editing, the frequent disagreement between genotype and phenotype is notorious, so the existence of genetic compensation mechanisms as the *Transcriptional Adaptation* is already known and remains open. The loss of function of the *THRAB* gene was promoted in order to establish the relationship between the functions of the nuclear receptor TR α B expressed by it and the phenotype observed during embryonic neurodevelopment. Excluding the hypothesis of genetic compensation promoted by the *THRAA* paralog gene in the offspring of the mutant line. **Methods:** The results were analyzed in offspring of zebrafish whose *THRAB* gene was deleted by CRISPR-Cas9 (a deletion of 9.000 bp), verifying the efficiency of the method in producing the mutation by sequencing its product. The description of the phenotype of the mutant line was then carried out, establishing the relationship between the TR α B receptor and the phenotype observed in this period. **Results:** The efficiency percentage of the method was 7.2% in F0 injected generation and 26.8% in F1. Through the phenotypic analysis in F2 homozygous generation, it was observed that the expression of the TR α B receptor is important in the development of the anteroposterior axis of the Central Nervous System. Changes in eye size and hindbrain and midbrain dysfunctions/disruptions associated with cognitive impairment have been described. **Discussion/Conclusion:** The CRISPR-Cas9 method is effective in producing loss of gene function, however there are particularities to be considered, such as the different molecular mechanisms of compensation and the type of mutation produced. From the described phenotype, new hypotheses emerge about the implications of TR α B in neuroembryonic development. Furthermore, the study contributes to the advancement and improvement of promising technology for gene therapy of thyroid hormone resistance syndrome (RTH).

Omics Technologies Posters - Wednesday

PB2975. Efficient DNA sample contamination metric estimation using a novel variant representation and algorithm.

Authors:

W. Lu^{1,2}, L. Gauthier¹, T. Poterba^{1,2}, C. Vittal^{1,2}, D. King^{1,2}, C. Stevens^{1,2}, M. Daly^{2,3,4}, C. Seed^{1,2}, B. Neale^{2,3}, K. Karczewski^{1,2}; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²Massachusetts Gen. Hosp., Boston, MA, ³Broad Inst Harvard & MIT, Cambridge, MA, ⁴Inst. for Molecular Med. Finland, Helsinki, Finland

Abstract Body:

DNA sample contamination is a major issue in clinical and research applications of whole genome and exome sequencing. Even modest contamination can dramatically affect the quality of variant calls and lead to widespread genotyping errors (e.g., contamination leads to systematic undercalling of homozygous genotypes in gnomAD). Currently, the most popular tool for detecting contamination (i.e., DNA from two or more humans in a single sample) is VerifyBamID, which is computationally expensive as it operates on underlying short read data (CRAMs or BAMs). For some datasets, the reads are no longer accessible, necessitating a different approach.

The more efficient gVCF format scales effectively to large sequencing studies and retains information on variant calls and genotyping quality sufficient to evaluate contamination. Capitalizing on the gVCF format, we present the Scalable Variant Call Representation (SVCR), which jointly represents multiple samples in storage space that scales linearly in the number of samples and variants and stores reference blocks and variant calls separately. An implementation of the SVCR in the Hail system, the Variant Dataset (VDS), provides the opportunity for devising gVCF-based metrics using only non-reference variant calls at a much lower computational cost.

Inspired by the VDS format and the empirical observation that contaminated samples contain more heterozygous calls with unusually high allele balance, we devise a novel gVCF-based contamination metric. Using only the non-reference component from the VDS, we compute the mean reference allele balance of the homozygous alternate genotypes of autosomal and biallelic SNPs with high coverage (sequencing depth $\geq 10\times$). We evaluate this metric on 73,456 whole genome samples from gnomAD v3 with 58,983 samples from the Broad Institute and 14,473 from other sequencing centers. Both platforms demonstrate strong correlations ($r=0.99$, $r=0.98$; $p < 2.2e-16$) between the new metric and the VerifyBamID contamination estimate. The correlation is consistent when computing the metric using common variants at various allele frequency cutoffs. This metric can be obtained through fast computation using only the variant calls from a VDS or gVCF and provides an accurate evaluation of sample quality. Our results demonstrate that the metric can efficiently replace the existing tools for estimating DNA sample contamination, which, in turn, will improve the accuracy and efficiency of downstream analyses of ultra-large whole genome and exome sequencing datasets in the future.

Omics Technologies Posters - Thursday

PB2976. Empowering Discovery in Childhood Cancer: Genomic Harmonization at the Kids First Data Resource Center

Authors:

D. Miller¹, M. Brown¹, Y. Zhu¹, B. Zhang¹, D. Higgins¹, B. Farrow¹, M. Mattioni², Z. Li³, K. Wang¹, A. Heath¹, A. Resnick¹; ¹The Children's Hosp. of Philadelphia, Philadelphia, PA, ²Seven Bridges, Cambridge, MA, ³Sentieon Inc, San Jose, CA

Abstract Body:

The Gabriella Miller Kids First Pediatric Research Program (GMKF) is an NIH Common Fund initiative focused on providing large-scale clinically annotated genomic data for pediatric cancer and structural birth defect cohorts.

Kids First is challenged with taking genomic data from a wide range of sources and providing researchers and physicians with harmonized genetic data. Preparing large-scale harmonized datasets presents unique challenges in scalability, reproducibility, and transparency. The Kids First Data Resource Center (DRC) tackles these challenges using open-source, community-standard workflows that are deployed in cloud-based HPC environments. Our workflows, written in Common Workflow Language (CWL), are modeled after established best practices workflows, optimized and verified through internal benchmarking, and, when possible, made in collaboration with our network of researchers. These workflows are made freely available both as open-source code on Github and as public apps via CAVATICA, an Amazon Web Services (AWS) based cloud computing platform associated with the Kids First DRC Portal co-developed by Seven Bridges Genomics, where workflows feature scatter-gather parallelization, conditional execution, and AWS resource optimization.

Today, the DRC has six production level workflows producing harmonized datasets at scale for whole genome sequencing, exome sequencing, and RNA-seq technologies. Additionally, we have over a dozen non-production workflows for everything from germline to tumor-only to long reads applications. These workflows have been run across the 24 Kids First studies and 20,000 participants already released on the Kids First Data Resource Portal. In total, we have more than 1.0 PB of data, with more being released yearly. Here we present our process for creating, validating, and distributing these workflows.

Omics Technologies Posters - Wednesday

PB2977. Enabling flexible low throughput sample preparation for multiple sequencing platforms using the Miro Canvas

Authors:

A. Day¹, M. Gildea¹, R. Kafrawi¹, A. Barner², S. Kumar², E. Carvajal², K. Cunningham², C. Ramnarine¹, E. LaRoche¹, K. Larkin¹, T. Howd¹, C-C. Lee², A. Chavan², T. DeSmet¹, N. Lennon¹, S. Gabriel¹; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²Miroculus, San Francisco, CA

Abstract Body:

Over the last five years, declining sequencing and library preparation costs have driven the demand for whole genome sequencing for both research and clinical applications. The Broad Institute's Genomics Platform has built the capacity to process thousands of long read samples and hundreds of thousands of short read samples per year. Advances in both automated liquid pipetting and quantification technologies have transformed library preparation into an efficient and more cost effective process. While high throughput processing is essential for meeting large project demand, there is a market need for an automated low throughput workflow that provides on-demand sample processing for time-critical applications. Ideally this workflow includes expedited processing to facilitate quick turnaround times and increased flexibility while maintaining high quality results.

Miroculus's Miro Canvas is an emerging technology that constructs one DNA library per run. This compact, digital microfluidic platform requires twenty minutes for sample preparation, after which the Miro Canvas runs independently. Using the same universal cartridge, it performs all enzymatic reactions, incubations and purifications that are required to produce sequencer-ready libraries for both short read and long read platforms. This instrument has the potential to fill the Genomics Platform's need for flexible, low throughput processing.

Here, we present data on the performance of PCR-free whole genome libraries made using the Miro Canvas and sequenced on Illumina, Pacific Biosciences, and Oxford Nanopore platforms. While the Miro Canvas has the functionality to construct libraries from mechanically sheared DNA, it also has the ability to perform enzymatic fragmentation. Samples prepared on the Miro Canvas performed equivalently to libraries prepared both manually and with automated liquid handling in terms of library yield (33.5% vs 29.9%) and sequencing quality (average Q35.3 vs Q35.9) even with a fraction of the reagents.

Furthermore, the Miro Canvas maintains line balance with similar run times and less touch points than manual or other automated processes. The instrument is straightforward to set up and requires minimal training translating to fewer user errors. The ability to alter protocols allows for rapid integration of R&D efforts in less time.

In the future, we see the Miro Canvas as a way to reduce turnaround times and increase lab flexibility especially when applied to niche applications, namely rapid WGS processing. The ability for the platform to achieve both high and low throughput demand will open a breadth of genomic capabilities and resources for the community.

Omics Technologies Posters - Wednesday

PB2979. Ensure sample identity in sequencing workflows with the Twist Sample ID Kit.

Authors:

S. Oh, M. Bocek, P. Wani, C. Littler, D. Murphy, E. Toro; Twist BioSci., South San Francisco, CA

Abstract Body:

Target enrichment systems, such as hybridization capture using the Twist Exome 2.0 panel, increase sample throughput by focusing on genomic regions of interest. Further, recent advances in automation and sequencing output have made it possible for laboratories to test hundreds or even thousands of samples per day. As the number of samples grows, however, so does the likelihood of an accidental sample swap. In some environments, such mistakes can be costly, and have a significant impact on drug development or research decisions. It is therefore important to develop methods that ensure the identity of samples can be accurately linked to their cognate results. Sample swaps can occur at any point along the route from collection to sequencing. At Twist, we developed a solution to ensure sample identity starting from an early step in the workflow. Here, an aliquot is taken directly from a blood collection tube and as little as **2 µl of whole blood can be directly amplified** in a multiplex PCR reaction. This simplifies the workflow, stakes sample identity before processing begins, and eliminates another source of sample swapping. The included primer pool and single-tube polymerase mix target a panel of SNPs, described by Pengelly, et al. and by The European Society of Human Genetics (ESHG) EuroGentest committee, that are useful for sample tracking and identification in human samples. Additionally, several loci, including AMELX/Y, are included for sex determination. The multiplexed product can then be barcoded and readied for sequencing using one of the Twist library preparation kits and the Twist 10 bp Unique Dual-Indexed (UDI) system. To confirm sample identities, the Sample ID panel SNPs are compared to the matched loci in the test samples. The Twist Sample ID kit may be used directly with the Twist Exome 2.0 panel. Alternatively, a sample ID target enrichment panel containing the targets of interest may be blended into your custom hybrid capture panel. Here we demonstrate the application of the workflow on a combination of different sample conditions: sample purity (whole blood vs purified genomic DNA), blood collection tube preservative (K2EDTA and Streck cfDNA), and sex (female and male). Based on population segments with the lowest measured minor allele frequencies (MAFs), we calculate the probability of encountering the same SNP profile on a 96-well plate is approximately 1 in 70,000, with probabilities for other populations commonly calculated to be at 1 in 950,000. With its direct-from-blood capability and power of discrimination, the Sample ID kit provides high confidence that the right results are matched to the right samples.

Omics Technologies Posters - Thursday

PB2980. Establishing a baseline transcriptomic profile of human eccrine sweat glands via single-cell RNA-sequencing with validation by 3D imaging

Authors:

A. Eastman, G. Rosson, N. Kim, G. Cutting, N. Sharma; Johns Hopkins, Baltimore, MD

Abstract Body:

Eccrine sweat glands (ESGs) are essential for human thermoregulation. While several genetic diseases are known to have ESG-related phenotypes, further knowledge of molecular processes involved in sweat production are required to elucidate etiologies. The immense atlases of transcriptomic information generated by single-cell RNA-sequencing (scRNA-seq) are based on tissue dissociated *en bloc*, which does not well capture cells from microstructures within tissues. Thus, only a minor fraction of skin scRNA-seq datasets is comprised of ESG cells. Consequently, no published scRNA-seq datasets currently contain all known ESG cell types. Therefore, we have performed scRNA-seq on micro-dissected ESGs to identify novel marker genes for ESG cell types and pathways implicated in ESG-related diseases. Individual ESGs were micro-dissected from skin and processed for scRNA-seq or fixation and optical clearing for immunofluorescent protein labeling (IF) and 3D imaging. For scRNA-seq, glands were enzymatically dissociated, 3' 10X cDNA library preparation was used, NovaSeq was used to sequence, and Seurat was used for analysis. The CLARITY protocol was followed for fixation and optical clearing, proteins of interest were labeled by IF, and a multi-photon microscope was used for 3D imaging. For scRNA-seq, we captured approximately 7,300 cells and used dimensionality reduction and clustering in Seurat to successfully separate them into all expected eccrine sweat gland cell types - clear cells, dark cells, myoepithelial cells, and ductal cells - as well as fibroblasts, pericytes, endothelial cells, and immune cells based on variable expression of published marker genes. Interestingly, several genes we show to be highly expressed in ESG dark cells are in the arachidonic acid catabolism pathway. This suggests that an arachidonic acid-based mechanism may trigger sialomucin release, which influences ESG lumen pH. Our optimized 3D imaging technique has allowed us to visualize ESG proteins - including *ACTA2*, *CFTR*, and multiple keratins - within the native geometry of the tissue, demonstrating its utility as a validation measure.

While previously published skin scRNA-seq datasets have captured 0, 1, or 2 ESG cell types, our technique has allowed us to identify all 4 major cell types. Enriching for microstructures in tissues throughout the body will greatly add to the level of detail of transcriptomic atlases. IF and 3D imaging of optically cleared ESGs allows for the validation of novel marker genes found in our scRNA-seq data, and for the establishment of the spatial organization of gene products.

Omics Technologies Posters - Wednesday

PB2981. Establishing best practice for structural variant discovery with long read sequencing in the Gabriella Miller Kids First Pediatric Research Program using Sentieon's haplotype-resolved variant evaluation tool

Authors:

Z. Li¹, M. Ahsan², H. Chen¹, D. Freed¹, K. Wang², Y. Zhu²; ¹Sentieon Inc., San Jose, CA, ²Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

As a NIH common fund project and one of the largest childhood cancer cohort programs, the Gabriella Miller Kids First (GMFK) Pediatric Research Program has sequenced 48,000 genomes representing over 20,000 patients with 44 childhood cancer and birth defects cohorts in the past 15 years. To take advantage of recent accuracy and throughput improvements in long read sequencing platforms, GMFK has initiated long reads pilot projects to enable sensitive and accurate discovery of genetic structural variation (SV) underlying childhood cancers and structural birth defects.

A key objective of the pilot program is to establish a production long read data processing pipeline that can reproducibly deliver high-quality data for cohort research. However, evaluation and qualification of long read SV pipelines has been particularly challenging due to limited truth sets and a lack of accurate evaluation tools. Additionally, popular SV evaluation tools perform a pairwise comparison of SVs, which results in the mischaracterization of more complex rearrangements, especially in tandem repeat regions.

As haplotype-resolved genome assemblies are becoming feasible, the development of better evaluation tools that can perform a faithful comparison with base-resolution has become critical.

Here, we present hap-eval, a new open-source haplotype-resolved SV evaluation tool. Instead of comparing each pair of calls, SV calls within a cluster are first assembled into haplotypes based on available genotype and phase information. The assembled haplotypes for the truth set and calls are then comprehensively compared, and the best-matching haplotypes are evaluated.

We used this tool to thoroughly benchmark and evaluate different SV discovery pipelines for both PacBio HiFi and Oxford Nanopore sequencing platforms. We have evaluated different long-read SV callers in combination with different aligners, and compared with SV truthsets for the HG002 sample developed by the Genome in a Bottle consortium. Platform-specific SVs callers were also benchmarked. Evaluation with haplotype-resolved SV callset on NA12940 were also performed. The haplotype-based comparison implemented in hap-eval is able to properly assemble and evaluate clusters of distinct SV calls providing a more accurate benchmark of SV discovery tools relative to earlier evaluation tools. Stratified analysis was performed for different genomic regions and SV lengths. We were also able to compare SV calls from different pipelines, which is essential in developing ensemble-based approach for SV discovery with further improved accuracy.

Omics Technologies Posters - Thursday

PB2982. Estimating and testing cell-type-specific gene co-expression from single-cell data: A robust and efficient method that mitigates technical confounding.

Authors:

C. Su¹, Z. Xu¹, X. Shan¹, B. Cai¹, J. Zhang², H. Zhao¹; ¹Dept. of Biostatistics, Yale Univ., New Haven, CT, ²Dept. of Management Sci., Univ. of Miami, Coral Gables, FL

Abstract Body:

Gene co-expression networks encode the functional organization of genes, and the single cell RNA-seq technology enables understanding such networks at the individual cell type level. However, sparse UMI counts from single cell data are measured with strong technical variations across cells, posing unique challenges to co-expression analysis. In this presentation, we first demonstrate that most existing co-expression estimation methods fail to address the confounding due to heterogeneous sequencing depths, resulting in falsely inferred co-expressions. For example, for a pair of independently simulated genes (5,000 cells, 100 simulation runs), at the statistical significance level of 0.05, Pearson correlations on log transformed and scaled data had an inflated type I error of 0.99 (SD=0.01), while state-of-the-art methods, such as locCSN and BZINB, also had inflated estimates of 0.66 (0.03) and 0.51 (0.01). In addition, at a fixed correlation strength, correlation estimates from these methods spuriously varied with mean expression levels (i.e., a mean-correlation relationship). To address the above limitations, we propose a novel and general bivariate expression-measurement model that explicitly accounts for the heterogeneity in sequencing depths. We also propose a new metric on gene co-expressions based on the true underlying expressions, which separates technical variations from biological ones and is robust to technical artifacts and variations in mean expression levels. To efficiently estimate large co-expression networks, we propose a fast method-of-moment estimator and an asymptotic test for co-expression identification. With simulated gene pairs, we show that our method successfully removed unwanted variation with nominal type I error control regardless of the expression level and achieved high power to infer co-expressed gene pairs. In large network simulations, our method achieved the highest AUC and its estimates were not confounded by gene expression levels (Pearson's $r=-0.00$, $p=0.98$), while the estimates from other methods were notably confounded by gene expression levels, such as Pearson correlations on log normalized data ($r=0.92$, $p<2.2e-16$) and SpQn ($r=0.68$, $p<2.2e-16$). Finally, we applied our method to single-nucleus RNA-seq data on Alzheimer's disease patients, and uncovered microglia-specific co-expressions among genes differentially expressed in microglia in diseased individuals. Further clustering analysis based on inferred co-expression patterns identified critical biological pathways involved in Alzheimer's disease, such as immune response, that were not discovered by other methods.

Omics Technologies Posters - Wednesday

PB2983. Evaluating deep learning for predicting epigenomic profiles

Authors:

Z. Tang; Cold Spring Harbor Lab., Cold Spring Harbor, NY

Abstract Body:

Deep learning has been successful at predicting epigenomic profiles from DNA sequences. Most approaches frame this task as a binary classification relying on peak callers to define functional activity. Recently, quantitative models have emerged to directly predict the experimental coverage values as a regression. As new models continue to emerge with different architectures and training configurations, a major bottleneck is forming due to the lack of ability to fairly assess the novelty of proposed models and their utility for downstream biological discovery. Here we introduce a unified evaluation framework and use it to compare various binary and quantitative models trained to predict chromatin accessibility data. We highlight various modeling choices that affect generalization performance, including a downstream application of predicting variant effects. In addition, we introduce a robustness metric that can be used to enhance model selection and improve variant effect predictions. Our empirical study largely supports that quantitative modeling of epigenomic profiles leads to better generalizability and interpretability.

Omics Technologies Posters - Thursday

PB2984. Evaluation of cell type annotation for peripheral blood mononuclear cells using single-cell RNA versus a deconvolution approach with bulk RNA sequencing.

Authors:

X. Liu, T. M. Norden-Krichmar; Univ. of California, Irvine, Irvine, CA

Abstract Body:

Background: Deconvolution methods have been widely used to computationally estimate cell type composition in bulk RNA-seq data, in order to reveal cell-type-specific gene expression without conducting single-cell RNA sequencing (scRNA-seq). However, direct comparison of deconvolution methods is challenging because bulk RNA-seq samples are often collected from different participants than the scRNA-seq samples. The goal of this study was to compare and evaluate cell types of peripheral blood mononuclear cells (PBMCs) from the same participants, as determined from scRNA-seq data versus cell types inferred by deconvolution from bulk RNA-seq data. **Methods:** Biospecimens from an IRB-approved study by the Southern California Alcoholic Hepatitis Consortium (SCAHC) were collected at baseline with consent by the participants from alcohol-associated hepatitis (AH, n=3) and healthy controls (CT, n=2). ScRNA-seq was performed using 10x Genomics Chromium Single Cell technology and Illumina sequencing. Sequence data was mapped to the human genome (hg19) and integrated using cellranger 6.0.2 version. We then performed downstream analysis using Seurat (version 4.0.1). Bulk-RNA data was aligned and mapped to the human genome (hg19) using Bowtie2 release 2.3.4.1 and Tophat2 release 2.1.1. Deconvolution was performed using Absolute Immune Signal (ABIS) with 29 human immune cell types for PBMC transcriptomic data. **Results:** Using the ABIS deconvolution approach, we identified 17 cell types from the bulk RNA-seq data, including 8 out of 9 cell types identified by Seurat cell annotation of scRNA-seq data. Natural-killer T cells were the only scRNA-seq cell type that was not inferred, due to their absence from the reference dataset in the ABIS method. Percentages of CD8+ T cells, CD14+ monocytes, and CD16+ monocytes were consistent between the deconvolution and scRNA-seq results. In addition, both ABIS and scRNA-seq showed an increased percentage of monocytes in AH vs. CT. ABIS deconvolution identified very little naïve CD4+ T cells and natural-killer cells, whereas ABIS categorized a large percentage of terminal status of T cells, such as gamma-delta T cells, mucosal-associated invariant T (MAIT) cells, and memory CD4+ T cells. **Conclusion:** This exploratory study assessed the performance of deconvoluted bulk RNA-seq against scRNA-seq data using samples of PBMCs from the same participants, enabling direct comparison of these two methods. Our results showed that the robustness of the deconvolution method is dependent upon the reference dataset selected, and therefore, the deconvoluted results require careful interpretation.

Omics Technologies Posters - Wednesday

PB2985. Evaluation of single nuclei RNA sequencing in brain tissue using Pre-templated Instant Partitions (PIPseq)

Authors:

R. Meltzer, K. Kugler, S. Pandey, A. May-Zhang, A. George; Fluent BioSci., Watertown, MA

Abstract Body:

Single nuclei RNA sequencing (snRNA-seq) is an essential method for transcriptional characterization of cells when viable cell scRNA-seq is not an option due to fragile, difficult, or frozen cell samples. PIPseq is a novel, single cell sequencing technology capable of instantly partitioning, lysing, and capturing mRNA from extracted nuclei. PIPseq has benefits over existing single cell methods by allowing researchers to capture and instantly partition their nuclei of interest without the use of microfluidics, and is intrinsically scalable from hundred to hundreds of thousands captured nuclei in a single sample. Here, we extracted nuclei from various mouse tissues using optimized nuclei isolation methods for each tissue. We performed snRNA-seq as well as scRNA-seq on each tissue in order to compare the nuclei transcripts to the whole cell transcripts in each tissue type. For neuronal tissue, we explored two different nuclei isolation techniques and analyzed the effect of each method on obtained transcripts and differentially expressed genes. In summary, these experiments allowed us to assess the similarities and differences between fresh, whole cells and frozen, single nuclei after single-cell/nuclei sequencing. This study emphasizes the importance of understanding how input material affects data quality and the biological interpretations that are derived from these preparations- especially when access to one sample type or the other is limited.

Omics Technologies Posters - Thursday

PB2986. Evercode™ Whole Transcriptome Single-Cell RNA-Seq with Focal Barcoding Enables Single-Cell Transcriptomes Coupled with Gene or Vector-Enriched Readouts

Authors:

J. Pangallo, L. Kenyon, A. Sova, R. Koehler, C. Roco, A. Rosenberg; Parse BioSci.s, Seattle, WA

Abstract Body:

A limitation of single-cell RNA sequencing is the inability to detect every expressed gene in a given cell without exhaustive sequencing, which can be cost prohibitive due to the sequencing depth required per cell. This is problematic for applications such as single cell CRISPR screens where it is essential to detect the guide RNA (gRNA) in every cell in order to associate a transcriptome-wide perturbation effect. To address this we demonstrate an approach, known as “Focal Barcoding”, that enables enrichment and highly sensitive detection of individual genes or transgenes in single cell experiments. This allows dissection of complex pathways and biological mechanisms associated with one or more specific markers and is completely compatible with Parse’s current suite of Evercode™ single-cell RNA sequencing kits that enable profiling of up to one million cells. By enriching a target of interest, focal barcoding dramatically reduces the sequencing needed to robustly detect that target across different cells. We demonstrate that enriched genes or transgenes of variable expression levels, some of which are only detected in a small percentage of cells, can be detected in > 90% of cells after using focal barcoding (even when sequenced at shallow depth). We will present this proof of principle data and discuss additional applications that will be enabled by Evercode™ focal barcoding, such as single cell CRISPR perturbation experiments.

Omics Technologies Posters - Wednesday

PB2987. Exome sequencing reanalysis complemented with combined multi-omics approach reached to 60% diagnostic yield in previously undiagnosed rare disease cohort

Authors:

K. Ounap^{1,2}, K. Oja^{1,2}, K. Reinson^{1,2}, K. Muru^{1,2}, T. Reimand^{1,2}, M. H. Wojcik^{3,4}, I. A. Osei-Owusu⁴, G. Michelotti⁵, A. O'Donnell-Luria^{3,4}, S. Pajusalu^{1,2}; ¹Univ. of Tartu, Tartu, Estonia, ²Dept. of Clinical Genetics, Genetics and Personalized Med. Clinic, Tartu Univ. Hosp., Tartu, Estonia, ³Boston Children's Hosp., Boston, MA, ⁴Broad Inst. of MIT and Harvard, Cambridge, MA, ⁵615 Davis Drive, Suite 100, Metabolon, Morrisville, NC

Abstract Body:

Rare diseases have heterogeneous underlying etiologies, and despite an application of different genomic analyses in clinical setting, only 30-50% of cases will receive a precise etiological diagnosis. In this study, we reanalyzed exome sequencing (ES) data in unsolved Estonian families from 2013-2019. In reanalysis negative cases, we carried out trio genome (GS), RNA sequencing, and/or untargeted metabolome analysis for the identification of new Mendelian disorders. We enrolled 154 individuals in whom single or trio ES was performed in clinical setting, but did not result in a confirmed diagnosis. Chromosomal microarray analysis, metabolic tests and fragile X syndrome testing in males are also regularly performed as part of the clinical workflow. The final study group consisted of 116 individuals; in all of them reanalysis of ES data was performed using the seqr platform, in collaboration with the Broad Institute Center for Mendelian Genomics. Among cases that remained negative after ES reanalysis, trio/quad GS was performed in 56 families, RNA sequencing in 49 individuals, and untargeted metabolome analysis in 35 individuals. We identified the disease cause in 70 (60%) individuals, of which 35 involved gene variant(s) or copy number variation in a previously known gene and 32 involved a novel gene. All novel gene variants were submitted to the Matchmaker Exchange. Among novel gene findings, 18 of them are published as a part of international collaboration (*RORA*, *OGT*, *H3F3B*, *RAB11A*, *POU3F3*, *CYFIP2*, *HIST1H1E*, *JAG2*, *POLRMT*, *TET3*, etc.). Reanalysis of ES data showed highest diagnostic efficacy - 49 (70%), among them 25 known and 24 novel genes. GS and RNA sequencing alone or in combination reached diagnostic efficacy of 16 (23%), 10 known and 6 novel genes. Metabolome data added special input in at least two cases (2 novel genes). Importantly, in one case, somatic known gene variant was identified and in two patients acquired etiology was confirmed after extensive diagnostic odyssey. Using this combined approach, our diagnostic yield increased by at least twofold. In the clinical setting in Estonia, the diagnostic yield of ES is 29% (Pajusalu et al., 2018). With this study, an additional 60% of previously unsolved cases received a genetic diagnosis with the total diagnostic yield reaching to approximately 70%. Our results highlight the importance of regular reanalysis of ES data as 70% of cases were solved by reanalysis. The added detection rate of GS and/or RNA sequencing is similar to previously published studies. Funding: Estonian Research Council grants PRG471, PUT355, and PSG774.

Omics Technologies Posters - Thursday

PB2988. Exome technology innovations advancing personalized medicine.

Authors:

K. Walker, J. Hu, Q. Meng, Q. Wang, X. Liu, Y. Dai, D. Pasham, M. Kaur, H. Shen, G. Weissenberger, S. Bhamidipati, K. Kottapalli, Z. Momin, G. Metcalf, H. Dinh, Y. Han, J. Santibanez, H. Doddapaneni, R. A. Gibbs, D. M. Muzny; Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX

Abstract Body:

Oligonucleotide capture-based DNA sequencing technologies have been a driving force for clinical diagnostics, novel disease gene discovery and identification of low frequency alleles in common disease and cancer. Advancement of exome sequencing in the clinical diagnostic arena has required technical innovations to ensure optimal DNA sequence coverage of clinically relevant genes, while maintaining overall high-quality, together with optimal speed and cost. To improve upon our previous generation of exome-sequencing reagents and methods, we generated a new exome capture design (HGSC-ClinExD) that targets 36.8Mb GENCODE and RefSeq coding regions on the HG38 reference sequence, utilizing TWIST capture technology. In addition to targeting gene coding regions, we incorporated clinically relevant non-CDS regions such as ClinVar pathogenic or likely pathogenic sites, COSMIC coding point mutations, pharmacogenomic sites, *TERT* promoter and key cardiovascular PRS sites. Exome sequencing performance using this new design was assessed utilizing 470 KAPA HiFi libraries from control, research consented and de-identified clinical samples. A subset of 210 samples with known pathogenic and benign variants (SNVs, Indel, CNVs) in genes with a range of genomic context (i.e., *PMS2*, *GBA*) were used to establish sensitivity, specificity, and reproducibility for clinical validation. Libraries were pooled (10-plex) for capture overnight at 70°C and sequenced as 70-plex in one lane of the Illumina NovaSeq S4 flowcell. This capture and sequencing format resulted in an average 12 Gbs of raw data generated for individual samples and >99% of the total targeted bases sequenced to a minimum depth of 20-fold with 65% of the reads mapping to target regions. Titration analyses showed that with only 5Gbs of data (40X average coverage) this design will achieve >95% of the target bases at 20X coverage, allowing for increased plexing of up to 130 samples per Illumina NovaSeq S4 lane and significantly decreasing sequencing costs. A survey of low coverage bases (<20X coverage) showed that the majority (53%) are in regions of <10bp encompassing a full range GC content. These results, along with the clinical validation data, demonstrate that the HGSC-ClinExD design has unprecedented utility for variant determination, novel gene discovery, deep somatic sequencing and orthogonal validation. Incorporation of the reagent in routine applications in both clinical and research settings enables comprehensive inclusion of clinical targets and general cost effectiveness, without loss of performance.

Omics Technologies Posters - Wednesday

PB2989. Extensive differential cell type-specific gene expression and regulation by sex in human skeletal muscle.

Authors:

S. Hanks¹, D. Ciotlos¹, A. Varshney¹, M. R. Erdos², N. Manickam¹, A. U. Jackson¹, J. Okamoto¹, H. M. Stringham¹, P. Orchard¹, N. Narisu², L. Bonnycastle², M. D. Sweeney¹, M. Laakso³, J. Tuomilehto⁴, T. A. Lakka³, K. L. Mohlke⁵, M. Boehnke¹, H. Koistinen⁴, F. S. Collins², S. C. J. Parker¹, L. J. Scott¹; ¹Univ. of Michigan, Ann Arbor, MI, ²NHGRI/NIH, Bethesda, MD, ³Univ. of Eastern Finland, Kuopio, Finland, ⁴Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, ⁵Univ. of North Carolina, Chapel Hill, NC

Abstract Body:

Human skeletal muscle exhibits sex differences in its size, composition, and physiology that may contribute to disparities in muscle-related diseases such as type 2 diabetes, obesity, and osteoporosis. Here, we quantify sex differences in cell type composition, cell type specific gene expression and chromatin accessibility, and miRNA abundances to better understand biological variation in this tissue. We profiled single nucleus-resolution gene expression (snRNA-seq, n=287) and chromatin accessibility (snATAC-seq, n=287), as well as bulk miRNA expression (miRNA-seq, n=290), in vastus lateralis skeletal muscle biopsied from Finnish individuals. We jointly clustered snATAC-seq and snRNA-seq nuclei into 13 cell types, including 3 main muscle fiber types (Type 1, 2A and 2X). We tested for sex differences in cell type composition using a negative binomial model, and found being female was associated with more non-nerve neuronal ($p=2 \times 10^{-6}$) and Type 1 slow twitch oxidative fiber nuclei ($p=2 \times 10^{-5}$) and fewer Type 2X fast twitch glycolytic fiber nuclei ($p=3 \times 10^{-12}$). We tested for sex differences in gene expression and found that 11-15% of genes showed significant differences by sex in the Type 1, 2A and 2X fiber types (FDR<5%). Gene set enrichment analyses showed that genes in mitochondrial activity and energy metabolism pathways were more highly expressed in males than females in all three fiber types. Similarly, we found that 13-32% of ATAC-seq peaks were differentially accessible (DA) by sex in the muscle fiber types. We tested for enrichment of autosome DA peaks in muscle chromatin states and found the strongest enrichment the active enhancers (1.4-3.0 fold for sex differences ($p < 1 \times 10^{-100}$)). We then tested for an enrichment of DA peaks in promoters (<1 kb from TSS) of differentially expressed genes. The odds of differential expression by sex increased 2.4-3.4X ($p < 1 \times 10^{-25}$) for each additional DA promoter ATAC-seq peak in the same direction, indicating potential sex differences in transcriptional regulation. We next tested for sex differences in miRNA expression levels in bulk muscle as miRNAs regulate gene expression post-transcriptionally. We found that 359/1,049 (34%) miRNAs are differentially expressed by sex, indicating potential differences by sex in post-transcriptional regulation. Overall, we found extensive differences in gene expression by sex at the cell type level and potential evidence of sex-based regulation at both the transcriptional and post-transcriptional levels. These findings may help uncover the mechanisms of sex differences in muscle physiology and disease.

Omics Technologies Posters - Thursday

PB2990. Extracellular protein monitoring in the ResolveOME genomic and transcriptomic dual workflow to uncover cancer pathology mechanisms in single cells

Authors:

T. Morozova¹, V. Weigman¹, J. Blackinton¹, J. Croteau², A. Fernandes², K. Taylor², M. Fabani³, J. West¹, G. Harton¹, J. Zawistowski¹; ¹BioSkryb Genomics, Durham, NC, ²BioLegend Inc, San Diego, CA, ³Singular Genomics Systems, San Diego, CA

Abstract Body:

Cancer is a disease of remarkable cell heterogeneity and is driven by complex, interconnected omic tiers. Single cell research has become an instrumental method in interrogating these multiple levels, as the bulk sequencing does not have enough resolution to reveal the true heterogeneity of the samples. The ResolveOME workflow aids understanding of cell-to-cell heterogeneity by providing unified genomics and transcriptomic information—including assessment of genome-wide single nucleotide variation, copy number changes, regulatory variants, splice isoform variation and cell state changes from the same single cell. Here, we have adapted the existing ResolveOME genomic/transcriptomic workflow to include the detection of cell-surface protein expression using the TotalSeqTM-A Human T-cell, B-cell, Natural Killer (TBNK) panel (BioLegend) of antibody-conjugated oligonucleotides. Primary peripheral blood mononuclear cells (PBMCs) were processed using the ResolveOME workflow which unifies template-switching single-cell RNAseq chemistry and Primary Template-directed Amplification (PTA) for whole genome amplification (WGA). We incorporated the TotalSeqTM-A antibody-oligo cocktail into this workflow, whereby both cellular mRNAs and antibody-derived oligos specifically bound to PBMC antigens anneal to oligo dT primers, followed by template switch-based reverse transcription. This resulted in the creation of first-strand cDNA molecules and antibody-derived tag molecules that could be affinity purified and pre-amplified following whole genome amplification by PTA. Separate protein+RNA and DNA fractions are processed as distinct library preparations at the end of the workflow. Sequencing was performed with the novel G4 Sequencing Platform from Singular Genomics. Using this strategy with PBMC cells, we detected all nine barcode IDs corresponding to the specific antibodies in the TotalSeqTM-A panel. The total number of detected barcodes varied between antibodies as well as between single cells, whereby CD19, CD3, CD4, CD16 and CD56 showed the most significant variation between individual single cells. We are extending these analyses to other cellular systems as well as to primary cancer cells, using antibody-conjugated antibody panels tailored to the application. Our study has devised a method for simultaneous detection of three omic tiers: whole genomic and transcriptomic signatures coupled to the assessment of defined protein panels at the individual cell level, empowering insights into the interplay between the three tiers not possible in isolation.

Omics Technologies Posters - Wednesday

PB2991. faiGP: an evolutionary approach to discover governing equations in high-dimensional genomic data

Authors:

S. Razavi¹, E. Gamazon²; ¹vanderbilt Univ. Med. center, nashville, TN, ²VUMC Clare Hall, Univ. of Cambridge, Nashville, TN

Abstract Body:

Here, we develop an integrative approach incorporating a convolutional variational autoencoder (for dimensionality reduction to reduce the search space), a bayesian multilabel classifier (for posterior inference and generation of the prior to guide the search), and a genetic programmer (for evolution of programs towards optimal fitness), to extract governing equations from data. We demonstrate an application of the framework in a theoretical account of ligand-receptor interactions with immediate relevance to transcription factor binding to regulatory DNA sequence. We propose a model of ligand-receptor binding kinetics that is generated from Hill dose-response experimental observations, providing a transcription-factor-mediated model of gene expression. Finally, the development of a model of transcription factor regulatory range with high fidelity to the original data, as an application, suggests the framework can facilitate discovery of governing equations in high-dimensional genomic data.

Omics Technologies Posters - Thursday

PB2992. Fast and gentle microfluidic cell sorting upstream of single cell transcriptomics

Authors:

M. Ciarlo¹, E. Rodriguez-Mesa², R. Barhouma², V. Tran³, A. Gadkari³, J. Musmacker³; ¹Miltenyi Biotec, San Diego, CA, ²Miltenyi Biotec, Santa Barbara, CA, ³Parse BioSci.s, Seattle, WA

Abstract Body:

While single-cell RNA sequencing (scRNA-seq) has become a widely used technique across many disciplines over the last decade, the high cost of sequencing, time-consuming cell preparation protocols, and extensive data analysis requirements have posed challenges to mainstream adoption. One way to reduce the cost and save time is to isolate the cell type of interest before performing scRNA-seq using cell-sorting techniques such as jet-in-air/droplet sorters or MACS® Magnetic Cell Sorting Technology. The challenge is that any upstream manipulation of the cells can lead to RNA degradation and perturbations in the transcriptome as well as the generation of cellular debris and/or dead cells which will negatively affect the quality of the transcriptomic data. Therefore, it is imperative to minimize sample preparation time and handle the cells as gently as possible. Jet-in-air/droplet sorters expose the cells to harsh conditions, such as high pressure, long fluidics pathways, electrostatic charges, and lengthy processing times. While the MACS Cell Separation Portfolio provides gentle, fast, and low parameter cell separation, the technology is not always sufficient for the isolation of specific cell types. This study demonstrates a multiparameter, fast, and gentle workflow for cell sorting with downstream scRNA-seq. In this example, the MACSQuant® Tyto® Cell Sorter, a microfluidic cartridge and microchip-based sorting technology, isolates B cells, CD16+ natural killer (NK) cells, and regulatory T cells (Tregs) from human peripheral blood as well as neurons, astrocytes and microglia from dissociated adult mouse brain. With this sorting technique, highly pure populations of distinct cell subtypes are obtained quickly while still maintaining high cell viability. Sorted cells are then analyzed via scRNA-seq using Evercode whole transcriptome technology allowing the profiling of single-cell whole transcriptomes from a bulk-sorted sample on any NGS device — eliminating the need for a droplet-based sequencing instrument. Sorted and unsorted samples were compared by looking at the number of reads obtained per cell as well as clustering of the data. This analysis showed a high number of reads per cell as well as distinct cell clusters of the sorted populations that overlaid directly on the clusters from the unsorted sample demonstrating that this method of sorting had no significant effect on the cell's transcriptome. This approachable and time-efficient process of combining microfluidic cell sorting with single-cell transcriptomics can be easily applied to other cell and tissues types and used in a broad range of applications.

Omics Technologies Posters - Wednesday

PB2993. FastRNA: an efficient exact solution for PCA of single-cell RNA sequencing data based on a batch-accounting count model

Authors:

B. Han, H. Lee; Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Abstract Body:

[Background] The analysis of single cell RNA sequencing (scRNA-seq) data almost always begins with the generation of the low dimensional embedding of the data by principal component analysis (PCA). Because PCA requires normally distributed data, researchers have traditionally applied a log-normal transformation to the count data of scRNA-seq. However, log-normalization can induce an artificial bias to the data. Studies have shown that the log-normalization can exaggerate the effects of the small counts. As a result, the final clustering can be in low quality in terms of the resolution. **[Challenge]** To address this bias, recent studies have developed methods that directly assume a count model. In these methods, a Poisson or negative binomial distribution is assumed for the observed count, and then normally-distributed residuals of the counts are calculated. The most popular method is scTransform (Genome Biology 2019), and a more recent method is the analytic Pearson residual (Genome Biology 2021). These methods were superior to naïve normalization in both theoretical and practical perspectives. However, they are extremely slow for large datasets, often taking more than several years of computation time. For this reason, for large datasets, naïve normalization is still being widely used. **[Solution]** Here we present FastRNA, an efficient exact solution for PCA of scRNA-seq data based on a full count model accounting for both batches and cell size factors. We assume the same general count model as previous methods, but we obtain the exact solution using two orders of magnitude less time and three orders of magnitude less memory than other methods. This achievement results from our unique algebraic strategy that utilizes several key observations, such as (A) that the PCA of the count residual matrix is equivalent to the PCA of the covariance of that matrix, (B) that the latter covariance matrix can be decomposed into a sum of sparse matrices, and (C) that the batch-wise count summation can serve as a sufficient statistic for calculation of batch-corrected residuals. These unique algebraic ideas, as a combination, work seamlessly together to completely remove the need for storing the actual residual matrix in memory. Importantly, our solution does not use an approximation but provides an exact solution without any sacrifice of the precision. Generating a batch-accounted PC of an atlas-scale dataset with 2 million cells takes less than 1 minute and 1GB memory using our method. By allowing the analysis of the largest data inside an ordinary laptop environment, we believe that FastRNA can revolutionize scRNA-seq analysis.

Omics Technologies Posters - Thursday

PB2994. FedExosome: mitochondrial DNA & microRNA profiling of neuronal-derived packages in Alzheimer's disease

Authors:

C. Hall, I. Garlotte, K. Subasinghe, R. Barber, N. Phillips; UNTHSC, Fort Worth, TX

Abstract Body:

Background: Alzheimer's disease (AD) is a complex neurodegenerative disorder characterized by impaired memory, diminished cognitive abilities, and accumulation of amyloid beta plaques and hyperphosphorylated tau tangles in the brain. AD progression is insidious, with the earliest clinical symptoms appearing decades after key neuropathological changes have occurred. Accurate diagnosis during the transitional period between disease onset and clinical manifestation could provide critical insight into AD pathogenesis and enable development of effective therapeutic strategies to prevent irreversible neuronal death. Evidence suggests that early alterations in the AD brain (oxidative stress, neuroinflammation, aberrant gene expression) can propagate to local and distal cells through the biological packages secreted by neurons. These neuronal-derived exosomes (NDEs) can cross the blood-brain barrier and are thus capable of mediating systemic inflammation in response to CNS-sourced stress through their mitochondrial DNA (mtDNA) and microRNA (miRNA) cargo. Circulating NDEs may serve as an easily accessible, early indicator of the forthcoming neuropathological changes in the AD brain. This study aims to develop a high-throughput workflow for profiling mtDNA and miRNA in plasma NDEs. **Methods:** Human plasma samples (500 μ L) were processed in duplicate using a two-step method involving (1) precipitation of total exosomes with ExoQuick (SBI) and (2) immunocapture of NDEs with a biotinylated antibody against the well-established neuronal surface marker, CD171. Intact NDEs were characterized via nanoparticle tracking analysis and flow cytometry. DNA was then extracted with the DNeasy blood & tissue kit (Qiagen) and mtDNA load was estimated using real-time qPCR. RNA was extracted with the EVeryRNA (SBI) and miRNeasy (Qiagen) purification systems and prepared for sequencing on the Illumina NextSeq 550 using the QIAseq miRNA library kit (Qiagen). **Results:** The protocol modifications implemented in this project were compared based on particle size, distribution, and purity as well as absolute mtDNA load and spike-in normalized miRNA read counts. These data not only demonstrate the feasibility of profiling the mtDNA and miRNA from 500 μ L of plasma NDEs but have also resulted in a high-throughput workflow for processing patient samples. **Conclusion:** This work represents the first successful attempt to simultaneously quantify the mtDNA and sequence the miRNAs in the relatively small subpopulation of plasma exosomes that originate from neurons. Future studies will harness the protocol developed herein to assess the biomarker potential of NDEs in AD.

Omics Technologies Posters - Wednesday

PB2995*. Fixing falsely duplicated and collapsed regions of the GRCh38 reference genome

Authors:

S. Behera¹, J. LeFaive², P. Orchard³, M. Mahmoud¹, J. Farek¹, D. Soto⁴, A. Smith³, M. Dennis⁴, J. Zook⁵, F. Sedlazeck¹; ¹Baylor Coll. of Med., Houston, TX, ²Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, ³Univ. of Michigan, Ann Arbor, MI, ⁴Univ. of California Davis, Davis, CA, ⁵Natl. Inst. of Standards and Technology, Gaithersburg, MD

Abstract Body:

The GRCh38 reference is the current standard in human genomics research. However, the current version of GRCh38 (p.13) contains a number of errors including 1.2 Mbp of falsely duplicated and 8.04 Mbp of collapsed (e.g., missing copies of segmental duplication) regions. This impact variant calls for 33 protein-coding genes, including 12 with medical relevance such as *KCNE1*, *CBS*, and *MAP2K3*. We generated a modified GRCh38 reference to correct these errors by masking the falsely duplicated regions and adding decoy sequences of falsely collapsed regions. To avoid re-analyzing data for the entire genome, we developed an efficient tool, FixItFelix, that extracts sequences from impacted regions of an existing BAM/CRAM and then remaps to the modified GRCh38 within 5~8 minutes. We evaluated the performance of our approach using the HG002 sample with the GIAB challenging medically relevant gene (CMRG) benchmark set and demonstrated improved accuracy of variant calling with the modified GRCh38. For falsely duplicated regions, a perfect recall score (1.0) with a precision of 0.96 was achieved for both SNV and indel call (0.007 and 0 respectively for GRCh38). For the falsely collapsed regions, remapping produced mainly improved precision by removing false variants caused by mismapped reads. Expanding our benchmark to genes not included in the CMRG set using variant calls from dipcall of phased HG002 assemblies, we also validated improved performance in these genes using the modified reference. Similar improvements were observed for SNV callings in exon regions when using the modified reference with HG002 whole exome and RNA sequencing data. Expanding our analysis beyond HG002, we used eight samples of different ancestries from the T2T Diversity panel— generating dipcall-based benchmark sets— and observed improvements in variant calling. To explore the impact of this new reference beyond variant calling, we performed cis-eQTL analysis using 449 lymphoblastoid cell line RNA-Seq datasets from the 1000 Genomes Project using either GRCh38 or the modified GRCh38 and showed that GRCh38 shows several apparently artifactual eQTL signals in the falsely duplicated regions. We applied FixItFelix to sequence data of 4,174 publicly available samples and showed that the mapping qualities for the erroneous regions were consistent with the evaluations for HG002. In summary, we present a modified GRCh38 reference that corrects errors while maintaining the same coordinates allowing us to leverage the extensive existing annotations of GRCh38, along with an efficient remapping approach that enables quick and efficient re-analysis of genomes to gain improved insights from existing data.

Omics Technologies Posters - Thursday

PB2996. Functional characterisation of the amyotrophic lateral sclerosis risk locus *GPX3/TNIP1*

Authors:

F. Garton¹, R. Restuadi¹, F. Steyn¹, E. Kabashi², S. Ngo¹, M. Nabais¹, M. Thompson³, T. Qi⁴, Y. Wu¹, A. Henders¹, L. Wallace¹, C. Bye⁵, B. Turner⁵, S. Mathers⁶, P. McCombe⁷, M. Needham⁸, D. Schultz⁹, M. Kiernan¹⁰, W. van Rheenen¹¹, L. van den Berg¹², J. Veldink¹³, R. Ophoff¹⁴, A. Gusev¹⁵, N. Zaitlen¹⁶, A. McRae¹, R. Henderson¹⁷, N. Wray⁴, J. Giacomotto¹⁷; ¹Univ. of Queensland, Brisbane, Australia, ²Imagine Inst., Inst. Natl. de la Santé et de la Recherche Médicale (INSERM) Unité 1163, Paris Descartes Université, Paris, France, ³UCLA, Los Angeles, CA, ⁴The Univ. of Queensland, Brisbane, Australia, ⁵Florey Inst. for NeuroSci. and Mental Hlth., Univ. of Melbourne, Melbourne, Australia, ⁶Calvary Hlth.Care Bethlehem, Parkdale, VIC 3195, Australia, Melbourne, Australia, ⁷Ctr. for Clinical Res., The Univ. of Queensland, Brisbane, Australia, ⁸Fiona Stanley Hosp., Perth, Australia, ⁹Neurology Dept. at Flinders Med. Ctr., South Australia, Adelaide, Australia, ¹⁰Brain & Mind Ctr., Univ. of Sydney, Inst. of Clinical NeuroSci.s, Royal Prince Alfred Hosp., Sydney, Australia, ¹¹Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, ¹²Univ. Hosp. Utrecht, Utrecht, Utrecht, Netherlands, ¹³Univ Med Ctr Utrecht, Utrecht, Netherlands, ¹⁴Univ California Los Angeles, Los Angeles, CA, ¹⁵Dept. of Med. Oncology, Dana-Farber Cancer Inst. and Harvard Med. Sch., Boston, MA, ¹⁶Dept. of Computer Sci., Univ. of California Los Angeles, Los Angeles, CA, ¹⁷Queensland Brain Inst., The Univ. of Queensland, Brisbane, Australia

Abstract Body:

Background: Amyotrophic lateral sclerosis (ALS) is a complex, late-onset, neurodegenerative disease with a genetic contribution to disease liability. Genome-wide association studies (GWAS) have identified ten risk loci to date, including the *TNIP1/GPX3* locus on chromosome five. Given association analysis data alone cannot determine the most plausible risk gene for this locus, we undertook a comprehensive suite of in silico, in vivo and in vitro studies to address this. **Methods:** The Functional Mapping and Annotation (FUMA) pipeline and five tools (conditional and joint analysis (GCTA-COJO), Stratified Linkage Disequilibrium Score Regression (S-LDSC), Polygenic Priority Scoring (PoPS), Summary-based Mendelian Randomisation (SMR-HEIDI) and transcriptome-wide association study (TWAS) analyses) were used to perform bioinformatic integration of GWAS data ($N_{\text{cases}} = 20,806$, $N_{\text{controls}} = 59,804$) with ‘omics reference datasets including the blood (eQTLgen consortium $N = 31,684$) and brain ($N = 2,581$). This was followed up by specific expression studies in ALS case-control cohorts (microarray $N_{\text{total}} = 942$, protein $N_{\text{total}} = 300$) and gene knockdown (KD) studies of human neuronal iPSC cells and zebrafish-morpholinos (MO). **Results:** SMR analyses implicated both *TNIP1* and *GPX3* ($p < 1.15 \times 10^{-6}$), but there was no simple SNP/expression relationship. Integrating multiple datasets using PoPS supported *GPX3* but not *TNIP1*. In vivo expression analyses from blood in ALS cases identified that lower *GPX3* expression correlated with a more progressed disease (ALS functional rating score, $p = 5.5 \times 10^{-3}$, adjusted $R^2 = 0.042$, $B_{\text{effect}} = 27.4 \pm 13.3$ ng/ml/ALSFRS unit) with microarray and protein data suggesting lower expression with risk allele (recessive model $p = 0.06$, $p = 0.02$ respectively). Validation in vivo indicated *gpx3* KD caused significant motor deficits in zebrafish-MO (mean difference vs. control \pm 95% CI, vs. control, swim distance = 112 ± 28 mm, time = 1.29 ± 0.59 s, speed = 32.0 ± 2.53 mm/s, respectively, p for all < 0.0001), which were rescued with *gpx3* expression, with no phenotype identified with *tnip1* KD or *gpx3* overexpression. **Conclusions:** These results support *GPX3* as a lead ALS risk gene in this locus, with more data needed to confirm/reject a role for *TNIP1*. This has implications for understanding disease mechanisms (*GPX3* acts in the same pathway as *SOD1*, a well-established ALS-associated gene) and identifying new therapeutic approaches. Few previous examples of in-depth

ASHG 2022 Annual Meeting Poster Abstracts

investigations of risk loci in ALS exist and a similar approach could be applied to investigate future expected GWAS findings.

Omics Technologies Posters - Wednesday

PB2997. Gene regulatory network inference using single-cell multiome ATAC-seq and RNA-seq data.

Authors:

Y. Wang¹, K. Chen², Z. Cai³, H. Zhao¹; ¹Yale Univ. Sch. of Publ. Hlth., New Haven, CT, ²Univ. of Sci. and Technology of China, Hefei, China, ³Nanjing Univ., Nanjing, China

Abstract Body:

Gene regulatory network (GRN) inference has been a major topic in the field of computational biology and bioinformatics. Constructing GRNs is crucial to understanding regulatory relationships between transcription factors (TFs) and their target genes under different biological conditions. Previous GRN inference methods mainly use transcriptomic data, and many of them have been proposed to analyze RNA sequencing (RNA-seq) data. Popular methods such as GENIE3 and SCENIC that rely on regression trees have shown success in GRN inference using bulk RNA-seq and single-cell RNA-seq (scRNA-seq) data, respectively.

Although inferring GRNs from transcriptomic data has been widely adopted, it has been shown that even the top methods' performance was moderate. This may result from the inherent shortcomings to the assumption that regulatory interactions can be fully extracted from expression patterns. However, gene regulation does not only involve the expression of necessary TFs, but also other factors such as the accessibility of important chromatin regions (promoters and enhancers) around the transcription starting (TSS) site of a gene. In recent years, the emergence of new sequencing technologies has enabled the measurement of expression levels by scRNA-seq and chromatin accessibility by scATAC-seq at the same time. Such co-assayed data provide new opportunities for us to infer more accurate GRNs.

In our presentation, we will introduce a new framework for GRN inference using single-cell multiome ATAC-seq and RNA-seq data. We first collect ATAC peaks within 500 kb around the TSS of a target gene, and pair TFs with candidate ATAC peaks by applying FIMO, a TF binding site analysis tool. Next, we run GENIE3, a random forest algorithm, to build a model between TF-ATAC pairs and the target gene using the multiome data. Last, the weight for each TF is calculated as the summation of the importance scores of all TF-ATAC pairs that involve the TF. We analyze a single-cell multiome data from a healthy donor's blood sample and benchmark the performance of our framework with the method that is based on expression data only using GRNs from existing databases (DoRothEA, TRRUST and RegNetwork) as the ground truth. We show that our model infers GRN with higher precision and AUPRC values. We also extend our approach by considering the strength of TF binding on an ATAC peak, which can further improve GRN inference. Overall, our proposed method can effectively use the ATAC data for more accurate GRN inference.

Omics Technologies Posters - Thursday

PB2998. Generating long-range sequencing information without long-read sequencing

Authors:

Z. Chen, L. Pham, V. Mikhaylova, T-C. Wu, G. Mo, P. Chang, Y. Xia, C. Heberling, I. Bassets, Y. Wang; Universal Sequencing Technology Corp, Carlsbad, CA

Abstract Body:

Long-range sequencing information is required for de novo assembly, haplotype phasing and structural variation detection. Current long-read sequencing technologies can provide valuable long-range information but at high cost or low accuracy and require large amount of DNA input. How can we use highly accurate and relatively low-cost short reads to generate long-range sequencing data? We achieve it by using TELL-SeqTM linked read technology, which enables a short-read second-generation sequencer to generate over 100 kb long-range sequencing information with as little as 0.1 ng input material. In a PCR tube, millions of clonally barcoded beads are used to uniquely barcode long DNA molecules and generate sequencing-ready library in 3 hours. Using TELL-LinkTM assembler with these barcoded linked reads produced by a short read sequencer, we successfully assembled many microbes into one scaffold at high accuracy with 0.1 - 0.5ng input DNA. Other sizes of genomes ranging from megabases to gigabases were able to be de novo sequenced with 0.5ng to 5ng genomic DNA input. In addition, TELL-Seq reads were used for scaffolding of other long read sequencing assemblies and resulted in significantly improved assembly for various samples, such as, insects, marine organisms and humans. TELL-Seq reads also provide excellent haplotype phasing capability. We constructed whole genome sequencing linked read libraries using 5ng high molecular weight genomic DNA from GM12878 and GM24385 cells and sequenced them on a NovaSeq system. Analyzing TELL-Seq reads with TELL-SortTM phasing tool, we phased NA12878 and NA24385 human genomes into megabase long phased blocks with N50 phased block size up to 15Mb long. A notable example was that entire 4Mb Major Histocompatibility Complex (MHC) locus was phased into one phase block. Besides whole genome scale phasing, we demonstrate excellent phasing results on long-range PCR amplicons and enriched genomic regions ranging from 2kb to 200kb using linked reads. Furthermore, for the very first time we show that TELL-Seq is capable to generate average 40kb long-range sequencing information from FFPE samples and provide the first “long-read” approach to investigate this widely used clinical sample type.

Omics Technologies Posters - Wednesday

PB2999. Generative Adversarial Networks Create High-Quality Synthetic Prostate Cancer Magnetic Resonance Images.

Authors:

B. Ledesma¹, I. Xu¹, D. Van Booven¹, S. Goberdhan², S. Punnen¹, A. Mahne², R. Stoyanova¹, H. Arora¹; ¹Univ. of Miami, Miami, FL, ²Univ. of Central Florida, Orlando, FL

Abstract Body:

Introduction. The recent integration of open-source data to machine learning models, especially in the medical field, has opened new doors to studying disease progression and/or regression. However, one limitation of the medical data for machine learning approaches is the insufficient quantity and quality of data for a particular medical condition. In this context, synthetic data augmentation by using generative adversarial networks (GAN), could potentially generate high-quality data that preserve clinical variability. **Methods.** A GAN pipeline to curate synthetic prostate cancer magnetic resonant images (MRIs) was conducted. A total of 139 T2-weighted prostate images from public sources were used in an unsupervised GAN called SinGAN, to create synthetic prostate MRI images from a single image. Synthetic images with a high-level segmentation boundary of the prostate were filtered and used in the quality control assessment by a different machine learning segmentation pipeline. To assess final quality, participating scientists with varying levels of experience (more than 10 years, 1 year, or no experience) to work with MRI images were asked to study 60 MRI images and evaluate if they were conventional or synthetic images. **Results.** The most experienced group correctly identified conventional vs synthetic images with 67% accuracy, the group with 1 year of experience correctly identified the images with 58% accuracy, and the group with no prior experience reached 50% accuracy. Nearly half (47%) of the synthetic images were mistakenly evaluated as conventional images. Interestingly, the blinded quality assessment by a board-certified radiologist to differentiate conventional and synthetic images was not significantly different in the context of the mean quality of synthetic and conventional images. **Conclusion.** SinGAN model creates MRI images that are similar enough to the acquired MR images to be indistinguishable in some cases. This study shows promise that high-quality synthetic images from MRI can be generated using GAN. Such an AI model may contribute significantly to various clinical applications which involve supervised machine learning approach.

Omics Technologies Posters - Wednesday

PB3001*. Genetics guided approach to infer unobserved covariates in functional genomics data

Authors:

R. Yamamoto¹, M. Thompson¹, N. Zaitlen²; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²UCLA, Los Angeles, CA

Abstract Body:

High dimensional datasets used in functional genomics studies are often confounded by technical and environmental variation that occur during high-throughput experiments. These unobserved confounders bias the results of analysis significantly. Therefore, it is common practice to use methods to estimate these covariates and account for such effects in analysis. Since these variations are not directly observed, most commonly used methods, such as PEER and SVA, aim to predict covariates that explain the largest proportion of variability for the data and regard them as confounding. However, the extent to which these predicted confounders affect actual biological variations is not well understood. In fact, some studies suggest that using covariates carelessly from these methods could remove signals of interest. In this study, we revisit the problem of hidden confounder prediction and propose a new approach to estimate covariates while maximizing the genetically derived variability. We formulate this approach in the context of eQTL detection studies. Our neural net-based model takes in the matrix of gene expression and known covariates and transforms them into a set of hidden covariates. Our loss function takes in predicted hidden covariates and we aim to maximize the absolute value of t-statistics for a subset of eQTLs that are known to have strong associations. Once the training reaches convergence, we use predicted hidden covariates to discover eQTLs that are not included during the training. We first show through simulations that our method outperforms PEER, state of art hidden covariates predictor. Our model yields more than two fold increase in the number of eQTL discovered compared to when using PEER (246 new eQTL discovered vs 89 new eQTL discovered using PEER across simulations). Our predicted hidden covariates also explain more variance for true hidden covariates than PEER factors (14% increase in R^2 ($H_{true} \sim H_{pred}$) overall). Applied to the GTEx consortium, we also observe a significant improvement in t-statistics of eQTLs compared to previous approaches. Together, we show that our novel approach to estimate hidden covariates provides a valuable resource to improve functional genomics studies.

Omics Technologies Posters - Thursday

PB3002. *Genetics of Cardiovascular diseases: Mathematical Modeling & Development of a Predicting Simulator Prediction of the heart attack, the treatment to prescribe and the risk of secondary complications*

Authors:

H. Hassani Idrissi¹, H. AKODAD², R. HABBAL³, S. NADIFI¹; ¹Univ. Hassan II Sch. of Med. and Pharmacy casablanca Morocco, Casablanca, Morocco, ²CHU Hassan II, Fes, Morocco, Fes, Morocco, ³CHU Ibnou Rochd, Casablanca, Morocco, Casablanca, Morocco

Abstract Body:

Objective: Cardiovascular diseases are among the leading causes of death worldwide, affecting both the costs of medical and healthcare services, as well as the quality of patients' life. Since the majority of these pathologies are the result of an occlusive thrombosis, the therapeutic strategy adopted is essentially based on the prescription of an antithrombotic agent. However, inter-individual variability has -for long- been observed in these diseases, in terms of response to treatment: patients fail to derive the same degree of benefit from the prescribed dose, and side effects are observed in many of them. Therefore, our pharmacogenomics study aims -not only- to identify the clinical and genetic factors modulating the response of CVD patients to treatment, but also tries to suggest a solution to improve the quality of care and treatment, and life expectancy of these patients. **Methodology:** To do so, the clinical data of our patients were collected after obtaining signed consents. We have -then-used- - **Reverse transcription RT-qPCR analysis** (objective: quantification of mRNA copies), - **Sequencing (DNA:** identification of eight novel mutations), - **PCR-RFLP** (genotyping of patients for SNPs that are already known) - **Molecular Modeling** using bioinformatics tools (protein: study of the effect of the identified novel mutations on the protein's function and structure), to identify the genetic factors that would be responsible for the resistance to treatment, the development of secondary complications, and predisposition to the disease. **Important results obtained** Based on these identified clinical and genetic risk factors, we have developed, through a Mathematical Modeling approach, mathematical models (equations) that we have integrated into a Digital Simulator, the first-of-its-kind, in terms of Medical Decision Support Systems (MDSS), in the form of a digital platform (smartphone application and website). We made sure that the latter is easy to access and manipulate by the treating physician, and highly sensitive and specific in terms of predictions. The ultimate aim of our study was to help predict the patient's behavior towards the therapeutic molecule before its administration, the risk of developing secondary effects, in addition to that of developing the heart attack among normal subjects (genetic counseling). It is also good to mention that the analyzed pathology (cardiovascular diseases) was just an example and that the followed approach can easily be used to predict the same events for any other pathology such as cancer.

Omics Technologies Posters - Wednesday

PB3003. Genome-wide prediction of chromatin profiles from gene expression

Authors:

J. Fu¹, S. Sabri¹, J. Ernst²; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²UCLA, Los Angeles, CA

Abstract Body:

High-throughput sequencing based methods for chromatin profiling have been developed to map epigenetic modifications such as histone methylation and acetylation. Large compendia of these maps have been accumulated and used for a range of biological questions including analyzing the contribution of non-coding genetic variants to disease. However, mapping such modifications in the context of limited number of cells or at the single cell level remains challenging. In some cases it is easier to acquire gene expression data. This motivates the development of computational methods that can impute maps chromatin marks by integrating compendia of chromatin and gene expression data into new cell types with only gene expression data available. Here we present a regression based method that makes predictions based on integrating features defined based on both expression data of nearby genes within the cell type and the relationship of chromatin mark signal at the target position with gene expression across cell types. We applied and evaluated the method using a compendia of paired bulk RNA-seq and H3K27ac data in human and mouse. We compare this method against other related methods. We also demonstrate the potential application of this method in the deconvolution of bulk chromatin profiles into cell type specific chromatin profiles in the presence of single cell RNA-seq data.

Omics Technologies Posters - Thursday

PB3004*. *Genome-wide prediction of pathogenic gain- and loss-of-function variants from ensemble learning of diverse feature set*

Authors:

Y. Itan¹, A. Schlessinger¹, **D. Stein**¹, C. Bayrak¹, y. wu¹, P. Stenson², D. Cooper³; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²Cardiff Univ., Cardiff, United Kingdom, ³Cardiff Univ, Cardiff, Wales, United Kingdom

Abstract Body:

Gain-of-function (GOF) variants yield increased or novel protein function while loss-of-function (LOF) variants yield diminished protein function. GOF and LOF variants can result in markedly varying phenotypes even when occurring in the same gene. Experimental approaches for identifying GOF and LOF are slow and costly, and computational tools cannot accurately discriminate between GOF and LOF variants. We developed LoGoFunc, the first computational genome-wide GOF and LOF classifier. LoGoFunc is an ensemble machine learning method for predicting pathogenic GOF, pathogenic LOF and neutral variants, trained on 672 protein, gene, variant, and network annotations describing diverse biological characteristics. We analyzed features in terms of protein structure and function, splice disruption, and phenotypic associations, revealing previously unreported relationships between various biological phenomena and variant functional outcomes. For example, GOF and LOF variants demonstrate contrasting enrichments in protein structural and functional regions, and LOF variants are more likely to disrupt canonical splicing as indicated by splicing-related features employed by the model. Furthermore, by performing phenome-wide association studies (PheWAS), we identified strong associations between relevant phenotypes and high-confidence predicted GOF and LOF variants. LoGoFunc performs well on an independent test set of GOF and LOF variants, and we provide precomputed genome-wide GOF and LOF predictions for 71,322,505 missense variants.

Omics Technologies Posters - Wednesday

PB3005. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases: an analysis of the Canadian Longitudinal Study on Aging (CLSA).

Authors:

Y. Chen^{1,2}, T. Lu^{1,2}, U. Pettersson-Kymmer³, I. D. Stewart⁴, G. Butler-Laporte^{1,2}, T. Nakanishi^{1,5,2}, A. Cerani^{1,2}, K. Liang^{1,2}, S. Yoshiji^{1,5,2}, J. D. S. Willett^{1,2}, C-Y. Su^{1,2}, P. Raina⁶, C. Greenwood², Y. Farjoun^{2,7,8,9}, V. Forgetta^{2,8}, C. Langenberg^{10,4}, S. Zhou², C. Ohlsson^{11,12}, J. B. Richards^{1,2,8,13}; ¹McGill Univ., Montreal, QC, Canada, ²Lady Davis Inst., Jewish Gen. Hosp., Montreal, QC, Canada, ³Umea Univ., Umea, Sweden, ⁴MRC Epidemiology Unit, Cambridge, United Kingdom, ⁵Kyoto Univ., Kyoto, Japan, ⁶McMaster Univ., Hamilton, ON, Canada, ⁷The Broad Inst. of Harvard and MIT, Cambridge, MA, ⁸5 Prime Sci. Inc., Montreal, QC, Canada, ⁹Fulcrum Genomics, Boulder, CO, ¹⁰Berlin Inst. of Hlth.(BIH) at Charité – Univ.smedizin Berlin, Berlin, Germany, ¹¹Univ. of Gothenburg, Gothenburg, Sweden, ¹²Sahlgrenska Univ. Hosp., Gothenburg, Sweden, ¹³King's Coll. London, London, United Kingdom

Abstract Body:

Metabolic processes can influence disease risk and provide therapeutic targets, yet their genetic determinants and role in human diseases are not well understood. Here, we conducted a series of large genome-wide association studies of metabolites, including 1,091 blood metabolites and 309 metabolite ratios in 8,299 individuals from the Canadian Longitudinal Study on Aging. We identified associations for 690 metabolites across 248 loci; and associations for 143 metabolite ratios arising from 69 loci. Integrating metabolite-gene and gene expression information identified 94 effector genes for 109 metabolites and 48 metabolite ratios. We used Mendelian Randomization (MR) to estimate the causal effect of selected metabolites and metabolite ratios on 12 traits and diseases influenced predominately by three mechanisms: aging, metabolism and immune response. MR identified 22 metabolites and 20 metabolite ratios for these 12 outcomes, which included orotate for estimated bone mineral density (eBMD), alpha-hydroxyisovalerate for body mass index and ergothioneine for two immune-related traits: inflammatory bowel disease and asthma. Since increased orotate was associated with lower eBMD in MR studies, we measured its level in a separate nested case-control cohort study for osteoporotic fracture and demonstrated that, as predicted by MR, increased orotate was associated with increased incident hip fractures. This study provides a valuable resource describing the genetic architecture of metabolites and delivers insights into their role in common disease, thereby offering opportunities for therapeutic targets.

Omics Technologies Posters - Thursday

PB3006. Genotyping known disease-associated short tandem repeats in 7,000 Project MinE genomes reveals their relation to ALS and assists in disease-association analysis of unknown short tandem repeats.

Authors:

J. van Vugt¹, R. Zwamborn¹, E. Dolzhenko², Project MinE Sequencing Consortium, M. A. Eberle², J. H. Veldink¹; ¹UMC Utrecht Brain Ctr., Utrecht, Netherlands, ²Illumina Inc., San Diego, CA

Abstract Body:

Short tandem repeats (STRs) are causal to many neurodegenerative disorders such as the *C9ORF72* repeat expansion in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Some STRs that are causal to a disease are also a risk factor for another disease. For example, the *ATXN2* repeat expansion is causal to spinocerebellar ataxia 2 and a risk factor for ALS. To see whether more disease-associated STRs are a risk factor for ALS, we have analyzed ~40 known STRs in 7,000 whole genomes of Project MinE. STRs were genotyped with ExpansionHunter (EH), because of the high genotyping accuracy and ability to genotype consecutive STRs. This genotyping accuracy was validated with PCR and Sanger sequencing of 5,680 samples in three STRs, which revealed concordance with the wet lab results of 98.3%. Though EH genotyping accuracy is very high, it is limited by the fragment length, read length and read depth, and can be obstructed by motif changes, motif interruptions, indels and multiallelic genotypes. We therefore created a workflow with a manual and automated part to assess the STR genotyping, based on the results of EH, REViewer and ExpansionHunterDeNovo (EHdenovo). The manual part entails visual inspection of 6,500 read pileup plots, which allows for identification of genotyping errors and, if possible, correcting these. The automated part involves the analysis of STR parameters from EH and REViewer, for example fragment length limitation and allelic read depth, and identification of STR motif changes with EHdenovo. Combining the automated and manual workflows allowed for complete STR genotype assessment. With EHdenovo we identified motif changes in *BEANI*, *DABI*, *DMPK*, *FXN*, and *RFC1* of which some motifs have not been reported before. The pathogenic and premutation frequencies of the disease-associated STRs in Project MinE genomes were compared to that described in literature. Some misdiagnoses were resolved (i.e. Kennedy's disease), which supports the importance of genetic screening upon disease diagnosis. The ALS association analysis was performed by comparing the repeat sizes between 5271 cases and 1756 controls. The only STRs that revealed a significant effect with ALS after multiple testing correction were *C9ORF72* and *ATXN2*, which are known ALS-associated STRs. Notably, the repeat expansion in *Huntingtin (HTT)* was not ALS-associated (5 cases and 1 control with 40 or more repeats). More than 37 thousand unknown STRs were analyzed in Project MinE using the automated workflow and more will follow. The read pileup plots of potential ALS-associated STRs were visually inspected to complete the genotype assessment and validate the authenticity of the ALS association.

Omics Technologies Posters - Wednesday

PB3007*. GestaltMatcher supports classification of ultra-rare disorders and delineation of novel syndromes by facial phenotype descriptors

Authors:

T-C. Hsieh¹, A. Bar-Haim², S. Moosa³, K. W. Gripp⁴, H. Lesmann⁵, J-T. Pantel¹, N. Fleischer², A. Hustinx¹, B. Javanmardi¹, L. Averdunk⁶, M. M. C. Chui⁷, C. C. Y. Mak⁷, B. H. Y. Chung⁷, F. Distelmaier⁶, P. Krawitz¹; ¹Inst. for Genomic Statistics and Bioinformatics, Univ. Hosp. of Bonn, Bonn, Germany, ²FDNA Inc., Boston, MA, ³Div. of Molecular Biology and Human Genetics, Stellenbosch Univ. and Med. Genetics, Tygerberg, South Africa, ⁴DuPont Hosp, Wilmington, DE, ⁵Inst. of Human Genetics, Univ. Bonn, Bonn, Germany, ⁶Dept. of Gen. Pediatrics, Neonatology and Pediatric Cardiology, Med. Faculty, Univ. Hosp., Heinrich-Heine-Univ., Düsseldorf, Germany, ⁷The Univ. of Hong Kong, Hong Kong, Hong Kong

Abstract Body:

Introduction: Many monogenic disorders cause a characteristic facial gestalt. Next-generation phenotyping (NGP) can support physicians in recognizing these patterns by associating facial phenotypes with the underlying syndrome through training on thousands of patient photographs. However, this supervised approach means that diagnoses are only possible if the disorder was part of the training set, which often excludes ultra-rare disorders due to a lack of data. To improve recognition of ultra-rare and novel disorders, we developed GestaltMatcher, an encoder for portraits that is based on a deep convolutional neural network. GestaltMatcher goes beyond existing NGP tools by supporting not only classification of known disorders but cluster analysis of any photograph. By this means, GestaltMatcher can be used to lump and split phenotypes.

Methods: We compiled a dataset consisting of 17,560 patients diagnosed with 1,115 different rare disorders. For each individual, a frontal photo and the molecularly confirmed diagnosis were available. A deep convolutional neural network was trained on patients' frontal photos and can be used to encode photos to facial phenotype descriptors. By this means, a photo is positioned in the Clinical Face Phenotype Space (CFPS), in which distances between images define syndromic similarity. We presented two cohorts for the lump and splitter analysis: Cohort-1 consisted of five patients with disease-causing mutations in *Gene-X*, and Cohort-2 contained 33 patients with the disease-causing mutations in *Gene-Y*.

Results: Within the CFPS, the patients with a similar phenotype were located in close proximity. Ranking syndromes by distance showed that patients could be matched to others with the same molecular diagnosis even when the disorder was not included in the training set. With cluster analysis in the CFPS, we demonstrated how GestaltMatcher can lump *Gene-X* into the phenotypic series of Rothmund-Thomson syndrome and split *Gene-Y* into two phenotypes. On the molecular level, the pathogenic mutations of *Gene-Y* are subgrouped into two different exons differing in NMD. Moreover, as a tool for matching patients of unknown cause, GestaltMatcher could be seen as the analogon to GeneMatcher on the phenotypic level, that not only accelerates the clinical diagnosis of patients with ultra-rare disorders but also enables the delineation of new phenotypes.

Conclusions: GestaltMatcher opens the door to extend the coverage of syndromes recognized by NGP tools and to further enable exploration of unknown phenotype-genotype associations.

Omics Technologies Posters - Thursday

PB3008*. GTAGMe-seq reveals the genome-wide long-range coordinated cis-regulatory elements at the single-molecule level

Authors:

Y. Liu¹, H. Fu¹, H. Zheng¹, L. Muglia², L. Wang¹; ¹Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, ²Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH

Abstract Body:

Cis-regulatory elements (CREs), such as promoters and enhancers, play instrumental roles in regulating mammalian gene expression. Sets of transcription factors (TFs) bring together multiple CREs, which can be up to megabases away, to initiate and maintain the expression of their target genes. How distal CREs coordinate together to regulate the downstream target genes' expression is one of the central questions in gene regulation. Detection and quantification of the binding events that co-occur at distal CREs require the ability to measure TFs-DNA interaction at the single-molecule level with awareness of spatial proximity. Here, we report a method, GTAGMe-seq, to simultaneously capture the genetic variants, DNA methylation, GpC methyltransferase footprints, and chromosome conformation changes from a single DNA molecule, together with transcriptome in the same assay. To benchmark the performance, we applied GTAGMe-seq to the IMR-90 and GM12878 cell lines and found the high concordances with the state-of-art mono-omics assays across all the molecular measurements at the same cell types. We found that the genomic regions linearly separated but positioned in proximity in three-dimensional topology showed coordinated chromatin accessibility at CREs. We further identified three groups of CREs with different long-range allele-specific activities (FDR<0.05): Group 1 (n=2,149) showed concordant allele-specific activities in both local (with heterozygous SNPs nearby) and distal regions (>20kb distant to linked SNPs but in spatial proximity). Group 2 (n=3,285) showed allele-specific activities only in local but not in distal linked regions. Group 3 (n=2,963) showed allele-specific activities only in distal but not in local regions. In genome-wide association studies (GWAS), the local imbalanced CREs activities linked with the risk or non-risk alleles are often utilized to prioritize the potential causal SNPs within the same linkage disequilibrium blocks. The long-range allele-specific activities in CREs from Groups 1 and 3 may suggest that the switching of risk and non-risk alleles can also be correlated with the activities of distant CREs but in spatial proximity. In inflammatory bowel diseases, we observed the overlap of Group 3 regions (B-lymphoblastoid cell) with three significant GWAS loci, which are all annotated as heterochromatin regions locally but with imbalance chromatin accessibility at linearly distal and spatial proximal regions. Thus, GTAGMe-seq here provides an interesting approach to linking non-coding genetic variants with the distal CREs' activities in spatial proximity at the same DNA molecules.

Omics Technologies Posters - Wednesday

PB3009. HD-Seq: A novel method for high accuracy sequencing.

Authors:

S. Shore, A. King, E. A. LaMarca, K. Gouin III, R. Schultzaberger, T. Looney, D. Witters, M. M. Fabani, E. N. Glezer; Singular Genomics, San Diego, CA

Abstract Body:

Background Next generation sequencing (NGS) has enabled rapid progress in cancer diagnostics. However, challenges remain around detection of somatic variants at allele frequencies similar or below the error rate of on-market DNA sequencers (typically $\sim 10^{-3}$) when using standard library preparation methodologies. Here, we introduce HD-Seq, a novel library preparation and sequencing method that allows for efficient and high accuracy sequencing, achieving error rates in the range of $\sim 10^{-5}$ - 10^{-6} . The HD-Seq method relies on the physical linkage of original dsDNA templates in a sample, a connection that is maintained throughout library preparation, clustering and sequencing. By requiring consensus between the sequences of the linked strands, errors present in one of the original strands or introduced during amplification are readily identified and eliminated. In addition, error correction efficiency is maximized by guaranteeing that complementary strands of each molecule are represented in the sequence output. **Methods** Library preparation for the HD-Seq method is similar to library preparation for standard libraries, except for use of two custom adapters and a purification. Typically, 50 ng of fragmented gDNA (150 bp mean size, Covaris) is end-repaired and A-tailed (Quantabio SparQ DNA library prep kit), ligated to Singular Genomics adapters, and purified. Libraries are then amplified with 10 cycles of PCR and purified followed by optional capture based enrichment. Paired end 2x150 bp sequencing occurs on the Singular Genomics G4 platform followed by consensus error correction. Briefly, paired reads are aligned to each other and non-overlapping portions of the read pair are trimmed. Discordant basecalls between R1 and R2 are assumed to reflect library preparation or sequencing errors and are masked (called 'N'). **Results** The typical mismatch rate observed in single pass (non-consensus) sequencing is approximately 10^{-3} . The HD-Seq method enables $\gt;100$ -fold lower error rates by building a consensus sequence from linked strands. Prior to HD-Seq correction, Read 1 and Read 2 from a representative HD-Seq experiment (Salmonella gDNA input) show mean accuracies of 99.76% and 99.82% , respectively. When reads are combined to build a consensus sequence, the accuracy increases to 99.9991% . **Conclusions** The HD-Seq approach is a novel library preparation and sequencing method that relies on linking the original paired strands in a dsDNA fragment. DNA errors affecting one of the two strands can be easily identified during sequencing of the corresponding linked strands. Through consensus sequencing, the HD-Seq technique enables accuracies of $\gt;Q50$.

Omics Technologies Posters - Thursday

PB3010. High accuracy Sequencing By Binding (SBB) improves human germline variant calling

Authors:

D. Nasko, C. Kingsley, N. Pezeshkian, W. Rowell; Pacific BioSci.s, Menlo Park, CA

Abstract Body:

PacBio Sequencing By Binding (SBB) is a short-read sequencing technology that provides a >10X increase in read accuracy relative to Sequencing By Synthesis (SBS). This new level of read accuracy brings opportunities to improve variant detection, but requires germline variant callers that take full advantage of the higher accuracy. Here we compare variant calling with SBB and SBS reads and develop an error model that is specific to SBB reads using the highly controlled and characterized Genome in a Bottle samples from the National Institute of Standards & Technology (NIST). PCR-free libraries were prepared for the reference samples HG001, HG002, HG003, HG004, and HG005 and sequenced to 30-fold depth each on the PacBio SBB sequencing technology. Reads were first mapped to the GRCh38 human reference genome and variants were called using GATK HaplotypeCaller and DeepVariant with its SBS model (DeepVariant-SBS). The SBB data was also used to train SBB-specific models for both DeepVariant (DeepVariant-SBB) and Sentieon DNAscope. With GATK Haplotype Caller and DeepVariant-SBS, the SBB callsets had higher precision but lower recall than SBS callsets for both SNVs and indels. In comparison, DeepVariant-SBB and Sentieon DNAscope trained on SBB data improve both precision and recall for SBB. This demonstrates that improved accuracy of SBB short reads translates into improved accuracy for variant calling, opening new applications in human genomics.

Omics Technologies Posters - Wednesday

PB3011. High throughput single-cell epigenomic profiling through split-pool combinatorial barcoding.

Authors:

Z. Sayar, A. Sova, R. Koehler, C. Roco, A. Rosenberg; Parse BioSci.s, SEATTLE, WA

Abstract Body:

Understanding the chromatin dynamics which regulate gene expression and cellular differentiation can elucidate mechanisms responsible for key biological processes. Current methods for profiling the open chromatin landscape of single cells require highly specialized equipment to compartmentalize individual cells and are expensive and limited in throughput. Here, we demonstrate an approach for high-throughput profiling of chromatin accessibility in single cells which integrates an Assay for Transposase-Accessible Chromatin-sequencing (ATAC-seq) with split-pool combinatorial barcoding. Our approach produces chromatin fragments of comparable quality and quantity to existing platforms, while offering the ability to scale up to hundreds of thousands of cells and dozens of samples in a single experiment. Furthermore, the workflow uses only common laboratory equipment, and allows for samples collected and fixed across different time points to be processed in the same experiment. To demonstrate the utility of this platform, we profiled fixed single nuclei isolated from multiple cell lines and mouse brain tissue. Our assay consistently produced high unique fragment counts per cell, enabling the identification of active gene regulatory networks associated with different cell types. Our resulting data yielded high transcription start site enrichment scores, while maintaining low mitochondrial fragment counts and low doublet rates across multiple species. Overall, our approach provides a flexible and accessible option for scalable single cell profiling of accessible chromatin.

Omics Technologies Posters - Thursday

PB3012. High throughput workflow for human whole genome sequencing using PacBio HiFi.

Authors:

J. Rocha¹, J. Burke¹, R. Fedak¹, D. Kilburn¹, E. B. Iglesias², D. Laubscher², B. Ottenwalder², S. Dee¹, H. Ferrao¹, K. Liu¹; ¹PacBio, MP, CA, ²Hamilton, Bonaduz, Switzerland

Abstract Body:

Improved throughput and cost of long-read sequencing, driven by recent technological advances, has opened the door for investigation of human whole genomes across progressively larger populations. To support the rapidly growing capabilities of long-read sequencing, high throughput sample and library preparation solutions are necessary. While short-read sequencing workflows are well established, long-read sequencing workflows for handling of high molecular weight (HMW) DNA are not widely available. We present a fully automated 96 well plate-based high throughput HMW DNA extraction, shearing, size selection, and library preparation workflow for processing human whole blood samples for PacBio HiFi sequencing. First, HMW DNA extraction is performed utilizing Nanobind magnetic disk technology on the fully automated Hamilton NIMBUS Presto. Nanobind disks feature micro- and nanostructured silica wrinkles to shield bound DNA from damage during extraction. This method is capable of generating ~6 µg of DNA on average per 200 µL blood sample on a 96 well plate in only 2.5 hours. Alternatively, the stand-alone Thermo Fisher KingFisher instruments provide a semi-automated option with comparable metrics. After extraction, size selection is performed to reduce small DNA from HMW samples. Here a high throughput version of Short Read Eliminator (SRE) technology is used. Size selection can also be achieved through magnetic bead-based methods instead. Prior to library preparation, HMW DNA is sheared down to 15-20 kb using an automated 96 well plate-based platform. Finally, automated PacBio library and loading preparation is performed on the fully automated Hamilton NGS STAR. The methods presented utilize standard configurations of commercially available Hamilton instruments and can easily be incorporated into existing workflows. Data is presented using human whole blood for a workflow which can prepare 96 samples from blood to library ready for loading in ~10 hours. A single SMRTcell typically generates ~10-12X coverage of high quality sequence data sufficient for analysis including phasing, 5mC, and variant calling. However, the protocols can be readily adapted to support a variety of other sample types including cultured cells, bacteria, and tissue.

Omics Technologies Posters - Wednesday

PB3013. High-throughput CRISPR Inhibition Screen to Improve Interpretation of Noncoding Variants in Developmental Epileptic Encephalopathies

Authors:

E. Almanza Fuerte; St Jude Children's Res. Hosp., Memphis, TN

Abstract Body:

Developmental and epileptic encephalopathies (DEEs) are a heterogeneous group of disorders characterized by severe, early-onset epilepsies, cognitive/developmental disabilities, and movement disorders. *De novo* mutations in coding regions are the most common cause for DEEs, identified by next-generation sequencing, however, these mutations only account for ~ 50% of DEE cases. We postulate the remaining undiagnosed cases may arise from mutations in the noncoding region that influence gene expression. Thus, a better understanding of the noncoding genome can help prioritize and interpret pathogenic variants. We are performing a high-throughput CRISPR-inhibition (CRISPRi) screen to identify non-coding regions and variants that regulate the expression of DEE-associated and neurodevelopmental genes. CRISPRi relies on catalytically deactivated Cas9 protein (dCas9), which is unable to cut DNA. dCas9 is linked to a repressive KRAB domain to bind to targeted DNA regions and provide steric hindrance, preventing the transcription of target genes. We identified 400 candidate genes for our full-scale CRISPRi screen. For each gene, we have identified regions of interest in noncoding regions surrounding (+/- 5kb) each candidate gene, including promoters, enhancers, putative enhancers, poison exons, and highly conserved regions. We designed multiple sgRNA for each candidate region, targeting both sense and antisense strands of DNA. We will leverage single-cell RNA sequencing to determine the consequences of CRISPRi perturbation. We will utilize a multiplex approach to introduce sgRNA into the cells, providing an unbiased and high-throughput approach to link regulatory regions to associated genes. We have completed our pilot screen targeting 20 DEE genes, including *CHD2*, *SCN1A*, and *SCN2A*. In total, we are targeting 20 promoter regions, 364 enhancer and putative enhancer regions and 151 highly conserved regions based on GERP score, with 3807 sgRNA. We included 25 positive control sgRNA against TSS of both DEE and non-DEE genes and 415 non-targeting controls. Currently, we are analyzing results from our pilot screen to identify differential gene expression corresponding to sgRNA influence. We will utilize data from this proof-of-concept screen to optimize the full-scale screen. Our high throughput CRISPRi screen will identify noncoding regions and variants that regulate the expression of 400 DEE-associated and neurodevelopmental genes. These findings will provide insights into variants identified by whole-genome sequencing in patients in uncovering underlying disease-causing mutations.

Omics Technologies Posters - Thursday

PB3014. High-throughput metabolomics enabling discovery of tens of thousands of small molecule biomarkers that link genetics and disease

Authors:

T. Long; Sapient, San Diego, CA

Abstract Body:

Metabolomics, or the measure of circulating small molecule factors, has the potential to transform our understanding of human disease as well as to uncover new biomarkers of drug response, target engagement, and disease risk even years in advance of disease onset. Small molecule biomarkers read out both genetic variations and exposures stemming from diet, lifestyle, and the environment, including the internal organs and the microbiome, and how these exposures interact with genetics to influence disease. Mapping small molecule biomarkers provides complementary information relative to genomics and can also be an intermediary to connect genetics to disease. Herein we describe a next generation, high-throughput rapid liquid chromatography-mass spectrometry (rLC-MS) system capable of capturing and measuring more than 11,000 circulating small molecule biomarkers - including thousands of uncharacterized factors - in a human biosample. This nontargeted platform was applied in population-level studies to mine genetic information on small molecule biomarkers via genome-wide association studies (GWAS) and identify biomarkers that are sensitive to changes in gene functions and that may be causal to disease development.

Biosamples from over 18,000 individuals across diverse populations were assayed using rLC-MS and matched with genotype data from the same cohort. Biocomputational integration of these two data-rich sources enabled the discovery of specific biomarkers that read out changes in gene function of disease-associated variants. Across the genome, we find tens of thousands of small molecule biomarkers that reach genome-wide thresholds of association with genetic variants. We also find that over 90% of the approximately 20,000 genes have genetic variants in the transcribed regions associated with one or more small molecule biomarker(s).

These findings emphasize the importance of nontargeted small molecule biomarker discovery as a complementary approach to genomics. Using rLC-MS technologies to probe the still largely unexplored non-genetic landscape of disease, we can leverage the discoveries in integrative analyses that elucidate links between genetics, environment/exposures, and disease. Performing these analyses at population scale provides the statistical power to identify robust, specific biomarkers across diverse individuals.

Omics Technologies Posters - Wednesday

PB3015. High-throughput RNA isoform sequencing using programmable cDNA concatenation.

Authors:

A. AlKhafaji¹, J. T. Smith¹, K. V. Garimella¹, M. Babadi¹, V. Popic¹, M. Sade-Feldman¹, M. Gatzert¹, S. Sarkizova¹, M. A. Schwartz¹, E. Blaum¹, A. Day¹, M. Costello¹, T. Bowers¹, S. Gabriel¹, E. Banks¹, A. Philippakis¹, G. M. Boland², P. C. Blainey¹, N. Hacohen¹; ¹Broad Inst., Cambridge, MA, ²Div. of Surgical Oncology, Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA

Abstract Body:

Alternative splicing is a core biological process that enables profound and essential diversification of gene function. Short-read RNA sequencing approaches fail to resolve RNA isoforms and therefore primarily enable gene expression measurements - an isoform unaware representation of the transcriptome. Conversely, full-length RNA sequencing using long-read technologies are able to capture complete transcript isoforms, but their utility is deeply constrained due to throughput limitations. Here, we introduce MAS-ISO-seq, a technique for programmably concatenating cDNAs into single molecules optimal for long-read sequencing, boosting the throughput >15 fold to nearly 40 million cDNA reads per run on the Sequel IIe sequencer. We validated unambiguous isoform assignment with MAS-ISO-seq using a synthetic RNA isoform library and applied this approach to single-cell RNA sequencing of tumor-infiltrating T cells. Results demonstrated a >40 fold boosted discovery of differentially spliced genes and robust cell clustering, as well as canonical PTPRC splicing patterns across T cell subpopulations and the concerted expression of the associated hnRNPLL splicing factor. High-throughput full-length cDNA sequencing will drive discovery of novel isoforms, bringing to light this rich genetic feature and make transcript isoform expression analysis the new standard in transcriptomics.

Omics Technologies Posters - Thursday

PB3016. High-throughput RNA sequencing directly from cell lysates enables reproducible phenotypic profiling for CRISPR treatment phenotyping and cell response screening applications

Authors:

A. O'Hara, M. Stephens, E. Louie, L. Turner, I. DeVito, C. Mozdierz, H. Latif, G. Zhou; Azenta, South Plainfield, NJ

Abstract Body:

Gene editing has the potential to revolutionize genetic approaches to personalized medicine and enable a new era of gene-based therapeutic treatment options for heretofore untreatable diseases. Recent advancements in targeted gene editing and genome editing techniques allow for highly-specific, targeted DNA modifications to genes of interest. This includes the CRISPR/Cas9 system, which can be used for gene knockout, gene mutagenesis and gain-of-function gene knock-in approaches to engineer cell lines and potential therapeutics. Downstream discovery of CRISPR editing effects relies heavily on high-throughput screening assays for measurement of phenotypic responses which can range from very targeted assays to wide-ranging unbiased approaches. Standard whole transcriptome RNA-seq is a preferred method for unbiased measurement of phenotypic responses but is often overlooked due to limited scalability and high sample screening costs. To address these challenges, we have developed a high-throughput gene expression (HT-GEx) assay that combines a variety of strategies to enable a low-cost and high-throughput alternative to standard RNA-seq. First, a simplified workflow removes upstream RNA isolation steps and tags transcripts directly from cell lysate. This is achieved by incorporating both a sample barcode and a unique molecular index (UMI) during the reverse transcription reaction. Next, we leverage a 3' end counting approach to enable a reduction in sequencing depth coverage (i.e. sample cost) without compromising gene detection sensitivity. With these key advances, a high-plex and high-throughput screening assay is achievable at a reduced cost by removing the need to purify RNA, tagging transcripts early in the workflow to allow pooling of samples and reducing sequencing depth. We have combined these high-throughput and cost-saving strategies and present results which confirm HT-GEx offers results on par with standard RNA-seq. Gene detection sensitivity based on number of genes detected per millions of reads is highly similar, with gene detection sensitivity saturated at approximately 2 million reads. Additionally, the number of genes detected between replicates for RNA samples and lysate samples demonstrate strong linear correlation, implying highly reproducible results. Thus, high-throughput gene expression is an ideal method for phenotypic screening and has applications in CRISPR edited cell lines following stand-alone or combined compound treatments.

Omics Technologies Posters - Wednesday

PB3017. High-Throughput Sample Processing workflow for rapid Vaginal microbiota profiling of women's health samples.

Authors:

J. Kelvekar, E. Zeringer, A. Cheng; ThermoFisher Scientific, Austin, TX

Abstract Body:

Vaginal microbiota is composed of a dynamically convoluted microecosystem that fluctuates throughout woman's life. For a healthy woman, the vaginal microbiota is highly influenced by Lactobacillus species which produce various antimicrobial compounds and helps to keep pathogenic bacteria in check. Conversely, Bacterial Vaginosis (BV), affecting up to 30% of women around the world and leads to increased risk for sexually transmitted diseases, HIV, and premature birth, is characterized by a sharp reduction in Lactobacillus species leading to an overgrowth of anaerobic bacteria with the potential for pathogenesis. However, despite the high importance of the vaginal microecosystem, there is a lack of knowledge regarding its function and how it helps in protecting the female reproductive system. This leads to a lack of development of an effective clinical treatment to improve women's health. The traditional techniques for vaginal microbiota identification, such as complex culture preparation, specialized microscopy and sequencing often take days or weeks and can have sensitivity limitations. To enable advancement of vaginal microbiota research, Thermo Fisher Scientific launched a robust detection workflow using the MagMax™ DNA Multi-sample Ultra Kit on the KingFisher Flex instrument together with the TaqMan OpenArray Vaginal Microbiota Comprehensive Plate. This is a simple end-to-end workflow that includes, nucleic acid extraction, purification and interrogation of 34 microbial targets with prokaryotic 16S rRNA and human RNase P gene targets as process controls. The vaginal microbiota profiling workflow is compatible with several sample collection and storage systems, including the HOLOGIC™ ThinPrep™ Pap Test, the BD SurePath™ Liquid-Based Pap Test, the Copan Diagnostics ESwab™ Specimen Collection and Transport System, and Thermo Scientific™ MicroTest™ M4RT kits and allows processing of up to 96 samples in less than 8 hours and enables simple and efficient sample processing with a rapid turnaround.

Omics Technologies Posters - Wednesday

PB3018*. High-throughput targeted enrichment for long read third-generation sequencing: A flexible, scalable, and cost-efficient method achieving similar throughput as for short read NGS.

Authors:

T. Steiert¹, J. Fuss¹, S. Juzenas¹, M. Wittig¹, M. Hoepfner¹, M. Vollstedt¹, G. Varkalaite², H. ElAbd¹, C. Brockmann³, S. Goerg³, C. Gassner⁴, M. Forster¹, A. Franke¹; ¹Inst. of Clinical Molecular Biology, Christian-Albrechts Univ. and Univ. Med. Ctr. Schleswig-Holstein, Kiel, Germany, ²Inst. for Digestive Res., Lithuanian Univ. of Hlth.Sci., Kaunas, Lithuania, ³Inst. of Transfusion Med., Univ. Med. Ctr. of Schleswig-Holstein, Kiel, Germany, ⁴Inst. of Translational Med., Private Univ. in the Principality of Liechtenstein, Triesen, Liechtenstein

Abstract Body:

Hybridization-based targeted enrichment is a widely used and well-established technique in high-throughput second-generation short read sequencing. Despite the high potential to genetically resolve highly repetitive and variable genomic sequences by, for example PacBio third-generation sequencing, targeted enrichment for long fragments has not yet established the same high-throughput due to currently existing complex workflows and technological dependencies. We here describe a flexible, scalable, and cost-efficient targeted enrichment protocol for fragment sizes of >7 kb. For demonstration purposes we developed a custom blood group panel of challenging loci. Test results achieved >65% on-target rate, good coverage (142.7x) and sufficient coverage evenness for both non-paralogous and paralogous targets, and sufficient non-duplicate read counts (83.5%) per sample for a highly multiplexed enrichment pool of 16 samples. We genotyped the blood groups of nine patients employing highly accurate phased assemblies at an allelic resolution that match reference blood group allele calls determined by SNP array and NGS genotyping. Seven Genome-in-a-Bottle reference samples achieved high recall (96%) and precision (99%) rates. Mendelian error rates were 0.04% and 0.13% for the included Ashkenazim and Han Chinese trios, respectively. In summary, we provide a protocol and first example for accurate targeted long read sequencing that can be used at similar high-throughput fashion as for short read NGS.

Omics Technologies Posters - Thursday

PB3019. Horizon - an ultrarapid platform for NICU/PICU whole genome clinical pathogenicity prediction

Authors:

G. Salih^{1,2}, M. Woodbury-Smith³, P. Rahman⁴, W. Kuebler⁵, A. Ahmad¹, H. Akter⁶, B. Berdiev¹, A. Islam⁶, N. Nassir¹, M. Uddin^{1,7}; ¹Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates, ²Vrije Univ.it Amsterdam, Amsterdam, Netherlands, ³Newcastle Univ., Jesmond, Newcastle upon Tyne, United Kingdom, ⁴Mem. Univ., St. John's, NL, Canada, ⁵Charité - Univ.smedizin Berlin, Berlin, Germany, ⁶NeuroGen Hlth.care Ltd., Dhaka, Bangladesh, ⁷GenomeArc Inc., Toronto, ON, Canada

Abstract Body:

The fast adaptation of whole genome sequencing technologies in the clinic provides major advantages in identifying clinically actionable variants. However, the functional interpretation of the identified variants poses a challenge in terms of accuracy and efficiency, especially in neonatal and pediatric intensive care units (NICU/PICU) where molecular diagnosis and clinical decision often need to be made very quickly. We developed an ultrarapid analytics, Horizon, an automated machine learning integrated platform, to predict pathogenicity of genetic variants (single nucleotide variants (SNV), indels, and structural variants (SVs)) from whole exome/genome mapped .vcf or .bed files. The platform implements i) a functional prediction module for mutations, ii) multiple machine learning modules, iii) large parallel processing and indexing strategies, and, iv) integration with multiple largescale proprietary highly accurate human annotated clinical genomic and pharmacogenomic databases. We have conducted run-time validation experiments on whole exome (64K variants) and whole genome (4 million variants) .vcf files for functional annotation and pathogenicity predictions according to American College of Medical Genetics (ACMG) guidelines. Using a small quad core server station, the computation time was 5 and 65 minutes to process entire exome and whole genome .vcf data, respectively. A patient with genome wide structural variants can be processed within 60 seconds. We are currently using population scale clinical genomic datasets to validate sensitivity and specificity of Horizon's pathogenicity prediction module. Using a cohort of 4,864 pediatric genetic disease patients (including NICU/PICU), the validation analysis will compare each patient's Horizon predicted pathogenicity with pre-assigned pathogenicity by multiple ACMG board certified clinical geneticists. Our initial results indicate that Horizon's ultrarapid implementation of functional annotation and pathogenicity prediction can be extremely useful for timely diagnosis and to make clinical intervention related decision of NICU/PICU patients.

Omics Technologies Posters - Wednesday

PB3020*. Human cytomegalovirus infection extensively re-organizes the human genome at human disease risk loci

Authors:

M. Weirauch¹, S. Khund¹, S. Parameswaran¹, M. J. Beucler², M. Freeman², O. Donmez¹, A. von Handorf¹, L. E. Edsall³, K. Kaufman⁴, B. Zhao⁵, J. B. Harley⁶, W. Miller², L. C. Kottyan⁷; ¹Cincinnati Children's Hosp., Cincinnati, OH, ²Univ. of Cincinnati Coll. of Med., Cincinnati, OH, ³Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, ⁴Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, ⁵Harvard Univ., Cambridge, MA, ⁶US Dept. of VA, Cincinnati, OH, ⁷Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH

Abstract Body:

Human infection with Cytomegalovirus (CMV) has been linked to several human disorders. However, the underlying mechanisms remain poorly elucidated. Here, using RNA-seq, ATAC-seq, ChIP-seq and Hi-ChIP-seq, we show that CMV extensively re-organizes the human genome upon infection of both fibroblast and epithelial cell lines. Comparison between uninfected and infected cells using RNA-seq and ATAC-seq data indicates thousands of differentially expressed genes and tens of thousands of human genomic regions with differential chromatin accessibility. Transcription factor binding site motif enrichment analysis reveals enrichment for CTCF and TEAD1 sites in the newly opened and closed chromatin of infected cells, respectively. ChIP-seq experiments confirm strong differential binding in infected vs uninfected cells of CTCF and TEAD1 at these sites. We confirm that the extensive loss of TEAD1 binding leads to inactivation of the Hippo pathway in infected cells. Regions of differential chromatin accessibility are highly enriched for genetic variants associated with diverse human diseases and phenotypes including several autoimmune diseases, end stage coagulation, and dysregulation of lipid metabolism. Our approach and data provide new avenues to dissect the mechanisms operating in host-pathogen interactions and characterize their effects on disease phenotypes.

Omics Technologies Posters - Thursday

PB3021*. Human microglia transcriptional changes associated with progression and development of Alzheimer's Disease.

Authors:

R. Kosoy¹, J. Bendl², S. Kleopoulos², Y. Ma¹, V. Haroutunian^{2,3}, D. Bennett⁴, J. Fullard², G. Hoffman¹, P. Roussos^{1,3}; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²Icahn Sch. of Med. at Mount Sinai, New York City, NY, ³James J. Peters VA Med. Ctr., Bronx, NY, ⁴Rush Univ. Med. Ctr., Chicago, IL

Abstract Body:

Microglia are resident macrophages of the brain with a range of functions, including immune surveillance, neuronal maintenance and synaptic remodeling. While microglia have been implicated in the etiology of Alzheimer's Disease, the relevance of animal model-based investigations is not clear and the contribution of microglia to disease progression in human patient samples is poorly understood. Importantly, we have very limited insights into the cellular and molecular processes of microglia in early versus advanced states of the disease, which is critical for developing appropriate therapeutic interventions. Here we present a transcriptional landscape associated with AD progression measures in primary microglia derived from 189 human postmortem brains, representing people with both healthy aging (61 individuals) and those with a range of neurodegenerative phenotypes (including 79 patients with AD). In addition to the short read RNAseq data generated in all of these patients, we also generated long-read IsoSeq data in 20 samples to identify the microglia specific transcripts, allowing us to query the expression patterns for transcribed entities beyond those present in the standard gene reference resources. We derived transcriptional changes associated with various AD relevant measures reflecting neurodegeneration, tau neurofibrillary tangles and dementia. Additionally, we utilized co-expression network approaches to describe biological structures intrinsic to human microglia, highlighting the diversity of biological processes associated with disease etiology and different AD phenotypes. We identify a molecular signature associated with disease severity allowing for identification of patient subtypes with different relationships between the objectively measured underlying neurodegeneration and the clinical manifestation of AD phenotypes. In particular, we observe pronounced indices of transcriptional dysregulation of immune functionality in microglia associated with AD progression. Furthermore, we observe independent patterns of biological functional annotations associated with AD etiology and progression, suggesting that the molecular and cellular processes leading to disease development are principally different from the processes relevant to disease progression and homeostatic maintenance. Finally, we identify a number of microglial key driver genes with most relevance to AD associated signatures and biological functions for further exploration of their suitability as therapeutic targets.

Omics Technologies Posters - Wednesday

PB3022. Identification of FZD3 and SEMA3A as potentially secreted proteins involved in proliferative sickle cell retinopathy.

Authors:

A. Lima Camargo, V. de Haidar e Bertozzo, S. da Silva Costa, J. Nicolielo Pereira de Castro, T. Adalton Rosa Rodrigues, P. Rodrigues Souza Cruz, M. Gonçalves Monteiro Viturino, M. C. Ozelo, F. Ferreira Costa, M. Barbosa Melo; UNICAMP, Campinas, Brazil

Abstract Body:

Background: Retinal alterations are the most important ocular morbidity in sickle cell diseases. They occur due to vaso-occlusion of the ocular microvasculature, which leads to hypoxia, ischemia, and neovascularization. Proliferative sickle cell retinopathy (PSCR) is the most severe form of retinopathy, leading to retinal detachment and visual loss, affecting 10 to 20% of the affected eyes. Endothelial Colony Forming Cells (ECFCs) are recruited into ischemic or damaged outer retina and play an important role during vascular repair and revascularization of ischemic retinopathies. **Aim:** To evaluate the differential gene expression profile and predict potentially secreted proteins (PSPs) in ECFCs from patients with HbSS genotypes with and without PSCR. **Methods:** RNA-seq data analysis comparing ECFCs from patients with PSCR (N=3) versus without PSCR (N=2) was performed with the Deseq2 package in R. *PANTHER* database was used for gene ontology (GO) analysis. Differentially expressed genes (DEGs) were combined for retina tissue-specific genes dataset by *Genotype-Tissue Expression (GTEx)*. DEGs obtained from GO terms and GTEx were compared to *The Human Protein Atlas* secretome database to identify PSPs. *miRWalk* 3.0 tool was used to evaluate predicted regulatory miRNAs. **Results:** The ECFCs expression profile revealed 501 DEGs. GO analysis demonstrated 129 DEGs related with angiogenesis, cell adhesion, MAP kinase activity, and PI3K signaling, pathways involved in PSCR. Among them, we highlight the overexpression of Frizzled Class Receptor 3 (*FZD3*), and underexpression of Semaphorin 3A (*SEMA3A*) which are PSPs also expressed in the retina. In the PSPs-miRNA network, the following edges were observed: *FZD3* with miR-20b-5p; *SEMA3A* with miR-15b-5p, microRNAs also expressed in the retina. Studies suggest that *SEMA3A* and Vascular endothelial growth factor (VEGF) have antagonistic effects on EC growth and survival. *VEGF* plays an important role during angiogenesis, and was overexpressed in our DEGs, unlike *SEMA3A*. High expression of *FZD3* and *VEGF* in EC has been associated with deregulation of the Wnt pathway. Interestingly, Wnt pathway ligands and receptors, such as *FZD3*, are key regulators of ocular angiogenesis in vascular eye diseases. Moreover, miR-20b-5p and miR-15b-5p were reported to be underexpressed and overexpressed, respectively, in the pathophysiology of diabetic retinopathy. **Conclusions:** We identified *SEMA3A* and *FZD3* as PSPs in ECFCs that may have essential roles in angiogenic processes and are also regulated by miRNAs. This data may contribute to a better understanding of the mechanisms involved in PSCR and indicate novel biomarkers of disease.

Omics Technologies Posters - Thursday

PB3023. Identification of novel long non-coding RNA with distinct expression patterns in different multiple myeloma subtypes.

Authors:

M. Bauer, D. Elsayed, C. Wardell, F. van Rhee, D. Ussery, C. Ashby, F. Zhan; Univ. of Arkansas for Med. Sci., Little Rock, AR

Abstract Body:

The biological role of long non-coding RNAs (lncRNAs) in multiple myeloma (MM) has recently been receiving more attention as we learn more of their vital role in all cellular functions. Despite advancement in therapies, a significant variation in survival with symptomatic MM still exists, indicating an enormous genomic heterogeneity and complexity of the disease. While lncRNAs do not carry information about proteins recent studies suggest that they regulate several molecular mechanisms, such as, supporting cellular homeostasis and regulation of gene expression. LncRNA expression is highly cellular specific and the list of discovered lncRNAs continues to grow.

In this study we explore the expression of novel lncRNA in different molecular subtypes of MM and examine their contribution to patient prognosis. The RNA-seq data of 643 newly diagnosed MM samples was analyzed to discover novel lncRNAs. We identified 8,556 potentially novel lncRNA transcripts. We used a number of bioinformatic approaches to infer their function and role in MM. Which included, looking for correlated expression with nearby protein coding genes and their pathways. We also compared k-mer content with known lncRNA to infer possible similar function. We then assessed significance through survival analysis.

We identified 1,264 novel lncRNA transcripts with significant differential expression between the different molecular subtypes of multiple myeloma. Hierarchical clustering defined 9 clusters of novel lncRNAs associated with different subtypes. Using nearby protein coding genes with significant expression correlation, we observed overlapping enriched pathways between clusters as well as unique pathways distinct to specific MM groups. The pathway, Signaling by BMP was the topmost significant enriched pathways for cluster 1 and 9 in which lncRNAs were highly expressed in MM defined by t(14;16) and t(14;20). K-mer content comparison with known lncRNA identified novel lncRNAs with similarity to HELLP associated long non-coding RNA (*HELLPAR*), *AC003984.1* and PDGFA Divergent Transcript (*PDGFA-DT / HRAT92*). These 3 known lncRNAs have been previously implicated in other diseases and may provide clues to the function of the novel lncRNAs. Lastly we did a survival analysis and identified 108 novel lncRNAs associated with significant overall survival when expressed.

The results of our study reveal that the noncoding portion of the genome is still largely unexplored and the potential role of lncRNAs in cellular maintenance and survival in MM. Disruption of these key molecules may have a big impact on disease progression and outcome and provide novel potential therapeutic targets.

Omics Technologies Posters - Wednesday

PB3024. Identifying high-confidence de novo mutations in somatic and germline cells through duplex sequencing of diverse tissue types.

Authors:

S. Lulla¹, J. Kunisaki², L. Hiatt¹, A. Quinlan³; ¹Univ. of Utah, Salt Lake City, UT, ²Univ. of Utah Sch. of Med., Salt Lake City, UT, ³Univ of Utah, Salt Lake City, UT

Abstract Body:

Identification of de novo mutations (DNMs) and the rate at which they accumulate have applications in research and clinical work, especially related to cancer and genetic disease risk. Several research efforts have used whole-genome sequencing (WGS) technologies to analyze DNMs. However, WGS has an error rate on the order of 10^{-3} , which presents challenges when distinguishing low-frequency mutations (signal) from sequencing artifacts (noise) in bulk tissue samples. TwinStrand Duplex Sequencing overcomes the limitations associated with WGS by reducing error rates to the order of 10^{-8} through exclusive reporting of variants that occur on complementary DNA strands, allowing more accurate quantification of DNM rates. Here, we present a computational workflow to quality control and examine DNMs detected with TwinStrand Duplex Sequencing across bulk somatic and germline tissue samples.

To determine general reliability of duplex sequencing data, we compare our observed mutation spectra to those expected from previous studies. We then further refine the data to high-confidence variant calls through several filtering strategies. Mutations that occur at frequencies of roughly 0.5 and 1 are considered heterozygous and homozygous inherited mutations respectively and are removed from analysis to isolate DNMs. Non-clonal variants that occur in more than one matched tissue sample or across multiple donors are likely to be artifacts and suggest error-prone loci. Variants with high no-call rates and homopolymer INDELS that occur adjacent homopolymers in the reference genome are also considered artifacts. Reasons for filtering are annotated for each variant, and variants are additionally annotated with information about mutation consequence and pathogenicity using ClinVar data. The workflow has been documented in a GitHub repository and used in a pilot study of longitudinal bulk sperm sample DNM rates.

We anticipate applying the pipeline to bulk sperm to measure germline mutation rates in the context of infertility. We also aim to apply the pipeline to somatic cells including blood, prostate, skin, and testicular tissue, as well as cancerous, pre-cancerous, and noncancerous colon. We will use these tissue data to evaluate mutagenic etiologies.

Omics Technologies Posters - Thursday

PB3025*. Illuminating Dark Proteins using Reactome Pathways

Authors:

L. Matthews¹, G. Wu², R. Haw³, T. Brunson², N. Sanati², S. Shorser³, D. Beavers², P. Conley², L. Stein³, P. D'Eustachio¹; ¹NYU Langone Med. Ctr., New York, NY, ²Oregon Hlth.& Sci. Univ., Portland, OR, ³OICR, Toronto, ON, Canada

Abstract Body:

Diseases are often the consequence of proteins or protein complexes that are non-functional or that function improperly. An active area of research has focused on the identification of molecules that can interact with defective proteins and restore their function. While 22% percent of human proteins are estimated to be druggable, less than fifteen percent are targeted by FDA-approved drugs, and the vast majority of untargeted proteins are understudied or so-called "dark" proteins. Elucidation of the function of these dark proteins, particularly those in commonly drug-targeted protein families, may offer therapeutic opportunities for many diseases. Reactome is the most comprehensive, open-access pathway knowledgebase covering 2585 pathways and including 14246 reactions, 11088 proteins, 13984 complexes, and 1093 drugs. Placing dark proteins in the context of Reactome pathways provides a framework of reference for these proteins facilitating the generation of hypotheses for experimental biologists to develop targeted experiments, unravel the potential functions of these proteins, and then design drugs to manipulate them. To this end, we have trained a random forest with 106 protein/gene pairwise features collected from multiple resources to predict functional interactions between dark proteins and proteins annotated in Reactome and then developed three scores to measure the interactions between dark proteins and Reactome pathways based on enrichment analysis and fuzzy logic simulations. Literature evidence via manual checking and systematic NLP-based analysis support predicted interacting pathways for dark proteins. To visualize dark proteins in the context of Reactome pathways, we have also developed a new website, idg.reactome.org, by extending the Reactome web application with new features illustrating these proteins together with tissue-specific protein and gene expression levels and drug interactions.

Omics Technologies Posters - Wednesday

PB3026. Improved data quality for renal cell carcinoma samples with HIVE scRNAseq integrated storage.

Authors:

E. Sergison, L. Wasson, T. Gierahn, J. Flanigan; Honeycomb Biotechnologies, Waltham, MA

Abstract Body:

Honeycomb Biotechnologies offers a new solution for single-cell technology: HIVE scRNAseq. The HIVE is a portable, handheld, single-use device that enables gentle capture, RNA stabilization without fixation, and easy processing for the analysis of single-cell samples without specialized instrumentation. The HIVE uses an array of 65,000 picowells that are preloaded with uniquely barcoded 3' transcript capture beads to quickly capture single cells. Once loaded, the Cell Preservation Solution is added to stabilize RNA and lock in molecular signals. Cell-loaded HIVEs can be shipped or stored until ready for Honeycomb's simplified and scalable HIVE processing and library preparation workflow. Quality control metrics (genes, transcripts, % mitochondrial reads), marker gene expression, and cell-type proportions are not significantly different between freshly processed cells and cells stored in the HIVE device for up to 9 months at -20°C.

Here, we demonstrate robust capture and profiling of cells isolated from a renal cell carcinoma biopsy. Freshly dissociated cells were either loaded directly into the HIVE (pre-cryo) and stored for ~3 weeks or cryopreserved for ~3 weeks prior to loading into the HIVE (post-cryo). Beads with captured transcripts were extracted from the HIVEs, and the remaining HIVE library preparation steps were conducted in a 96-well plate format. HIVE scRNAseq libraries were sequenced on an Illumina NovaSeq 6000 and count matrix files were generated using BeeNet, Honeycomb's software specifically designed for HIVE scRNAseq libraries. Seurat was used for downstream analysis. We provide data demonstrating the increased viability and quality of single-cell RNAseq data generated using the HIVE scRNAseq Solution with integrated storage compared to data generated from samples stored using cryopreservation. The HIVE propels forward clinical research using blood, bone marrow, fine needle aspirates, and other minimally invasive biopsies in the context of infectious diseases, blood cancers, allergy/asthma, and autoimmune diseases and inflammation. Stable HIVE storage enables longitudinal studies, time-courses, sporadic or end-of-day samples, collection sites with centralized processing, and multi-site collaborations. Honeycomb aims to expand single-cell opportunities across basic, translational, and clinical research throughout the world.

Omics Technologies Posters - Thursday

PB3027. Improved detection of functionally relevant aberrant splicing using the intronic Jaccard index

Authors:

C. Mertes, I. Scheller, K. Lutz, V. Yepez, J. Gagneur; Technical Univ. of Munich, Garching, Germany

Abstract Body:

Detection of aberrantly spliced genes from RNA-seq data is an important step of RNA-seq based diagnostics of rare diseases. We have recently developed FRASER¹, a denoising autoencoder-based method that outperformed alternative approaches for the detection of aberrant splicing on precision-recall analyses. When systematically investigating FRASER results on 303 rare disease samples² and 16,213 GTEx v8 samples, we noticed however that many of the detected events captured aberrant weak and cryptic splice site usage that did not result in substantial major isoform change.

Here, we introduce the intronic Jaccard Index, a new intron excision metric that combines alternative donor, alternative acceptor, and intron retention signal into a single value. The intronic Jaccard index captures variations in major isoform usage more closely than other split-read based metrics while retaining the advantage of being annotation-agnostic. As with FRASER, we modeled this ratio using a beta-binomial based denoising autoencoder thereby controlling for potential confounding sources of variations. Compared to FRASER, using a single metric also reduces the multiple testing burden and speeds up the fitting procedure. In addition, we implemented a goodness-of-fit metric, which we use to filter out false positives caused by model misspecification. Collectively, we call this new approach FRASER2.

On the rare disease dataset, FRASER2 considerably reduced the number of splicing outliers (by 25% - 75% per sample) for a mild loss of sensitivity (only 2 out of 24 pathogenic splicing cases not recovered with default cutoffs). Moreover, in GTEx, FRASER2 splicing outliers were 5 times more enriched for rare splice affecting variants. Including the goodness of fit metric further reduced the number of false positive calls. Also, outliers called using the intronic Jaccard index are less sensitive to sequencing depth. By introducing a new splicing metric, the intronic Jaccard index, FRASER2 is able to maintain the sensitivity of FRASER while providing more functionally relevant outlier calls. References:¹Mertes, Scheller et al., Nat Commun (2021)²Yépez, Gusic, et al., Genome Med (2022)

Omics Technologies Posters - Thursday

PB3028. Improved sequence mapping using a complete reference genome and lift-over

Authors:

N-C. Chen¹, L. Paulin², F. Sedlazeck^{3,4}, S. Koren⁵, A. Phillippy⁶, B. Langmead¹; ¹Johns Hopkins Univ., Baltimore, MD, ²Baylor Coll. of Med., Huston, TX, ³Baylor Coll. Med., Houston, TX, ⁴Rice Univ., Houston, TX, ⁵Natl. Human Genome Res. Inst., NIH, Bethesda, MD, ⁶NIH/NHGRI, Bethesda, MD

Abstract Body:

Complete, telomere-to-telomere genome assemblies promise improved analyses and the discovery of new variants, but many essential genomic resources remain associated with older reference genomes. Thus, there is a need to translate genomic features and read alignments between references. Here we describe a new method called levioSAM2 that accounts for reference changes and performs fast and accurate lift-over between assemblies using a whole-genome map. In addition to enabling the use of multiple references, we demonstrate that aligning reads to a high-quality reference (e.g. T2T-CHM13) and lifting to an older reference (e.g. GRCh38) actually improves the accuracy of the resulting variant calls on the old reference. By leveraging the quality improvements of T2T-CHM13, levioSAM2 reduces small-variant calling errors by 11.4-39.5% compared to GRC-based mapping using real Illumina datasets. LevioSAM2 also improves long-read-based structural variant calling and reduces errors from 3.8-11.8% for a PacBio HiFi dataset. Performance is especially improved for a set of complex medically-relevant genes, where the GRC references are lower quality. The software is available at <https://github.com/milkschen/leviosam2> under the MIT license.

Omics Technologies Posters - Wednesday

PB3029*. Improving gene detection sensitivity from clinically relevant low input and FFPE RNA samples utilizing a rapid whole transcriptome library prep workflow.

Authors:

T. Sanders¹, L. French², J. Walker¹, C. Ross¹, T. Harrison¹, K. Reed¹, R. Wadsworth², B. Kudlow¹, J. Pavlica¹; ¹Watchmaker Genomics, Boulder, CO, ²Watchmaker Genomics, Cape Town, South Africa

Abstract Body:

mRNA-seq is a powerful tool for transcriptome profiling. However, it does not detect long noncoding RNAs, which are valuable disease biomarkers, and it is not amenable to certain clinically relevant sample types, such as blood and FFPE. Whole transcriptome sequencing (WTS), which specifically depletes uninformative overabundant transcripts, preserves noncoding species and is compatible across a wide range of sample types and qualities. As WTS becomes more broadly applied in the clinical and translational space, there is an increasing demand for solutions that alleviate the laborious library preparation bottleneck while also meeting increasingly stringent sensitivity requirements. To address this need, we undertook a complex development effort combining enzyme engineering and streamlined WTS library preparation workflow design and optimization.

Traditional RNA depletion workflows are lengthy, complicated, and difficult to automate - requiring long incubations and reagent additions at elevated temperatures. To streamline the protocols and improve depletion efficiency and specificity, we developed algorithms to computationally design novel depletion probes. To improve sensitivity, we leveraged an engineered reverse transcriptase to maximize conversion and tailored buffers throughout the workflow to further enhance performance. Cleanup steps were reduced to minimize sample loss, and a novel de-crosslinking step was integrated to improve FFPE performance.

With RNA extracted from whole blood and multiple FFPE blocks, we compared our solution to commercially available kits using inputs ranging from 1 to 500 ng. We observed a significant increase in sensitivity with low input and FFPE samples in comparison to other workflows, as indicated by an increase in the number of unique genes detected and more reads supporting a given gene call. Further, Watchmaker libraries showed excellent inter-input correlation, reproducibility, and improved sequencing economy with fewer bases wasted due to lack of alignment or mapping to rRNA regions.

This highly automatable and rapid solution generates libraries in under 5 hours and improves on the data quality currently attainable with existing commercial kits - making more clinically relevant samples addressable with higher success rates.

Omics Technologies Posters - Thursday

PB3030. Improving sensitivity of combinatorial barcoding-based single cell sequencing.

Authors:

S. Schroeder, G. S. Kim, V. Tran, P. Matulich, J. Pangallo, A. Sova, A. Gadkari, L. Kenyon, E. Papalexi, Z. Sayar, A. Sapre, R. Koehler, A. Rosenberg, C. Roco; Parse BioSci.s, Seattle, WA

Abstract Body:

While adoption of single cell sequencing has grown rapidly over recent years, microfluidic based approaches remain costly and the need to isolate individual cells in droplets limits throughput. These limitations can be overcome by using combinatorial barcoding to label individual cells in single cell experiments. This approach uses fixed cells or nuclei as starting material and generates single cell transcriptional profiles through a series of split-pool barcoding steps that uniquely label individual cells without any need for costly microfluidic instruments. Furthermore, the ability to use fixed cells and nuclei provides flexibility and consistency in single cell RNA-seq experiments, as samples can be frozen and stored for up to 6 months, then processed in a single reaction, eliminating potential batch effects. With both droplet and combinatorial based scRNA-seq, a key consideration is the number of RNA molecules that can be detected per cell. Over the last several years there have been considerable improvements in the sensitivity of both droplet and combinatorial based single cell sequencing. This added sensitivity makes it possible to detect lowly expressed genes that may be critical in pathway analysis or distinguishing cell types and cell states. However, many samples and cell types have inherently low RNA content due to cell size and transcriptional activity and would benefit from further improvements in sensitivity. Here we demonstrate that by using an updated and optimized chemistry, we are able to dramatically improve the number of genes detected per cell in single cell experiments using Parse's Evercode™ single cell products. We observe consistent improvement in transcript and gene detection across cells and nuclei from a variety of different sample types tested, including mouse brain nuclei and human PBMCs. In PBMC samples we see an approximate two-fold improvement in transcripts detected per cell. We anticipate these improvements will enable researchers to uncover underlying biology that was previously obscured with lower sensitivity single cell RNA-seq.

Omics Technologies Posters - Wednesday

PB3031. Imputing functionally impactful *CYP2A6* structural variants from SNP array data.

Authors:

A. W. R. Langlois^{1,2}, M. Chenoweth^{1,2}, C. Lerman³, N. L. Nollen⁴, L. S. Cox⁴, J. S. Ahluwalia⁵, R. F. Tyndale^{1,2,6}; ¹Dept. of Pharmacology & Toxicology, Univ. of Toronto, Toronto, ON, Canada, ²Campbell Family Mental Hlth.Res. Inst., CAMH, Toronto, ON, Canada, ³Norris Comprehensive Cancer Ctr., Univ. of Southern California, Los Angeles, CA, ⁴Dept. of Population Hlth., Univ. of Kansas Sch. of Med., Kansas City, KS, ⁵Dept.s of Behavioral and Social Sci. and Med., Brown Univ., Providence, RI, ⁶Dept. of Psychiatry, Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

CYP2A6 is the primary nicotine-metabolizing enzyme, responsible for ~90% of nicotine inactivation. Faster CYP2A6 activity is associated with greater smoking, poorer smoking cessation outcomes and higher lung cancer risk. The *CYP2A6* gene is polymorphic, with known functional variants including structural variants (SV) such as a gene deletion (*CYP2A6*4*), a gene duplication (*CYP2A6*1x2*), and hybrids with the highly homologous *CYP2A7* pseudogene (*CYP2A6*12*, *CYP2A6*34*). Compared to single nucleotide polymorphisms (SNP), which comprise the other functional *CYP2A6* variants, SVs are also functionally impactful but challenging to genotype due to their complex genetic architecture. Work in the major histocompatibility complex region found that SVs are associated with SNP haplotypes using SNP array data (Sekar et al. 2016). Our aim was to develop a reference panel for imputation of *CYP2A6* SVs from SNP array data.

European- (EUR; n=935) and African-ancestry (AFR; n=964) individuals from smoking cessation clinical trials were genotyped for SNPs on an Illumina array and for *CYP2A6* SVs using Taqman copy number assays. SV diplotype and SNP array (with the region affected by SVs excluded) data were integrated and phased using Beagle 5.2 (as in Sekar et al. (2016)); ancestry-specific panels were generated. Leave-one-out cross-validation was used. Briefly, a target sample was set, and a panel was created using all other samples. SV diplotype was imputed for the target; this was repeated for all samples. Accuracy was measured by comparing imputed calls to known SV diplotypes.

In AFR, diplotypes (i.e. the presence or absence of SVs for both alleles) were accurately imputed in 97% of all individuals (932/964), while among those with SVs 67% were identified (11/11 *CYP2A6*12*, 27/44 **4*, 1/2 novel hybrid SV, 2/2 **34*, and 16/26 **1x2* alleles).

In EUR, diplotypes were accurately imputed in 98% of all individuals (912/935), while among those with SVs 76% were identified (41/42 *CYP2A6*12*, 2/6 **4*, 9/10 novel hybrid SV, and 4/16 **1x2* alleles). False positive (FP) rates were low in both ancestry groups (<1% overall).

Our data confirm that *CYP2A6* SVs are associated with identifiable SNP haplotypes. Over 70% of SV alleles were positively identified with very low FP rates; panel expansion could improve this rate. Use of this reference panel could simplify SV genotyping for prediction of CYP2A6 activity, particularly for clinical use of weighted genetic risk scores. The method could also be applied widely to other genes with SVs. While external validation of the panel is necessary, array imputation may be useful for predicting *CYP2A6* SVs in large biobanks and clinical trials.

Omics Technologies Posters - Thursday

PB3032. Imputing the whole metagenomic shotgun sequencing data using the 16s amplicon sequencing data

Authors:

S. Jang, B. Han; Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Abstract Body:

The importance of the human microbiome is emerging to analyze various diseases current days. There are two methods for profiling the microbiome: 16s rRNA gene sequencing (16s sequencing) and whole-genome shotgun sequencing (WGS). Although the cost of NGS has become much cheaper than before, 16s sequencing is still widely used to identify the profile of microorganisms since it is much cheaper than WGS. Unlike 16s rRNA sequencing, WGS can read all metagenomic DNA, not specific DNA. Therefore, WGS has multiple advantages compared to the 16s sequencing including higher taxonomic profiling resolution and functional profiling analysis. Our study provides a method that can use machine learning to impute the WGS profiled count data from 16s profiled count data. It is quite challenging that profiled data has lots of zero inflation. Our approach can help studies characterize and quantify accurate genus and species levels of microbiome profile cost-effectively without WGS.

Omics Technologies Posters - Wednesday

PB3033. Increased mitochondrial expression is associated with increased *APOEε4* expression and affected by ancestry and sex.

Authors:

M. Muniz Moreno¹, K. Celis¹, F. RAJABLI², P. Whitehead³, K. Hamilton-Nelson⁴, G. Beecham², D. Dykxhoorn², K. Nuytemans⁵, L. Wang², O. Gardner⁴, D. Dorfsman², E. Bigio⁶, M. Mesulam⁶, D. Bennett⁷, T. Schuck⁸, S. Weintraub^{9,10}, C. Geula⁶, M. Gearing¹¹, C. Alba¹², C. Dalgard¹³, D. Davis², W. Scott², J. L. Haines¹⁴, M. Pericak-Vance⁴, A. Griswold², J. Young¹⁵, J. Vance²; ¹HHG, Miami, FL, ²Univ. of Miami, Miami, FL, ³Univ. of Miami, MIAMI, FL, ⁴Univ. of Miami Miller Sch. of Med., Miami, FL, ⁵John P. Hussman Inst. for Human Genomics, Miami, FL, ⁶Feinberg Sch. of Med., Chicago, IL, ⁷Rush Alzheimer's Disease Ctr. - Rush Univ., Chicago, IL, ⁸University of Pennsylvania, Pennsylvania, PA, ^{9,10}Weinberg Coll. of Arts and Sci., Chicago, IL, ¹¹Emory Univ. Sch. of Med., Atlanta, GA, ¹²The Henry M Jackson Fndn., Bethesda, MD, ¹³USUHS - TAGC, Bethesda, MD, ¹⁴Case Western Reserve Univ, Cleveland, OH, ¹⁵Univ. of miami, Miami, FL

Abstract Body:

Alzheimer disease (AD) is a multifactorial neurodegenerative disease that affects 6.5 million Americans. The *APOEε4* allele is the strongest genetic risk factor and is reported to lead to mitochondrial dysfunction (MTD). The AD risk is lower when the allele is carried on an African local ancestry (ALA) compared to European local ancestry (ELA). Since two thirds of AD patients are female, we sought to get a deeper understanding of ancestry-specific gender alterations in carriers of the strongest genetic risk factor, *APOEε4*. Single nuclei RNA sequencing (snRNA-seq) the frontal cortex (Brodmann area 9) in 17 homozygous AD *APOEε4/4* carriers of ALA (N=7, 4 females) and ELA (N=10, 7 females) was used to identify ancestry gender-based alterations and the specific MTD landscape of *APOEε4/4* carriers. The snRNA-seq was performed using 10X Chromium system. We found that the astrocyte clusters with high expression of *APOEε4*, also had the highest levels of mitochondrial (MT) markers. Next, we analyzed the 35 MTD processes in the MitoXplorer database (MX) and identified higher MTD alteration in cell types with the highest *APOEε4* expression, namely microglia, endothelial cells, and L2/3 layer excitatory neurons. When analyzing gender differences, all *APOEε4*/MT cell types, except one microglial subtype, showed lower MTD ratios in female *APOEε4* ALA carriers compared to *APOEε4* ELA. This ratio was inverted in males. The increase in MTD in males points to an overload of MTD leading to MT stress and oxidation that is associated with worsening AD pathology. The MX processes showing the greatest differences in both sexes were apoptosis, amino acid metabolism, calcium signaling, and MT dynamics. The expression of MT encoded genes linked to oxidative phosphorylation was decreased in females with *APOEε4* ALA compared to *APOEε4* ELA and increased in male samples of ALA compared to ELA. In conclusion, we observed specific mitochondrial processes significantly altered in the cell types which showed the highest expression of *APOEε4*, but differing in direction and degree by both sex and ancestry. This suggests further studies of the *APOEε4* effect on mitochondrial dysfunction are warranted and highlights the importance of considering sex and ancestry in the pathological mechanisms of AD.

Omics Technologies Posters - Thursday

PB3034. Increasing throughput and decreasing cost for single-cell profiling using combinatorial indexing

Authors:

M. Nakamoto, P. Boyd, N. Pacheco, J. Thomas, D. Pokholok, L. Lin, F. Schlesigner, T. Guettouche, F. Steemers; ScaleBio, San Diego, CA

Abstract Body:

The ability to examine populations at the single-cell level has led to novel advances across many fields, but despite the increasing number of systems available the cell throughput of single-cell experiments has not scaled as rapidly due to high costs and complex workflows. Studies that require hundreds of thousands of cells or more, such as those involving screens, rare cell populations, or large patient cohorts, can be prohibitively expensive.

To address this, we increase the throughput of on-market systems by adding upstream combinatorial indexing technology using the ScaleBio ATACseq pre-indexing kit. In this workflow nuclei from up to 24 samples are distributed across 24 wells of a provided plate for barcoded tagmentation, introducing an additional cell-based index to each molecule. Nuclei are then pooled and can be superloaded onto existing single-cell capture systems. Despite loading of 100,000 nuclei per channel the additional tagmentation barcode creates a low effective doublet rate, enabling recovery of the majority of multiplet nuclei for downstream analysis.

We demonstrate this technology first using a human/mouse barnyard experiment in which human and mouse cell lines were mixed prior to loading onto the tagmentation plate. Nuclei were then pooled and 100,000 nuclei were loaded onto a single lane of a droplet system. Analysis of these data show effective recovery of nuclei across multiple sites and users (~55-60%) with low effective doublet rates (<5%). Barnyard analysis using ATACseq data shows clean distinction of human and mouse populations with no decrease in sensitivity or increase in background with increasing nuclei per droplet. We next tested this technology with PBMCs with a similar loading rate of 100,000 nuclei per channel. We observed high barcode and mapping rates, high recovery of both fresh and frozen PBMCs, and after QC filtering and clustering we show successful identification of the major PBMC subsets using ATACseq data. Taken together these data show that the ScaleBio ATACseq pre-indexing kit can massively increase throughput of on-market systems while decreasing cost, enabling single-cell analysis for even the most high-throughput needs.

Omics Technologies Posters - Wednesday

PB3035. In-depth analysis and comparison of different whole genome methylation sequencing library preparations with low DNA input

Authors:

Z. Sun¹, S. Behati¹, P. Wang², A. Bhagwate¹, V. Wang¹, S. McDonough¹, W. Taylor¹, J. Cunningham¹, J. Kisiel¹; ¹Mayo Clinic, Rochester, MN, ²Mayo Clinic, Scottsdale, AZ

Abstract Body:

Background. Whole genome methylation sequencing (WGMS) is an unbiased and comprehensive way to measure the modification of cytosines in the genome. The most common approach is through bisulfite treatment which converts an unmethylated cytosine to uracil (thymine). However, the harsh treatment leads to DNA loss and artifacts. A recent enzymatic methyl-seq (EM-seq) overcomes these limitations and is becoming a popular alternative. WGMS can also theoretically be used for single nucleotide or structural variant detection; however, limited data are available on variant calling accuracy from WGMS data. **Materials and methods.** NA12878 was sequenced by EM-seq, QIAseq Methyl Library Kit (QIA-seq), Swift Accel-NGS Methyl-Seq (Swift-seq) and WGS with 25ng input (additional replicates of 25ng and 10ng for EM-seq). Traditional quality metrics (read alignment, genomic coverage, CpG capture, conversion rate, and etc), DNA methylation measurement, SNP call accuracy using Genome in a Bottle (GIAB) as reference and copy number calls using several benchmarks were compared. Two data processing pipelines, Bismark and SAAP-BS, and two methylation data specific variant callers BisSNP and BSSNPER were used to compare the analytical variability. CNVpytor was used to call copy number variation (CNV). **Results.** EM-seq demonstrated the best performance in almost all measurements. It had the highest mapping rate, number of CpGs captured at 5X, and the lowest duplication rate. QIA-seq had the highest conversion rate (0.999) but the highest duplication rate and lowest CpGs at 5X, particularly from Bismark processed data. Swift-seq performance was between EM-seq and QIA-seq. For the commonly detected CpGs within a processing pipeline, the highest correlation was between EM-seq and Swift-seq and followed by EM-seq replicates. Across processing pipelines, the data from the same library always clustered together and had correlation coefficient greater than 0.97. In SNP detection, EM-seq could reach a recall of 0.83, precision of 1, and F1 score 0.9 in which only the recall was inferior to WGS (0.98). Both Swift and QIAseq had low recalls and F1 scores although precision was high for the genomic positions with coverage. The recalls for CNVs in all libraries including the WGS were below 20% while an EM-seq library performed as well as the WGS. **Conclusions.** EM-seq demonstrates the best performance in CpG capture, genomic coverage evenness and DNA methylation measurement. It provides almost similar information for SNP and CNV to WGS. Poor CNV performance for all libraries is mostly due to lack of a gold standard of a truth set and remains challenging in genomics research (not unique to WGMS).

Omics Technologies Posters - Thursday

PB3036. Individualized cellular ancestry: efficient reconstruction of cell lineage trees

Authors:

Y. Jang¹, L. Fasching², T. Bae¹, L. Tomasini², J. Schreiner², A. Szekely², T. Fernandez², J. Leckman², F. Vaccarino², A. Abyzov¹; ¹Mayo Clinic, Rochester, MN, ²Yale Univ., New Haven, CT

Abstract Body:

Mosaic mutations can be used to track cell ancestries and reconstruct high resolution lineage trees starting from the very first cell divisions of the zygote. However, this approach requires sampling and analyzing the genomes of multiple cells, which can be redundant in lineage representation, limiting the scalability of the approach. We describe a strategy for cost- and time-efficient lineage reconstruction using clonal induced pluripotent stem cell (iPSC) lines from human skin fibroblasts. The approach leverages shallow sequencing coverage to assess the clonality of the lines, clusters redundant lines and sums their coverage to accurately discover mutations in the corresponding lineages. Only a fraction of lines needs to be sequenced to high coverage. We show that using shallow-sequenced 72 iPSC lines from a living individual, this strategy can recapitulate and extend a lineage tree generated using high coverage data. We discuss and propose an optimal experimental design for reconstructing lineage trees.

Omics Technologies Posters - Wednesday

PB3037. Induced pluripotent stem cells (iPSCs) for functional genetic screen applications.

Authors:

N. Isachenko, D. Hu, A. Chenchik, D. Tedesco; Cellecta, Inc., Mountain View, CA

Abstract Body:

Induced pluripotent stem cells (iPSCs) are widely used for disease modeling, drug discovery, and cell therapy development. However, iPSCs are difficult to engineer with an efficient CRISPR/Cas9 system for functional genetic screens. We used WTC11 human iPS cells to generate functionally validated lentiviral transduced Cas9 lines, characterized by high gene-editing activity and sustained potential to differentiate. Flow cytometry was used to check for the expression of the pluripotency markers Oct3/4, TRA1-60, and SSEA-4 in the iPS-Cas9 cells, and their actual pluripotency was then confirmed by testing the ability to differentiate into the three germ layers: ectoderm, mesoderm, and endoderm. Furthermore, genome-wide transcriptome analysis was performed to confirm the expression of germ-layer specific markers in the differentiated cells. Sustained Cas9 activity was confirmed in the differentiated cells. This study provides proof-of-principle that patient-derived iPSCs can be used to enable CRISPR/Cas9 functional genetic screening technology in reconstituted patient-specific tissues/disease models.

Omics Technologies Posters - Thursday

PB3038. Infection with SARS-CoV-2 leads to variant-specific changes in gene expression in airway cell lines and primary cell cultures.

Authors:

A. Rustagi¹, D. T. Bravo¹, M. Leong¹, A. Beck¹, T. Cormier², J. V. Nayak¹, C. A. Blish¹; ¹Stanford Univ., Stanford, CA, ²Parse BioSci.s, Seattle, WA

Abstract Body:

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has caused a global pandemic of illness, disability, and death unmatched since the flu epidemic of 1918. People infected with SARS-CoV-2 exhibit symptoms ranging from asymptomatic to mild airway infections through to severe impacts on lung function and possibly death. The virus spreads through aerosols and droplets exhaled by infected individuals that, in turn, come into contact with mucous membrane tissues of uninfected individuals. The virus can infect a range of cell and tissue types within the human body primarily by using the Spike (S) protein on its surface to bind to angiotensin-converting enzyme 2 (ACE2) displayed on cell surfaces. Once the virus releases its RNA and proteins into the cell cytoplasm, the cell begins to copy and translate the viral genome and assemble new viral particles, which ultimately escape the cell and propagate the infection within the body and between individuals. Questions remain about the impacts of viral infection on host gene expression in different cell types, including the induction of innate immune genes. To investigate these questions, we cultured a human lung epithelial cell line (A549-ACE2) as well as primary cultures of nasal tissues from healthy donors. We induced the primary cell cultures to differentiate into cultured nasal cells and then infected both the epithelial cell line and the cultured differentiated nasal cells with various strains of SARS-CoV-2 virus. We used the Parse Biosciences Evercode™ Whole Transcriptome RNA-sequencing kit to investigate gene expression in these infected cells. We identified the cells present in the differentiated nasal cultures using canonical cell markers and examined changes in gene expression. Our current research focuses on the impacts of SARS-CoV-2 infection on the human transcriptome in hopes of better understanding the biology of disease progression and immune response.

Omics Technologies Posters - Wednesday

PB3039. Inference of menopausal status from female reproductive tissue gene expression signatures

Authors:

E. Theusch; Univ. of California, San Francisco, Oakland, CA

Abstract Body:

Unfortunately, the extensive endocrine changes that accompany menopause result in physiological changes throughout the body that can negatively influence women's health. One major barrier to uncovering the mechanisms of these changes is the lack of menopausal status information in many existing datasets, including most human tissue gene expression datasets. Due to the responsiveness of reproductive tissues to female sex hormones and the correlations of ovarian follicle-produced and female sex hormone-regulated genes with age that we observed in our initial studies of uterus and ovary data, we sought to infer women's menopausal status using female reproductive tissue gene expression data. Since GTEx (Genotype-Tissue Expression) includes many tissues contributed by each donor, this can then be applied to other tissues for downstream analyses. To extract a strong menopause-related signal from library size-adjusted and variance-stabilized gene expression data, we used PEER (probabilistic estimation of expression residuals) to infer "hidden factors" that corresponded to unmeasured covariates, while including known covariates. Importantly, we withheld age as a covariate since it is strongly correlated to menopause status, our unmeasured variable of interest. In the uterus, hidden factor 2 (HF2) showed the greatest correlation with age, and, importantly, this relationship was not linear but had an inflection point around age 51, the average age of menopause, supporting the likelihood that it is a proxy for menopause status. 96% (72/75) of uterus tissue samples with $HF2 < 0$ were collected from women under 54 years old (inferred premenopausal), and 95% (64/67) of samples with $HF2 > 0$ were from women over 48 years old (inferred postmenopausal), demonstrating the high consistency of menopausal status inferences based on the $HF2 = 0$ threshold and age. We further validated uterus HF2 as a viable menopausal status marker by cross referencing with pathology notes recorded during ovarian tissue collection from many of the same donors. 96% (27/28) of donors for which ova, follicles, and/or corpora lutea were noted in the ovaries had uterus $HF2 < 0$. Furthermore, 92% (11/12) of donors with ovaries noted to be atrophic had uterus $HF2 > 0$. Interestingly, ovaries that were noted to have a postmenopausal appearance had a wide range of uterus HF2 values, potentially reflecting that ovarian changes precede gene expression changes in the uterus. Future extension of premenopausal versus postmenopausal gene expression comparisons to non-reproductive tissues could provide molecular-level insights about the changes in health and disease risk that women experience post-menopause.

Omics Technologies Posters - Thursday

PB3040. Insights from HLA transcriptome analysis of a cohort from the Qatar Genome Program

Authors:

A. Fadda, H. Naeem, B. Lo, Y. Mokrab; Sidra Med., Doha, Qatar

Abstract Body:

The Human Leucocyte Antigen (HLA) is an essential player in the immune system, controlled by a set of genes on chromosome 6 in the human genome. This set of genes expresses proteins presented on the cell surface, for the main role of presenting foreign antigens for targeted destruction. Hence, an array of diseases has been associated with HLA, such as cancer, autoimmune diseases, and transplant rejection. HLA genes are highly polymorphic for the purpose of fine tuning the immune response, with thousands of alleles documented so far. There is evidence that in addition to allelic identity, expression levels contribute to disease status as well: around 50% of reported loci in the GWAS catalogue are in intronic and regulatory regions of HLA genes. Therefore, we aimed to study the transcriptomics of 13 class I & II genes and their association with phenotype. Through Qatar Biobank and the Qatar Genome Program, around 2000 Qatari samples have genomic and RNAseq data, demographic, biochemical, proteomic, and medication data, and various questionnaires, providing a very rich database of over 2000 phenotypic parameters per subject. We used a specialized analysis pipeline to infer the HLA haplotype from the genomic data of each subject, then used a personalized reference to quantify the per-allele expression from the RNAseq data. We have ~ 400 alleles in total for the 13 genes, with HLA-B, -A, -DRB1 having the most alleles respectively. One-way-ANOVA shows allele-specific expression. Alleles' expression per gene were summed up to create gene-level expression. Our results show that, apart from HLA-G, class I genes show higher expression, with HLA-C being the most abundant, followed by HLA-B, -A, -E, and -F. Surprisingly, association testing between individual gene expression and phenotypic data was not significant (with correction for multiple testing), with few exceptions including thrombosis, chest pain, stroke, angina, asthma, and hemorrhoids. To evaluate non-gene-specific HLA expression, RNAseq data were quantified against the entire MHC region. This total-HLA expression was tested for associations with the phenotypic data. The most striking association was with gender, with females having slightly higher total HLA expression level than males. In conclusion, our dataset is unique in its richness and the ethnicity of its subjects, and the analysis will result in more distinctive findings.

Omics Technologies Posters - Wednesday

PB3041. Insights into Molecular Pathways Associated with Juvenile Myositis through Single-Cell Transcriptome Profiling

Authors:

C. Kao, H. Qu, J. Garifallou, R. Pellegrino di Silva, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Inflammatory myopathies (IM)-including dermatomyositis (DM), polymyositis (PM), and inclusion body myositis (IBM)-are a group of rare autoimmune diseases most often characterized by chronic muscle inflammation leading to progressive skeletal muscle weakness and loss of function, where cutaneous involvement may be present (in DM) and systemic problems occur in severe cases. IM presentation is heterogenous, where amyopathic forms (without apparent muscle involvement) can manifest. Juvenile-onset IMs are generally considered distinct but closely related clinical entities from their adult-onset forms. The etiology of IM is unknown, although there is a clear genetic/heritable component. Steroids and other immunomodulatory agents have been transformative, but some cases remain difficult to treat/manage. We are collecting a biobank of samples from families with juvenile IM cases, with the aim of using "omics" technologies to find potentially informative biomarkers to guide use of existing and possibly new therapies. One of the technical challenges posed with blood sample collection is that it may not always be practical to process and harvest cells immediately, since the collection site may not have the necessary equipment, tools, and staff, where instead unprocessed blood may need to be shipped first, which then introduces a delay of ~2-3 days between collection and processing. The impact on single-cell gene expression signatures of peripheral blood mononuclear cells (PBMC) upon prolonged storage prior to processing has not been examined, thus we collected blood from healthy volunteers into EDTA-coated tubes and isolated (and froze down) PBMCs immediately or stored the blood at 4°C for 24, 36, or 48 h prior to PBMC isolation (and freezing). The samples were then analyzed following single-cell transcriptional profiling using the 10x Genomics Chromium platform and sequencing via Illumina NovaSeq (i.e. single-cell RNASeq, or scRNASeq). The impact of prolonged delays in processing on the transcriptome signatures of PBMCs was assessed, and a "delay" signature was identified which was used as a means for correcting and filtering out genes most sensitive to delays in sample processing. We then analyzed the scRNASeq dataset derived from PBMCs collected from juvenile IM subjects which had experienced processing delays and evaluated the differential expression signature to healthy controls before and after filtering/correction for the delay signature. The impact on type I IFN and mTOR pathway signature was also examined.

Omics Technologies Posters - Thursday

PB3042. Integrated genetic analysis of transcriptome sequencing data in congenital diaphragmatic hernia.

Authors:

L. Qiao¹, H. Zhao², R. Hernan³, E. Byington², G. Zhong⁴, J. Wynn⁵, L. Yu⁶, X. Zhou⁷, D. Potoka⁸, A. Wagner⁹, B. Warner¹⁰, G. Mychaliska¹¹, M. Danko¹², D. Chung¹², R. Cusick¹³, T. Crombleholme¹⁴, F. Lee¹⁵, G. Aspelund¹⁶, V. Duron¹⁶, Y. Shen⁴, W. Chung⁴; ¹Dept. of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY, ²Dept. of Systems Biology, Columbia Univ. Irving Med. Ctr., New York, NY, ³CUIMC, New York, NY, ⁴Columbia Univ., New York, NY, ⁵New York Presbyterian Columbia, New York, NY, ⁶Columbia Univ, New York, NY, ⁷Columbia Univ Med Ctr, New York, NY, ⁸Div. of Pediatric Surgery, Dept. of Surgery, Univ. of Pittsburgh, Pittsburgh, PA, ⁹Div. of Pediatric Surgery, Children's Hosp. of Wisconsin, Med. Coll. of Wisconsin, Milwaukee, WI, ¹⁰Div. of Pediatric Surgery, Washington Univ. Sch. of Med., St. Louis, MO, ¹¹Section of Pediatric Surgery, Fetal Diagnosis and Treatment Ctr., Univ. of Michigan Hlth. System, Ann Arbor, MI, ¹²Dept. of Pediatric Surgery, Monroe Carell Jr. Children's Hosp. at Vanderbilt, Nashville, TN, ¹³Boys Town Natl. Res. Hosp., Omaha, NE, ¹⁴Med. City Children's Hosp., Dallas, TX, ¹⁵Div. of Pediatric Gen. and Thoracic Surgery, Cincinnati Children's Fetal Care Ctr., Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, ¹⁶Dept. of Pediatric Surgery, Columbia Univ. Med. Ctr., New York, NY

Abstract Body:

Congenital diaphragmatic hernia (CDH) is a life-threatening birth defect characterized by incomplete development of the diaphragm that is often accompanied by other congenital anomalies. Although rare *de novo* and inherited coding variants identified from exome and genome sequencing contribute to about 20% of CDH patients and have identified a number of candidate risk genes, the genetic causes of most CDH cases remain unexplained. Here, we performed RNA sequencing (RNA-seq) of diaphragm cells of 153 CDH cases to investigate the utility of RNA-seq in genetic analysis. All cases were previously exome or whole genome sequenced. Forty-nine cases (28.8%) have putative pathogenic *de novo* variants. We normalized the expression level across samples to remove technical batch effects, and then quantified cis- and trans-effects of putative pathogenic variants to confirm their functional role at the cellular level. We confirmed that most genes with likely gene disrupting variants have nonsense mediated decay (NMD) that affects expression levels for that gene. We performed an annotation-free algorithm able to detect splice sites *de novo* to identify aberrant splicing events by testing for differential splicing in each subject against the others. We found functional effects of three important RNA processing genes, *HNRNPC*, *ALYREF* and *CSTF2*. The sample from the carrier of a *HNRNPC* deleterious missense variant has the highest global rate of intron retention among all samples. Additionally, *GATA4* and *GATA6*, key transcription factors in developing heart and diaphragm, have extremely low expression levels in ~10% CDH subjects, all without known deleterious variants in *GATA4* or *GATA6*. The samples with low *GATA4* expression have low expression of genes potentially regulated by *GATA4*. In conclusion, we demonstrated that RNA-seq data of patient tissues can confirm the functional impact of putative pathogenic variants at the cellular levels and that *GATA4* and *GATA6* regulatory networks are a potential convergent pathway in CDH.

Omics Technologies Posters - Wednesday

PB3043. Integrated Proteotranscriptomics Analysis of Breast Cancer using Variational Autoencoders to Prioritize Pathogenic Genes

Authors:

J. Jhee¹, M-Y. Song², B. G. Kim³, H. Shin^{4,5}, S. Lee¹; ¹Ctr. for KIURI Bio-Artificial intelligence, Ajou Univ. Sch. of Med., Suwon, Korea, Republic of, ²Div. of Cardiology, Dept. of Internal Med., Kyung Hee Univ. Hosp., Kyung Hee Univ., Seoul, Korea, Republic of, ³Dept. of Brain Sci. and Neurology, Ajou Univ. Sch. of Med., Suwon, Korea, Republic of, ⁴Dept. of Industrial Engineering, Ajou Univ., Suwon-si, Korea, Republic of, ⁵Dept. of Artificial Intelligence, Ajou Univ., Suwon, Korea, Republic of

Abstract Body:

Because cancer phenotypes with high heterogeneity are controlled by diverse mechanisms following multiple points in the Central Dogma, various deep learning approaches have been applied to identify biomarkers that have representative characteristics of integrated multi-points omics data. To overcome the imbalance which multi-omics data generally have a characteristic that has high dimensions compared with relatively few patient samples, many researches applied variational autoencoder (VAE) that is known to have strength in handling high dimensional data with few samples for its innate generative modeling process. This study develops methods to prioritize key breast cancer regulator genes based on a unified representation applying the compressed features learned through VAE using proteotranscriptomic data of identical 105 breast cancer patients from TCGA and CPTAC. We computed encoding weighted scores (EWS) which deduced each lower-dimensional encoding feature with VAE using pre-processed omics data, and conducted survival analysis using Cox hazard model with the survival period and EWS to select significant encoding features (SEF). The SEF were decoded through VAE, and we computed the p-value per each gene based on decoding weight score, and calculated the accuracy using breast cancer-related genes from the knowledgebase. Survival analysis results show that each PT, P and RNA data contains 12.9%, 12% and 19.3% of significant encoding features, respectively. To verify the effect of the unified representation, we compared accuracy measurements, significant encoding features which belong to proteotranscriptomic data represented not only represent the highest accuracy of 0.98 at significance level of 0.01 but also showed an increasing trend of accuracy as decreasing significance level. Based on the results, we can conclude that our methods based on VAE to prioritize cancer-associated gene set showed remarkable performance at integrated multi-omics than mono-omics data.

Omics Technologies Posters - Wednesday

PB3045. Integrating Knowledge in Breast Cancer Subtyping Deep Learning Classifiers

Authors:

J. Davidson, M. Kressman, J. Tang, H. Lakshman Kumar, N. Zarate, L. Garabedian, B. Aduaka, C. Janssens, P. Kim, P. Anderson; Cal Poly, SLO, San Luis Obispo, CA

Abstract Body:

The large amounts of data present in biomedical research provides an incredible opportunity for precision medicine, and also a complicated question on how to use that data in a biologically and diagnostically meaningful way. Deep learning neural networks have been utilized to improve precise diagnostics and treatment options in many cancer fields, but often fail due to issues with reproducibility, underspecification and subtle adjustment to certain parameter combinations which may perform poorly when the test distribution differs from the training distribution. We seek to address these complications by embedding prior knowledge from the biological research literature. We hypothesize that embedding domain specific knowledge into classifiers can boost predictive models and increase their functional utility. To test this hypothesis, we focus first on breast cancer due to the wealth of knowledge available, especially related to subtype biomarkers and genetic signatures. Here we will present our work translating clinical and genetic data into graph knowledge which is then embedded into our modified neural network classifier. We embed this prior knowledge into the loss function and can easily test the impact of various levels of clinical data (mutation rate, patient metadata, tumor gene expression, etc) obtained from the METABRIC study. Our results show that our approaches reduce variability in our deep learning architecture and increase consistency. We demonstrate the specificity of these knowledge integrations to the outcomes of biomarkers identified through pathway enrichment studies. In general, our novel model can apply to a broad range of biomedical questions to access machine learning classifiers with improved biological confidence that the results are meaningful and reproducible.

Omics Technologies Posters - Thursday

PB3046. Integrating multiple human toxicogenomics datasets for drug repositioning

Authors:

M. Leclercq; Université Laval, Québec, QC, Canada

Abstract Body:

Introduction: A few publicly available toxicogenomics databases exist, which include many molecules exposed on various human cell lines from multiple tissues, with different time and dose expositions, generated from different technologies. By integrating such data we aim to identify molecules providing the same transcriptomics effect (molecular repositioning) in various human cells by automated learning. However, to perform such a task, the cell lines must present equivalent transcriptomic profiles after an exposition to a same molecule, at equivalent dose and time. Here, we evaluated different approaches to integrate such data and expose the identified groups of molecules presenting consistent transcriptomic profiles across datasets. We integrated experiments from various platforms, including L1000 from Connectivity Map, microarray data from Open TG-GATEs and RNA-seq from local datasets. Methods and results: The tested approaches to integrate the datasets included computation of differential expression (e.g. DeSeq2, GeoDE), data transformations and standardisation, and batch effect correction methods using various approaches (e.g. Combat, Harmony). We also tested deep learning algorithms (e.g. variational auto-encoders, domain adversarial neural networks) to integrate the data together. We succeeded in integrating microarray experiments and RNAseq, but L1000 was mitigated. Some molecules demonstrated a better integration than others. We finally used unsupervised learning to find molecules having the same corrected transcriptomic profiles. Integrating these data allowed us to evaluate the consistency of the impact of tested molecules on the various tested cell lines through gene signatures. Conclusions: Integration of toxicogenomics databases is very challenging, mostly due to the heterogeneity of the data, multiple batch effect sources, and experimental and biological variations. Many attempts to correct batch effects by the state of the art methods demonstrated a mitigated success to group together equivalent experiments. However deep learning showed promising results to handle batch effects and we were able to propose closely related molecules based on their impact on transcriptomic profiles in various cell lines.

Omics Technologies Posters - Wednesday

PB3047. Integration of single nucleus and bulk adipose RNA-seq reveals distinct adipocyte subtypes for obesity and insulin resistance.

Authors:

M. Alvarez¹, E. Rahmani², Z. Chen¹, O. Avram¹, B. van der Kolk³, K. Mohlke⁴, K. Pietiläinen³, E. Halperin¹, M. Laakso⁵, P. Pajukanta¹; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²Univ. of California, Berkeley, Berkeley, CA, ³Univ. of Helsinki, Helsinki, Finland, ⁴Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁵Univ. of Eastern Finland, Kuopio, Finland

Abstract Body:

Obesity predisposes to cardiovascular disease and type 2 diabetes. Several transcriptomic studies have characterized adipose tissue (AT) gene expression changes across cardiometabolic conditions. However, AT consists of a heterogeneous assembly of cell-types confounding bulk tissue analyses. Thus, we sought to map transcriptomic changes associated with cardiometabolic traits at cell-type resolution, and performed one of the largest single-nucleus RNA sequencing (snRNA-seq) of subcutaneous adipose biopsies to date from 111 individuals across 4 cohorts, resulting in 104,669 nuclei. Additionally, we integrated bulk subcutaneous AT RNA-seq of 335 participants from the METSIM cohort, of which 84 overlapped with the snRNA-seq samples. First, we searched for cell-type composition changes across metabolic phenotypes by using negative binomial regression and snRNA cell-type counts. We found that adipocyte and AT-resident myeloid cell-type proportions were negatively and positively correlated, respectively, with obesity, insulin resistance, and serum lipid levels. To validate these changes in the 335 bulk AT samples, we estimated cell-type proportions using Bisque and observed the same directions of correlations with traits. To search for subcell-type associations, we performed cell-type-specific differential expression (DE) in adipocytes, stromal, endothelial, and myeloid cells using snRNA-seq-derived pseudo-bulk. Among these tissue-resident cell-types, adipocytes demonstrated the highest number of DE genes, with 127 and 105 significant (FDR<0.05) for BMI and Matsuda index, respectively. Interestingly, extracellular matrix pathway genes were significantly up-regulated in obese and insulin resistant individuals' adipocyte nuclei, while their lipid genes such as *FADS3* were down-regulated. We validated these results by performing DE analysis in the 335 bulk AT METSIM samples after regressing out estimated cell-type proportions. We found that 112 (93%) and 86 (90%) overlapping snRNA adipocyte DE genes for BMI and Matsuda index showed the same directional effect in the bulk AT data, indicating that single-cell level references can help retrieve biological signals in larger bulk RNA-seq sets. Finally, canonical correlation analysis (CCA) between snRNA adipocyte and bulk AT expression showed that the first canonical vector recapitulated the observed single-cell-level adipocyte differences between the obese and insulin resistant individuals. Overall, our integrative transcriptomic approach revealed that obese and insulin resistant individuals' adipocyte subtypes differ from those of metabolically healthy individuals.

Omics Technologies Posters - Thursday

PB3048*. Integrative computational analysis of long-read transcriptomes identifies alternatively spliced poison exons in iPSC-derived brain organoids

Authors:

M. Broad, J. Hong, K-M. Lamar, J. Calhoun, G. Carvill; Northwestern Univ., Chicago, IL

Abstract Body:

Poison exons (PEs) are naturally occurring, evolutionarily conserved alternative exons that regulate the expression of genes by introducing a premature truncation codon (PTC), signaling nonsense-mediated decay (NMD) to degrade the transcript. Computational mapping and deep RNA-sequencing of the human and mouse cortex have revealed that ~33% of all alternatively spliced cassette exons are predicted to involve PEs. Additionally, PEs are cell type and tissue-specific, emphasizing their role in regulating gene expression spatially. For example, the PE of *SCN1A* is included in the mature mRNA transcript in neural progenitor cells (NPCs) and degraded by NMD. However, in mature neurons, the PE is skipped, as the full-length sodium channel is required for normal neuronal electrophysiological function. Pathogenetic variants in genes like *SCN1A* and *FLNA* can cause dysregulated PE splicing during neurodevelopment and are associated with neurodevelopmental disorders (NDDs) like epilepsy and autism spectrum disorder. Although PEs are integral in transcriptional regulation and proper neurodevelopment, little is known about the locations of PEs in the genome. Therefore, we have developed a computational analysis pipeline to annotate and identify alternatively spliced isoforms that include PEs in the human genome with a particular emphasis on neurodevelopment. We achieved this using whole transcriptome long-read sequencing of brain organoids generated from healthy female- and male-derived iPSCs at multiple time points. The brain organoids were either untreated as control or treated with the drug cycloheximide (CHX) treatment. CHX inhibits global protein translation and increases the abundance of PE-containing isoforms, thus increasing the likelihood of detection. We have previously successfully identified increased inclusion of PEs in known NDD-associated genes, including *SCN8A*, *SCN1A*, and *CHD2*. This experimental paradigm will identify PEs in human neurodevelopment that can be used to interpret variants from genome sequencing in individuals with NDDs. Ultimately, we will create a computational resource of PEs that the scientific community can use to identify therapeutic targets and create effective therapies to treat individuals suffering from severe NDDs.

Omics Technologies Posters - Wednesday

PB3049. Integrative single-nucleus multi-omics across 287 skeletal muscle biopsies reveals context-specific e/caQTL and extensive caQTL-GWAS colocalization

Authors:

S. Parker¹, A. Varshney¹, N. Manickam², M. Erdos³, P. Orchard², A. Jackson², N. Narisu⁴, H. Stringham², M. Laakso⁵, J. Tuomilehto⁶, T. Lakka⁷, K. Mohlke⁸, M. Boehnke², H. Koistinen⁶, F. Collins⁹, L. Scott¹⁰; ¹Univ MICHIGAN, Ann Arbor, MI, ²Univ. of Michigan, Ann Arbor, MI, ³NHGRI/NIH, Bethesda, MD, ⁴NIH, Bethesda, MD, ⁵Inst. of Clinical Med., Kuopio, Finland, ⁶Finnish Inst. for Hlth. and Welfare, Helsinki, Finland, ⁷Univ. of Eastern Finland, Kuopio, Finland, ⁸Univ North Carolina, Chapel Hill, NC, ⁹NHGRI, NIH, Bethesda, MD, ¹⁰Univ Michigan, Ann Arbor, MI

Abstract Body:

Skeletal muscle is the largest organ by weight (>40%) and is relevant for several polygenic phenotypes including type 2 diabetes (T2D) and related metabolic traits. Identifying genetic mechanisms underlying these traits requires pinpointing the set of causal variants, the target genes these variants regulate, and the specific cell populations in which they function. Here, we generated single nucleus chromatin (snATAC) and transcription (snRNA) profiles across 287 frozen human skeletal muscle biopsies. We integrated 414,982 nuclei across the modalities with LIGER and identified 13 cell types, ranging in abundance from 35% (type 1 fiber) to <1% (adipocyte) of all nuclei. We integrated genotypes and identified expression and chromatin accessibility quantitative trait loci (e/caQTL) in the five most abundant cell types (type 1, type 2a, type 2x, mesenchymal stem cell, endothelial). We identified 6,844 eQTL (250kb from TSS) and 100,928 caQTL (10kb from peak) at 5% FDR, ranging from shared to cell type-specific, and fine-mapped 268 eQTL and 4,502 caQTL signals to a single causal variant (95% credible set). We identified 1,637 eQTL-caQTL colocalizations (coloc5 posterior probability $H_4 > 0.5$); 977 (58%) were unique to one cell type. We performed heritability enrichments, colocalization, and chain-of-causality inferences of cell-specific e/caQTL with fine-mapped signals from UKBB, T2D, and related metabolic trait genome-wide association studies (GWAS). For all GWAS, we observed more colocalizations with caQTL compared to eQTL. For example, for waist-to-hip ratio (WHR) adjusted for BMI, 246 GWAS signals had caQTL colocalizations, with 86 (35%) detectable in only one cell type, highlighting the importance of sn-caQTL maps for GWAS functional studies. We identified multiple three-way GWAS-eQTL-caQTL colocalizations. One example is at a WHR adjusted for BMI GWAS signal, eQTL for SEMA3C, and a 20kb upstream intergenic caQTL peak, with strong evidence for chromatin accessibility being causal on gene expression (MR Steiger $P=0.00002$). Notably, SEMA3C is associated with exercise induced improvements of metabolism. These results inform the causal chain of molecular events that influence the complex genetic regulatory architecture of skeletal muscle at high-resolution epigenomic, transcriptomic, and cell state scales.

Omics Technologies Posters - Thursday

PB3050. Integrative single-nucleus multi-omics analysis identified candidate regulatory elements and variants and their target genes in Alzheimer's disease brains

Authors:

O. Chiba-Falek, D. Gingerich, J. Gamache, J. Barrera, M. Garrett, A. Ashley-Koch, G. Crawford; Duke Univ., Durham, NC

Abstract Body:

The genetic architecture of late-onset Alzheimer's disease (LOAD) is yet to be fully explored. While genome wide association studies (GWAS) discovered numerous LOAD-associated loci, the causal variants and their target genes remain largely unknown. Our overarching goal is to advance our understanding of the genetic underpinning LOAD. Here we applied a pipeline for systematic interrogation of regulatory elements and variants underlying gene dysregulation in LOAD in a cell-subtype resolution. We conducted a single-nucleus (sn)multi-omics study in *parallel* on the 10X Genomics platform using 24 frozen brain tissues, 12 normal and 12 LOAD. The *parallel* snRNA-seq and snATAC-seq collected from the same nuclei samples and at the same time were used in multimodal analysis to profile in gene expression and chromatin accessibility. The 202,223 snRNA-seq and 79,771 snATAC-seq nuclei were grouped into 33 and 26 clusters, respectively. We identified cell subtype specific LOAD-associated differential expressed genes (DEGs; total of 5,910 in all clusters), differential accessible peaks (DAPs; total 82,742) and *cis* co-accessibility networks (CCANs; 25,020) in LOAD. The *parallel* experimental design facilitated the performance of multi-omics integrative analysis and the following analyses focused on 309 LOAD CCANs, where at least one peak overlapped the promoter/intron 1 of a LOAD DEG and contained >1 DAPs. Transcription factor (TF) enrichment analysis identified 2-93 enriched motifs in 9 of the cell subtypes, out of which 1-10 TFs were LOAD DEGs in the corresponding cell subtype. Several TFs including, ZNF135, SP1 and JUN, were enriched in multiple cell subtypes. A subset of 12 CCANs in 7 cell subtypes overlap with known LOAD-GWAS regions that contain 47 (~10% of the total CCAN peaks) LOAD-associated candidate *cis* regulatory elements (cCRE) linked to 16 LOAD DEGs. Finally, we catalogue 35 putative functional SNPs within cell subtype-specific LOAD cCREs that disrupt TF motifs. Several examples including, SNPs within cCRE linked to *MYO1E*, *CUTA* and *RPS15* genes that affect the binding of TFs in the corresponding cell clusters of the microglia, excitatory neuron, and oligodendrocytes, respectively. To our knowledge this study represents the most comprehensive integrative single-cell genomics study in LOAD. We provide new insights into the interactions between the genome, epigenome, and transcriptome in LOAD brains in an unprecedented cell-subtype specific resolution. In summary, our findings enhance the translation of LOAD-genetic risk into mechanistic understanding of causation.

Omics Technologies Posters - Thursday

PB3052. Investigating somatic mutation rates and signatures across the length of the colon with duplex sequencing

Authors:

L. Hiatt^{1,2}, J. Kunisaki^{1,2}, S. Lulla², X. Nie³, Y. Guo³, J. Ramsay³, J. Horns¹, J. Baldwin-Brown⁴, J. Guo⁵, N. Phadnis⁴, T. Jenkins¹, K. Aston¹, J. Hotaling¹, A. Quinlan²; ¹Univ. of Utah Sch. of Med., Salt Lake City, UT, ²Univ. of Utah Dept. of Human Genetics, Salt Lake City, UT, ³Univ. of Utah Dept. of Oncological Sci., Salt Lake City, UT, ⁴Univ. of Utah Sch. of Biological Sci., Salt Lake City, UT, ⁵Chinese Academy of Sci. Inst. of Zoology, Beijing, China

Abstract Body:

Somatic mosaicism refers to the accumulation of mutations in non-germline cells throughout an organism's lifetime. Somatic mutations exist in healthy tissues but play a central role in disease pathogenesis, from developmental syndromes to cancer. In the colon, somatic mosaicism is a causative factor in colorectal cancer (CRC) and has been proposed to contribute to inflammatory bowel disease (IBD). These diseases cause significant mortality and morbidity: CRC is the 2nd leading cause of US cancer death, and IBD affects one in 200 individuals of European ancestry. These diseases have well-established regional presentations, with distinct pathologies along the length of the colon, such as "right" and "left" CRC and differently-clustered subtypes of IBD. However, the genetic etiologies underlying these region-specific pathologies are unknown.

Recent experimental and computational advances have made it possible to investigate the somatic mosaicism of healthy organs and somatic evolution to disease. However, we still lack a detailed characterization of healthy colon and how mutagenic processes may contribute to pathogenesis. To investigate the mutational landscape of healthy colon, we will apply duplex sequencing to biopsies across the length of cadaver-sourced colons. Duplex sequencing reduces error rates from 10⁻³ to <10⁻⁹, allowing exceptionally accurate mutation detection and measurement of mutation clonality. By extracting mutational signatures, we will determine whether there are endogenous or exogenous processes that may be enriched in specific parts of the organ. We will also evaluate mutation rate gradients as they may vary across the colon. Investigating these processes will provide insight into regional pathologies and their genetic etiologies and guide future clinical strategies for disease prevention, screening, and management. We are currently collecting cadavers via the Utah-based organization DonorConnect and establishing a computational workflow to quantify and interrogate region-specific mutational landscapes in the colon. We expect to present preliminary data based on a pilot study on a subset of acquired colons that have undergone duplex sequencing.

Omics Technologies Posters - Wednesday

PB3053. ioSearch: A tool for searching disease-associated interacting omics.

Authors:

S. Das, D. Srivastava; St. Jude Children's Res. Hosp., Memphis, TN

Abstract Body:

Biomarkers are important predictors of disease onset/progression and hence play a vital role in the prediction of patient survival and/or response to therapy. But the key challenge in finding biomarkers for complex diseases is, decoding the intricate interplay of multiple omics data. Although information from multiple omics data intuitively provides substantial information on the disease compared to a single source of data, it is difficult to analyse such correlated information from multiple omics together due to the differences in the data structure emerging from different assays, less sample size than the number of features under study etc. On one hand, translation deregulation has been held responsible for disruption in the normal mechanisms of individuals while it is also well-known that aberrations in genes disrupt downstream protein formation. Thus, intuitively studying the alteration in protein may provide substantial insight into disease progression, when integrated with the transcriptomic data. But most analytical procedures are unable to integrate subject-specific multi-omics data for a disease outcome because of the high dimensionality and diversity in the nature of multi-omics data. We propose an algorithm ioSearch, to identify novel omics signatures that may differentiate between two subtypes or groups, in a single data integration framework. Our method is based on principal components that select important transcriptomic and proteomic features associated with disease and explores the inter-relation of the most informative features using a multivariate model, to identify disease-associated multi-omics biomarkers. Simulation results show that our method is powerful with a controlled type I error rate. On application of ioSearch to the protein and gene expression data of nearly 800 breast cancer patients from TCGA database, we identified a list of genes and proteins most of which are reported to have interacting functional implications across independent experimental studies.

Omics Technologies Posters - Thursday

PB3054. Is single nucleus ATAC-seq accessibility a qualitative or quantitative trait?

Authors:

Z. Miao¹, J. Kim²; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Single nucleus ATAC sequencing (snATAC-seq) is a technique that detects open chromatin for each individual cell. While it is a key assay for gene regulation analysis, existing computational methods for snATAC-seq show a fundamental inconsistency in how data is quantified. Most computational pipelines binarize the data while others retain quantitative information. By analyzing several public datasets, we show that even with sparse single cell data, quantitative counts are correlated with functional outputs such as different levels of gene expression. Thus, the quantitative counts are informative for estimating a cell's regulatory state, which calls for consistent treatment. By investigating the possible molecular events associated with snATAC-seq assays, we propose a new Paired-Insertion-Counting (PIC) as a uniform method for snATAC-seq feature characterization. PIC has theoretical and practical advantages over existing snATAC-seq feature counting methods and is conservative. In addition, we derive the distribution of PIC counts taking into account the experimental protocols (i.e., Tn5 insertions and amplification), which can be used to derive a model-based quantification for snATAC-seq data.

Omics Technologies Posters - Wednesday

PB3055. IsoMiGA: Combining long and short read RNA-seq in human microglia reveals novel isoforms and splicing events and insight into Alzheimer's disease

Authors:

E. Brophy¹, J. Humphrey², R. Kosoy³, B. Zeng¹, E. Coccia⁴, D. Mattei¹, A. Ravi⁵, A. G. Efthymiou⁴, B. Muller¹, G. J. Snijders², C-F. Tsai⁶, S. P. Kleopoulos¹, T. Liu⁷, G. E. Hoffman⁴, A. Goate⁴, L. de Witte¹, J. F. Fullard¹, P. Roussos⁴, T. Raj⁴; ¹Mount Sinai, New York, NY, ²Mount Sinai Sch. of Med., New York, NY, ³Mount Sinai Sch. Of Med., New York, NY, ⁴Icahn Sch. of Med. at Mount Sinai, New York, NY, ⁵Icahn Sch. of Med., Mount Sinai, New York, NY, ⁶Pacific Northwest Natl. Lab., Richland, WA, ⁷Rutgers Univ, Piscataway, NJ

Abstract Body:

Genome-wide association studies in Alzheimer's disease (AD) have identified multiple non-coding genetic loci nearby, or within, genes expressed in myeloid cells of the innate immune system. Microglia are the primary myeloid cell type in the brain and are therefore of high interest for the genetics of AD. We recently generated the Microglia Genomic Atlas (MiGA), a genetic and transcriptomic resource comprising 255 primary human microglia samples from 100 human donors. We found several AD-associated risk variants that act through either expression or mRNA splicing of specific genes in microglia. While we localized the expression-altering variants to microglia-specific enhancers, due to the limitations of short-read RNA-seq, it is more challenging to dissect how these variants alter splicing. Here we use long-read RNA-seq (IsoSeq; Pacific Biosciences) data to create a microglia-specific isoform reference, isoMiGA, to identify novel transcripts and give insight into genetic associations between AD risk and splicing. isoMiGA is constructed from 30 IsoSeq libraries from microglia purified from 27 post-mortem donors for a total number of 89 million long reads. We used the isoMiGA reference to impute isoform expression in 474 short read RNA-seq post-mortem microglia samples from 310 human donors from 3 different independent multi-ancestry cohorts. Additionally, we performed short-read RNA-seq on 3 iPSC-derived microglia lines stimulated with either lipopolysaccharide or interferon-gamma. After filtering, isoMiGA contained a total of 81,170 transcripts, of which 51% are novel compared to GENCODE v38. Interestingly, we identified 1,973 novel fusion transcripts that we explored further. A fusion transcript contains sequences from two or more neighboring genes spliced together by an unknown mechanism. Using microglia-specific ATAC-seq data, we see that 81% of the fusion transcripts have an overlapping transcription start site. 1,877 fusion transcripts have an open reading frame, of which 911 would potentially produce a protein product. We validated multiple fusion transcripts using RT-PCR to verify gene-spanning junctions and observed an overlap in fusion transcripts with those found in the human frontal cortex. We observed that multiple fusion genes and other novel genes were differentially expressed and/or differentially spliced in response to pro-inflammatory stimulation. Overall, long read sequencing of microglia enables the identification of novel isoforms and a deeper understanding of cellular and disease biology.

Omics Technologies Posters - Thursday

PB3056. Large-scale single-cell RNA-sequencing in Japanese identified biological and host genetic involvement of innate immune cells in COVID-19 severity.

Authors:

R. Edahiro¹, Y. Shirai¹, Y. Takeshima², S. Sakakibara², Y. Yamaguchi¹, T. Murakami¹, T. Morita¹, Y. Kato¹, Y-C. Liu², D. Motooka², Y. Naito², A. Takuwa², F. Sugihara², K. Tanaka², J. Wing², K. Sonehara¹, Japan COVID-19 Task Force, H. Namkoong³, H. Tanaka³, H. Lee⁴, K. Fukunaga³, H. Hirata¹, Y. Takeda¹, D. Okuzaki², A. Kumanogoh¹, Y. Okada^{1,5,6,7}; ¹Osaka Univ. Graduate Sch. of Med., Osaka, Japan, ²Osaka Univ., Osaka, Japan, ³Keio Univ. Sch. of Med., Tokyo, Japan, ⁴Keio Univ. Sch. of Med., Tokyo, Japan, ⁵Osaka Univ., Immunology Frontier Res. Ctr. (WPI-IFReC), Osaka, Japan, ⁶RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, ⁷the Univ. of Tokyo, Graduate Sch. of Med., Tokyo, Japan

Abstract Body:

While multiple studies have highlighted dysregulation of complex networks of peripheral blood immune responses in COVID-19 using single-cell RNA-sequencing (scRNA-seq) analysis, the immune response of the host to SARS-CoV-2 still remains unclear. COVID-19 GWAS have demonstrated the influence of human genetic background on pathogenesis of COVID-19. However, functional analysis has mostly focused on the variant at *LZTFL1* on 3p21 which showed the strongest severity association in Europeans but conferred a rare frequency of the risk allele in East Asian. We analyzed single-cell transcriptome and T/B cell receptor of over 600,000 peripheral blood mononuclear cells from 53 COVID-19 patients and 51 healthy controls of Japanese ancestry with host genetics data. Differential abundance analysis revealed that the proportion of non-classical monocytes (ncMono) decreased in COVID-19 patients, especially in severe COVID-19. RNA velocity analysis identified the downregulation of the cell transition from classical monocytes to ncMono in COVID-19 patients. We observed the downregulation of *CXCL10*, which is IFN- γ induced gene and reported to be involved in COVID-19 severity by proteomics analysis, in ncMono of severe COVID-19, and the depletion of *CXCL10/CXCR3* interaction between ncMono and plasmacytoid dendritic cells (pDC) or activated T cells despite IFN- γ signal from activated T cells. Cell-cell communication analysis inferred that the cellular interactions involving ncMono and pDC were reduced in severe COVID-19 compared to moderate COVID-19. Clonal expansions of B cell receptor were most evident in plasmablasts of severe COVID-19. The putative disease genes identified by the GWAS for severe phenotypes (i.e., hospitalized COVID-19 and very severe COVID-19) showed cell type-specific expressions in monocytes and dendritic cells, whereas no cell type was enriched in the self-reported infection GWAS. Context and cell type-specific expression quantitative trait loci (eQTL) effects of COVID-19-associated risk variants preferably enriched in monocytes of COVID-19 patients (e.g., *IFNAR2*). In summary, our data linked innate immune cells dysfunction, especially ncMono, with severe COVID-19 and demonstrated the enrichment of host genetic risk in innate immune cells, indicating biological and host genetic critical involvement of innate immune cells in COVID-19 severity.

Omics Technologies Posters - Wednesday

PB3057. Learning nonlinear causal relations in complex biological systems using a deep-learning approach and knockoff statistics

Authors:

H. Park¹, Z. Fan¹, K. Kernan¹, P. Benos¹, S. Canna², J. Carcillo¹, S. Kim¹; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Learning the causal structure helps identify risk factors, disease mechanisms, and candidate therapeutics of complex diseases for future evaluation. Although the complex diseases would render variable interactions highly nonlinear, most bioinformatic methods of causal inference cannot identify the nonlinear relationships and estimate their effect size. To address this limitation and enable a more realistic causal structure learning, we developed the first method that learns both linear and nonlinear causal relations and estimates the effect size, causal Directed Acyclic Graphs using deep-learning VArIable SElection (DAG-deepVASE). DAG-deepVASE formulizes a variable selection problem for causal learning and solves this problem using a deep-learning approach with knockoff statistics. We further demonstrated that it consistently outperforms existing methods in identifying known and new causal relations in the following simulation and biological data. First, in simulation data of various sample size (200 to 1000), number of variables (50 to 3000), and different nonlinearity degree (complete- or partial-nonlinear simulation), DAG-deepVASE identifies 97% of true causalities without false positives, while the other methods could identify less than half of true nonlinear causalities with several false positives. Second, using serum cytokine data of 404 children with sepsis, DAG-deepVASE identifies 98 linear and 117 nonlinear associations. Although existing methods could identify the linear associations, the nonlinear associations were discovered only by DAG-deepVASE. By incorporating the linear and nonlinear causal relations, DAG-deepVASE could rediscover complex causal relations validated in sophisticated experimental models, including a beneficial one from interferon-gamma to tumor necrosis factor-alpha mediated by lipopolysaccharide. Third, using a data set of 214 nutrient intakes, 87 gastrointestinal bacteria genera and body-mass index collected from 90 healthy volunteers, DAG-deepVASE identified not only validated but also novel causal interactions by estimating the nonlinear effect size. Fourth, using the gene expression and clinical variables of the TCGA breast cancer data, DAG-deepVASE enables to understand complex biological interactions not only within genes and clinical features separately, but also across genes and clinical features. Altogether, DAG-deepVASE discovers the complex pathobiological interactions involving both linear and nonlinear causal relations, which is not possible using other methods.

Omics Technologies Posters - Thursday

PB3058. Leveraging a multi-omics Parkinson's Disease dataset to understand the role of mitochondrial transcriptional control in the aetiology of Parkinson's Disease.

Authors:

A. Fairbrother^{1,2}, A. Gil Martinez¹, M. Ryten¹, A. Hodgkinson²; ¹UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom, ²Kings Coll. London, London, United Kingdom

Abstract Body:

Mitochondrial (MT) dysfunction is known to contribute to the pathogenesis of an array of neurodegenerative diseases, including sporadic Parkinson's Disease (PD) for which the causal genetics remain largely unknown. Mutations in the nuclear-encoded mitophagy genes PINK1 and PRKN are important causes of autosomal recessive PD, which is strong evidence for the involvement of nuclear-mediated MT dysfunction in PD aetiology and underscores the importance of nuclear-MT interaction for normal cellular function. We aim to understand whether nuclear genetic variation is associated with variance of MT transcriptional processes (*trans*-QTLs), and whether these are disease-specific. We focus on two types of MT phenotype related to MT function: gene expression of 15 genes and RNA methylation at 15 independent sites. MT RNA methylation events are known to promote RNA stability and affect the activity of mito-ribosomes and cleavage enzymes. Thus, they are essential to MT function, and to our knowledge, have not previously been implicated in neurodegenerative disease. We leverage the case-control AMP-PD consortium dataset, consisting of multi-omic data from the blood of 1483 PD patients and 965 healthy controls. In total, we identify 56 *trans*-eQTLs (expression QTLs) and *trans*-meQTLs (methylation QTLs). We replicate 42.2% of associations in an independent dataset. We identify *cis*-eQTLs known to regulate several important genes in general MT transcriptional processes, including the nuclear-encoded MT RNase *MRPP3*, and *FASTKD4*, which are involved in MT mRNA processing and regulation of MT mRNA stability. Interestingly, 27% (15/56) of our significant associations represent interaction effects where genetic regulation of MT transcription is significantly different between PD cases and controls (FDR<0.05), pointing to formation of novel control mechanisms in the disease state. Three of our peak genetic variants overlap with those from a recent PD GWAS meta-analysis (P<0.05), which is more than expected by chance (P=0.016752), implicating MT processes in the causal biology of disease. Genes linked to these variants include *GPR155*, a G-protein receptor involved in cognition and *MAGII*, encoding a cellular component of the pre-synapse, *WIPF1*, involved in actin cytoskeleton organisation. We also highlight *ARPP21* (associated with MT methylation in cases), a cAMP-regulated phosphoprotein enriched in basal ganglia. In summary, we observe differences in nuclear genetic regulation of the MT transcriptome between PD cases and controls, pointing to a role for MT-nuclear co-ordination in the pathogenesis of sporadic PD.

Omics Technologies Posters - Wednesday

PB3059. Linking GWAS risk variants to disease genes by epigenomic mapping and prediction of functional enhancer-promoter interactions.

Authors:

Y. Guo, G. Ramaswami, G. Xiang, L. Chen, B. Matthews, J. Weinstein, B. Akerberg, Y. Liu, D. Bumcrot; CAMP4 Therapeutics, Cambridge, MA

Abstract Body:

Genome wide association studies (GWAS) have identified thousands of noncoding loci associated with human diseases and complex traits. These loci are strongly enriched in gene regulatory elements such as enhancers. Identifying the genes regulated by these enhancers promises to further reveal disease mechanisms. However, it remains a major challenge to link functional enhancers to their target genes. We developed enhancer-promoter interaction characterization (EPIC), a machine learning model for predicting functional enhancer-promoter (E-P) pairs. EPIC integrates epigenomic data such as HiC/HiChIP, ChIP-seq, and ATAC-seq, and uses CRISPRi-based enhancer perturbation screening data in K562 cells to train a random forest model to score E-P pairs. The features in the EPIC model include genomic distances, chromatin interaction frequencies of E-P pairs, biochemical activities of enhancers and promoters, and interactions among these features. We compared EPIC with the ABC model (Fulco et al., 2019), one of the best performing published methods, for predicting functional E-P interactions. In hold-out testing, EPIC (AUPR=0.611, AUROC=0.919) outperforms the ABC model (AUPR=0.451, AUROC=0.885) and the difference is statistically significant ($P = 7.6e-11$, Delong's test on ROC curve). We generated epigenomic data in human primary hepatocytes and discovered about 30,000 E-P interactions using EPIC. We compared EPIC with the ABC model in discovering target genes of a set of "gold-standard" curated GWAS loci of liver-related diseases and traits (Mountjoy et al., 2021). We find that EPIC (AUPRC=0.643, AUROC=0.879) is more accurate than the ABC model (AUPRC=0.337, AUROC=0.877) in distinguishing causal genes from neighboring genes. In addition, the EPIC scores significantly associate with the liver eQTL status of the E-P pairs ($P < 2.2e-16$). These results demonstrate the functional relevance of the E-P pairs discovered by EPIC. We further expanded our analysis to discover liver-related GWAS loci-gene pairs and conducted in silico mutagenesis analysis based on the Enformer model (Avsec et al., 2021) to predict the effect of enhancer deletions on gene expression.

In summary, EPIC enables accurate cell-type-specific prediction of functional E-P interactions using epigenomic data. It outperforms an established method in predicting E-P interactions and in linking GWAS loci to causal genes in a new cell type. Applying EPIC to diverse human cell types may help discover disease-causing genes and enable development of novel therapeutics that target enhancers of disease-related genes.

Omics Technologies Posters - Thursday

PB3060. Long read transcriptomes identify features not found with very deep short read sequencing

Authors:

T. Salas-Morris, N. Prasad, B. Umylny; Discovery Life Sci., Huntsville, AL

Abstract Body:

The limited capacity of long-read sequencing technologies has significantly limited researchers' ability to efficiently apply long read sequencing data in drug development. Discovery Life Sciences (Discovery) has addressed this challenge by building a service laboratory that utilizes over 20 Sequel IIe devices from Pacific Biosciences (PacBio) and incorporates liquid handling and robotics to optimize efficiency, cost, and accuracy of large-scale, long-read sequencing projects. To demonstrate the ability of this offering to discover novel isoforms that cannot be detected using established short read RNA-Seq technologies, Discovery partnered with researchers to analyze 12 human tumor samples using between 1 and 5 SMRT Cells per sample (3-18 million reads). When compared to short read sequencing results of the same samples, the data shows that even with only 3 million reads, long read technology can identify validated isoforms missed by very deep short read sequencing. Analysis by quality control and annotation tools further shows that not only are the isoforms identified by short read sequencing much shorter, but they also are much more likely to be false positives, making analysis of the short read transcriptomes more difficult.

Omics Technologies Posters - Wednesday

PB3061. Long-amplicon variant-robust genome capture of SARS-CoV-2 using molecular inversion probes.

Authors:

E. Boyden, A. Patel, J. Vieira, J. Amaral, G. Porreca, P. Saunders; Molecular Loop BioSci.s, Inc., Woburn, MA

Abstract Body:

Combating pandemics such as COVID-19 requires routine and systematic complete viral sequencing of positive samples to identify and track emergent variant strains. The PCR-based ARTIC assay is effective but prone to amplification dropouts when new mutations arise, which has required multiple primer revisions. We previously developed an alternative short-read sequencing assay that uses tiled molecular inversion probes (MIPs) for high redundancy (7.5X mean tiling depth), which improves coverage uniformity and insures against amplification and mutation dropouts, thereby facilitating superior genome completion rates across a broad range of Ct values. However, short-read NGS assays may show reduced sensitivity to modestly sized indels and be unable to phase distant variants. Sequencers by Pacific Biosciences and Oxford Nanopore Technologies can generate reads that are several hundred to thousands of bases long, with more power to identify indels and phase variants, but capitalizing on these advantages requires sufficiently long library molecules, which is not a common feature of published MIP assays. To establish that our highly scalable and easily automatable chemistry can produce templates that are optimal for long-read sequencing, we designed panels of MIPs to capture 99.5% of the SARS-CoV-2 Wuhan-Hu-1 genome in overlapping elements of either 675 or 1215 bp, with mean tiling depths of 22.5X or 40.5X respectively, and evaluated their performance on synthetic RNA from multiple strains using either a single-day or overnight workflow. The results demonstrate that a) our MIP-based SARS-CoV-2 genome capture assays are sensitive, specific, and robust, even for mutation-rich strains such as Omicron BA.1, and b) MIPs can capture much larger elements than commonly reported, which will improve their utility for long-read sequencing platforms.

Omics Technologies Posters - Thursday

PB3062. Long-read capture with Twist target enrichment system

Authors:

T. Han, H. Corbitt, L. Arbiza, E. Toro, C. C. Locklear; Twist BioSci., South San Francisco, CA

Abstract Body:

Targeted resequencing allows for high-resolution characterization of gene regions at a scale and cost that is more accessible than whole genome sequencing. While long-read PacBio HiFi sequencing has been shown to accurately and comprehensively interrogate complex clinically actionable loci, studies have been primarily focused on single genes using PCR amplicon-based methods. Here we describe a method to leverage Twist Bioscience target enrichment workflow for gene panels sequenced with HiFi reads. We designed gene panels of various target sizes, ranging from 0.2-20 Mb. Our long-read hybrid capture protocol starts with 200-1000 ng of fragmented gDNA that were sheared using mechanical fragmentation. After end-repair and dA-tailing, truncated Y-shaped adapters were ligated to adapted gDNA. A pair of 10-bp unique dual indices (UDIs) for sample barcoding were added during PCR. Multiple samples can be pooled for an overnight hybridization. The post-capture libraries then undergo SMRTbell library preparation and sequencing on PacBio Sequel II. Depending on target size, up to 400 samples may be multiplexed and sequenced in one SMRT Cell with HiFi read length of 5-10 kb. We demonstrate that this method efficiently enables comprehensive coverage of gene targets using Coriell samples run with multiple gene panels of varying sizes, which include complex regions like *CYP2D6*, *HLA*, *SMN1*, and *LPA*. This long-read hybrid capture protocol can be utilized with Twist custom or fixed gene panels to efficiently capture genes of interest using long-read sequencing. Optional secondary panels (spike-ins) can also be easily added during hybridization for additional content. The demonstrated method allows for scalable and cost-efficient hybrid capture with long read lengths, minimizing coverage bias, and maximizing accuracy to fully capture all variant types. This includes structural variation and haplotype phasing information which are inaccessible to short-read and Sanger sequencing.

Omics Technologies Posters - Wednesday

PB3063. Long-read isoform sequencing reveals aberrant splicing of *PSEN2*, but not *PSEN1*, in individuals with sporadic Alzheimer's disease.

Authors:

P. Valdmanis¹, K. Gudsnuk¹, C. D. Keene¹, T. D. Bird¹, S. Jayadev¹, M. M. Course^{1,2}; ¹Univ of Washington, Seattle, WA, ²Colorado Coll., Colorado Springs, CO

Abstract Body:

Alzheimer's disease (AD) is a common neurodegenerative disease, characterized by dementia and premature death. Early-onset familial AD is caused in part by pathogenic variants in presenilin 1 (*PSEN1*) and presenilin 2 (*PSEN2*). Sporadic AD is much more frequently observed than familial AD and shares the same hallmarks of pathology including amyloid beta aggregates and tau neurofibrillary tangles, without pathogenic variants in these genes. We hypothesized that improper RNA splicing of *PSEN1* and/or *PSEN2* could generate aberrant transcripts that act mechanistically in a manner similar to pathogenic variants in these genes and provide a direct connection between sporadic and familial AD. Currently available RNA-seq data of *PSEN1* and *PSEN2* is limited to short reads that are fragments of complete transcripts, which may not readily reveal pathogenic isoforms. We utilized a probe-based pull-down strategy and leveraged PacBio isoform sequencing (Iso-Seq) to characterize thousands of complete *PSEN1* and *PSEN2* transcripts in the prefrontal cortex of individuals with sporadic AD, familial AD (carrying *PSEN1* and *PSEN2* variants), and controls. While *PSEN1* splicing patterns were consistent between sporadic AD, familial AD and controls, our results reveal alternative splicing patterns of *PSEN2* specific to sporadic AD. Alternative splicing events include a human-specific cryptic exon present in intron 9 of *PSEN2* and a 77bp intron retention product prior to exon 6 that are both significantly elevated in sporadic AD samples, alongside a significantly lower percentage of canonical full-length *PSEN2* transcripts versus familial AD samples and controls. Both alternatively spliced products are predicted to generate a prematurely truncated PSEN2 protein and were corroborated in an independent cerebellum RNA-seq dataset. In addition, we observed far fewer full-length transcripts carrying pathogenic alleles versus wild-type alleles in *PSEN2* variant carriers. Finally, we identified frequent RNA editing at Alu elements present in an extended 3'UTR in *PSEN2*. Overall, this work expands the understanding of *PSEN1* and *PSEN2* variants in AD, shows that transcript differences in *PSEN2* play a role in sporadic AD, and point to novel therapeutic strategies for AD.

Omics Technologies Posters - Thursday

PB3064*. Long-read RNASeq in human brains aligned to T2T CHM13 complete human genome reveals new gene bodies and new transcripts, exons, & exon junctions in known genes.

Authors:

B. Aguzzoli Heberle¹, J. Brandon¹, K. A. Nations¹, M. Page¹, M. E. Wadsworth¹, D. W. Dickson², P. T. Nelson¹, J. B. Miller¹, J. Fryer², M. T. W. Ebbert¹; ¹Univ. of Kentucky, Lexington, KY, ²Mayo Clinic, Jacksonville, FL

Abstract Body:

Background: RNASeq experiments are generally performed using short-read sequencing technologies that collapse all RNA isoforms for a given gene into a single expression measurement—a **major oversimplification of the underlying biology**. A critical next step in biology research will be to determine the distinct roles individual RNA & protein isoforms play across tissue and cell types. Long-read technologies can sequence entire RNA molecules, allowing researchers to accurately quantify expression for the complete set of RNA isoform species, including new RNA isoforms. Long-read sequencing is especially well suited to discovering novel transcripts when aligning to the first complete telomere-to-telomere (T2T) human reference genome (CHM13). Here, we identify entirely new gene bodies, RNA transcripts/isoforms, exons, and exon junctions in human brain tissue. We also identified major differences between CHM13 and GRCh38 based on gene annotations and RNASeq results.

Methods: We sequenced pre-frontal cortex tissue from five post-mortem human brain samples using Oxford Nanopore Technologies long-read sequencing (cDNA). Data were basecalled using Guppy, reads were aligned to the CHM13 human reference genome using minimap2, and transcripts were assembled and quantified using Bambu.

Results: Among other findings, we discovered 146 new, high-confidence gene bodies expressed in all five samples with 5+ reads in each sample, including 16 with 2+ isoforms. We also found 428 new RNA isoforms in known gene bodies, where 44 are from medically-relevant genes. In total, we discovered 1185 new exons and 668 new exon junctions. *We also identified thousands of non-genic regions with high RNA expression throughout the completed genome, and hundreds of the genes discovered by the T2T consortium had high expression in our samples.* Lastly, we found that 43 genes were transposed to different chromosomes between the GRCh38 and the CHM13 assemblies, and hundreds of additional genes were found to be duplicated on other chromosomes between the assemblies.

Conclusions: Our results suggest long reads combined with the completed CHM13 human reference genome has the potential to reveal exciting new biology relevant to human health and disease, including new gene bodies and RNA isoforms that are overlooked with standard approaches. These methods can provide a more complete picture of the transcriptomic landscape across all tissue and cell types, with potential for clinically-relevant discoveries. We also highlight key gene transpositions and duplications that could affect linkage disequilibrium blocks when aligning samples to the CHM13 reference genome.

Omics Technologies Posters - Wednesday

PB3065. LongReadSum: A fast and flexible quality control tool for long-read sequencing data.

Authors:

J. Perdomo, M. U. Ahsan, Q. Liu, L. Fang, K. Wang; Raymond G. Perelman Ctr. for Cellular and Molecular Therapeutics, Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract Body:

Recent advances in long-read sequencing technologies have generated reads tens to thousands of kilobases long with high accuracy. These long reads have a broad range of applications in genomics, including but not limited to uncovering previously missed genetic causes of human diseases, detecting variants in difficult-to-map regions, assembling repetitive regions of the human, and identifying novel splicing isoforms. Prior to analyzing long read data, quality control (QC) checks are needed to ensure that the raw data from sequencers do not contain substantial errors and biases and to understand the basic characteristics of the sequencing run. While short-read sequencing technologies have well-established QC tools such as FastQC, there have been several challenges in implementing QC frameworks for long reads. First, long read data can be generated by different types of sequencing technologies (such as PacBio and Oxford Nanopore) with unique data formats, and the currently available QC tools usually only support one specific sequencing platform or only support some data formats. Second, long read data can be orders of magnitude larger than short reads (for example, a single flowcell on the Nanopore platform can generate 5TB of signal data), and thus it remains imperative to develop a tool that enables fast, high-throughput, and comprehensive QC summary statistics of reads. LongReadSum is a tool that addresses these challenges: It supports the four main data format types used across sequencing technologies (FASTA, FASTQ, FAST5, unaligned BAM, and aligned BAM) and can generate a comprehensive summary of different aspects of sequencing data in a timely manner by executing programs in a flexible multi-threaded C++ framework. Outputs are compiled into both a static and a dynamic HTML report customizable to the user's needs which contains basic statistics such as the total number of reads, base pairs, maximum, mean, and median read length, percent guanine-cytosine (GC) content, and the N50. The report also includes histogram plots of read length, base quality across reads, and average base quality per read. In addition, for Nanopore reads, we include platform-specific QC measures such as sequencing throughput by time. These statistics provide a comprehensive overview of all major aspects of read quality, enabling the identification of significant errors that may preclude downstream analyses. In conclusion, LongReadSum is a computational tool for fast, comprehensive, and high throughput long read QC with support for all major sequencer data types, and it can be adapted to custom long-read sequencing pipelines.

Omics Technologies Posters - Thursday

PB3066*. Low cost, noninvasive RNA-sequencing to enable massive scaling of transcriptome studies.

Authors:

M. Martorella^{1,2}, S. Kasela², R. Garcia-Flores^{1,2}, K. Buschur¹, P. Hoffman², A. Gokden², A. Vasileva², S. Castel², T. Lappalainen^{3,2}; ¹Columbia Univ., New York, NY, ²New York Genome Ctr., New York, NY, ³KTH Royal Inst. of Technology, Stockholm, Sweden

Abstract Body:

Transcriptomic studies disentangle functional mechanisms of gene expression regulation, however, many applications are limited by invasiveness and cost of sample collection and processing, or, by disease-relevancy of samples collected. Optimizing RNA-sequencing of noninvasive biospecimens facilitates discovery by scalable access to other cell types.

304 noninvasive samples (hair follicles, buccal swabs, saliva, and urine cell pellets) collected from 19 individuals over four time points were processed using a reduced-cost, in-house TSO and tagmentation reaction, SMART-seq, and Illumina TruSeq commercial kits. We found low-input methods performed best, but samples passing QC showed comparable metrics and high replication regardless of preparation. Urine and hair follicles provided high-quality data equivalent to cell line RNA, while saliva samples showed a high percent of unmapped reads and typically failed QC. Remapping unmapped reads across all tissues showed microbial signatures characteristic to each, and saliva may be more valuable in microbiome study designs. Tissue was the main driver of gene expression variability across all samples, and sample donor contributed most within a tissue, indicating capture of biological differences rather than technical factors. Cell type deconvolution showed hair, buccal, and urine samples primarily captured epithelial cells, whereas saliva contained neutrophils. Across collections, cell type proportions in hair remained constant while cell quantities fluctuated in other tissues depending on the donor. To identify invasive tissue proxies, we compared noninvasive tissues to GTEx. By PCA, hair clustered with skin, buccal and urine bore similarity to a multitude of tissues, and saliva fell adjacent to whole blood. This clustering appeared to reflect shared cell type composition. We replicated GTEx eQTLs in our data, and saw hair captured cis-regulatory mechanisms present in most tissues, with skin performing highest, and urine best replicated eQTLs found in kidney cortex.

To assess common and rare disease applications, we analyzed gene expression using OpenTargets and OMIM. Hair follicles and buccal swabs showed enrichment for eczema and psoriasis genes, while urine additionally captured kidney disease genes. For Mendelian disease, clustering by median OMIM gene expression with GTEx recapitulated prior proxy observations.

Our findings support noninvasive sampling as a promising approach for scalable, cost-efficient, and disease-relevant transcriptome analyses. This may have value for large-scale cohort studies, longitudinal designs, and sampling from vulnerable populations.

Omics Technologies Posters - Wednesday

PB3067. Low-pass whole genome sequencing as a cost-effective and improved alternative to genotyping arrays

Authors:

P. Zhang, H. Ling, J. Paschall, B. Marosy, J. Gearhart, K. Doheny; Johns Hopkins Univ., Baltimore, MD

Abstract Body:

As the cost of sequencing continues to drop, low-pass whole genome sequencing (lpWGS) has become a potential alternative for SNP arrays in genome-wide association studies. The emergence of both new reference panels derived from large-scale diverse deep WGS data and more computationally efficient algorithms designed for imputation and phasing of lpWGS is now making lpWGS a viable alternative. Here we designed and analyzed an experiment to evaluate aspects of practical implementation including: library preparation methodology, impact across ancestry groups, DNA source, lpWGS coverage levels, imputation algorithm, and imputation reference. We created a validation ‘truth’ dataset by deep sequencing 92 samples with diverse ancestry background (down sampled coverage to 28X). The lpWGS ‘test’ dataset (~ 1X) was generated using a cost-effective library prep (plexWell LP384 from seqWell) for the same set of 92 samples. For cost-comparison modeling, we mimicked NovaSeq6000 run conditions of 768 samples in a 2 x 150 bp S4 flowcell. We used a recent imputation method designed for lpWGS, GLIMPSE, to perform imputation based on the 1000 Genomes (1KG) deep sequencing reference panel (NYGC). The SNP array ‘test’ dataset was *in silico* array data derived by extracting the illumina Global Screening Array (GSA, 654K variants) genotypes from the deep WGS validation dataset, and then performing imputation with the TOPMed reference panel. We evaluated both sensitivity and concordance of each method by comparing all imputed sites that exist in the validation dataset, as well as sites passed different QC filters (R-square from TOPMed imputation, INFO and GP from GLIMPSE imputation). When compared to the validation dataset for autosomal chromosomes for the 92 samples, our results showed that lpWGS recalled more non-reference variants than array (average sensitivity, 89.7% vs 84.1%). After applying filters, lpWGS has improved concordance over array and the gain is more prominent for non-European ancestry, the average precision for non-reference calls are 0.982 vs 0.922 for African, 0.981 vs 0.915 for Asian, and 0.982 vs 0.924 for European. When binned by minor allele frequency, we observed more gain for less common variants (e.g. MAF <10%). In addition, we were able to use lpWGS data to perform the necessary array-based quality checks of contamination (for 5% mixture or higher), genetic ethnicity by principal component analysis (PCA), and relatedness using identical by descent (IBD). Our study demonstrates the feasibility of using lpWGS as a cost-effective and more population agnostic alternative for array-based approaches using publicly accessible resources.

Omics Technologies Posters - Thursday

PB3068. Machine learning approach embedding an alternative splicing as a strategy for a multi omics integration

Authors:

Y. Lee¹, S. Han², J. Lee³, h. aycheh⁴; ¹Seoul Natl. Univ., Seoul, Korea, Republic of, ²Univ. of Utah, Salt Lake City, UT, ³Yonsei Univ., Seoul, Korea, Republic of, ⁴Univ. of Utah Asia campus, Incheon, Korea, Republic of

Abstract Body:

While individual analysis of omics datasets is valuable for identifying omic-phenotype associations, analyses using only one data type are not sufficient to fully elucidate complex diseases because such diseases are the end point of events cumulating with multiple variations through multi-omics biology. To better understand the genetic architecture of complex diseases, a relevant strategy for integrating multi-omics data (i.e. genomics and transcriptomics) is needed. In this study, we seek to develop a machine learning that can integrate genetic variants and expression based on a molecular mechanism - alternative splicing (AS). AS acts as a bridging mechanism connecting genetics to phenotype, passing genetic information encoded in DNA sequences into protein via mRNA sequences. Most of all AS is a key regulatory mechanism of gene expression. We developed a machine learning model that encompass a genetic regulatory elements of gene expression as feature variables sch as exon length, intro length, splice site, distance between genetic variants and exons, splicing regulatory elements, and splicing factor that bridge a exon-level expression and genetic variant. Among these features, intron length, splicing regulatory elements, and the distance were the highest contribution in the model. This study will summarize an evaluation of prediction power of various supervised machine learning classifier and will provide a feasibility of multi-omics integration strategy that incorporate a molecular mechanism into machine learning approach.

Omics Technologies Posters - Wednesday

PB3069. Major cell-types in imbalanced multiomic single-nuclei datasets impact statistical modeling of links between regulatory sequences & genes.

Authors:

F. Leblanc^{1,2}, G. Lettre^{1,2}; ¹Université de Montréal, Montreal, QC, Canada, ²Montreal Heart Inst., Montreal, QC, Canada

Abstract Body:

Introduction - Most variants identified by genome-wide association studies (GWAS) are located in non-coding regions of the genome without clear indications of their function. Enhancer activity disruption is presumed to be the underlying mechanism for most of these GWAS variants. Epigenomic profiling, including ATACseq, is one of the main tools used to define enhancers. Because enhancers are overwhelmingly cell-type specific, inference of their activity is greatly limited in complex tissues that include multiple cell types. Newly developed multiomic assays, probing both open chromatin and gene expression within the same single nuclei, enables the study of the correlations (links) between these modalities at single cell resolution. Current best practices to infer the regulatory effect of cis-regulatory elements (cCREs) involve removing biases associated with ATACseq peak coverage and GC content. As proposed by Ma et al. 2020, a null distribution of Gene-Peak correlations is built using ATACseq peaks of matching coverage and GC content, drawn from chromosomes excluding the one hosting the tested gene. The resulting distribution of Pearson correlation values is then scaled, providing Z-scores for the cCRE and each matched trans-links. This is done under the assumption that these trans-ATACseq peaks should not have a regulatory effect on the tested gene (i.e. they are independent). This strategy has been broadly adopted by popular single nuclei multiomic workflows such as Signac. **Results** - Here we uncovered limitations and confounders of this approach using the publicly available 10X PBMC multiomic dataset. We found a strong loss of power to detect a regulatory effect for cCREs with high read counts in the dominant cell-type. We showed that this is largely due to cell-type specific trans-ATACseq peaks correlations creating bimodal null distributions. Further, contrary to the expectation of increased power with higher cell counts, we showed that down-sampling the dominant cell-type generally increases Z-scores for gene-peak links specific to the dominant cell-type. We tested alternative models and concluded that the simplest (i.e. using directly the correlation coefficient without building a null distribution) has the strongest predictive value for the regulatory effect of a cCRE when compared to Epimap predictions and CRISPR validation data.

Omics Technologies Posters - Thursday

PB3070*. Making population-scale toxicogenomic analysis performant and cost-effective using REVEAL Biobank.

Authors:

S. Sarangi, U. Mudgal, M. Colosimo, M. Peterson, Z. Pitluk; Paradigm4, Waltham, MA

Abstract Body:

The scale and complexity of human beings' exposure to environmental toxicity has a major impact on public health across the world. Clinical trials, whether for investigative purposes or drug development, are complicated by the variable exposure of participants to environmental hazards, including their historical medicines usage. Understanding the interplay between genomic loci, environmental hazards and medicines may allow the selection of patients who are more likely to benefit from interventions. Additionally, validating the influence of toxins on diseases and disorders should help raise awareness and mitigate our footprint on the environment.

Performing large toxicogenomic studies using data sources such as the UK Biobank (UKBB) and the Comparative Toxicogenomic Database (CTD) can often be challenging due to extensive computational resources, high cost of analyses, and building complex pipelines. REVEAL Biobank offers a scalable, highly-performant, and cost-effective analytics solution that is easy to deploy and use. In the recently concluded MCBIOS 2022 conference, we had presented results of a toxicogenomic analysis performed using REVEAL Biobank, UKBB (Application ID: 51518), and the CTD. In that work, we calculated associations between 10,138 variants across 71 genes common among chemical-gene interactions for 5 chemicals (NO₂, NO_x, SO₂, Pb, and Asbestos), and 779 phenotypes (across cardiovascular, metabolic disorders, among others). The cost of this analysis was \$12.5 for ~10 million logistic regressions. In this study, we extended our previous work from two perspectives: 1) performance & scalability; 2) improve understanding of chemical-gene interactions. We demonstrated scalability and performance of REVEAL Biobank in computing a GWAS by creating a design of experiments space to measure time and cost contingent on three inputs - number of covariates, number of mutations, and number of phenotypes. The ranges used for the inputs were - covariates - 10-30; mutations - 10-100 thousand; phenotypes - 500 - 2500. To improve understanding of the influence exerted by toxins on disease-relevant genes, we conducted a Polygenic Risk Score (PRS) computation on the top significant associations identified from the GWAS and linkage disequilibrium (LD) analyses. Population-scale toxicogenomic studies using tools such as GWAS, LD, and PRS are imperative to build network maps and elucidate the complex mechanisms of chemical-gene interactions and their impact on public health.

Omics Technologies Posters - Thursday

PB3071. Massively parallel reporter perturbation assays uncover temporal regulatory architecture during neural differentiation

Authors:

A. Kreimer; Rutgers Univ., Piscataway, NJ

Abstract Body:

Gene regulatory elements play a key role in orchestrating gene expression during cellular differentiation, but what determines their function over time remains largely unknown. Here, we perform perturbation-based massively parallel reporter assays at seven early time points of neural differentiation to systematically characterize how regulatory elements and motifs within them guide cellular differentiation. By perturbing over 2,000 putative DNA binding motifs in active regulatory regions, we delineate four categories of functional elements, and observe that activity direction is mostly determined by the sequence itself, while the magnitude of effect depends on the cellular environment. We also find that fine-tuning transcription rates is often achieved by a combined activity of adjacent activating and repressing elements. Our work provides a blueprint for the sequence components needed to induce different transcriptional patterns in general and specifically during neural differentiation.

Omics Technologies Posters - Wednesday

PB3072. Max Read™: a novel high-throughput sequencing format for patterned flow cells

Authors:

N. van Wietmarschen, B. Brooks, L. LaMarca, D. Witters, R. Shultzaberger, M. Fabani, E. Glezer; Singular Genomics Systems Inc, San Diego, CA

Abstract Body:

Title Max Read™: a novel high-throughput sequencing format for patterned flow cells. **Background** Short read next generation sequencing (NGS) has advanced a multitude of applications, including RNA quantification, DNA fragment detection and counting, barcode sequencing, single-cell analyses, and proteomics. Typically, short-read NGS is performed by sequencing one round of single-end or paired-end reads on an array of DNA clusters. Here, we introduce the Max Read™ kit, allowing multiple sets of independent single-end or paired-end reads on the same patterned flow cell, enabling higher output of short reads. **Methods** A set of 8 poly-A-enriched RNA-seq libraries was prepared using the NEBNext Ultra II RNA Library Prep Kit (NEB) and the NEBNext Poly(A) mRNA Magnetic Isolation Module. For each library, 1 µg of input universal human reference (UHR), human brain (HBR), human kidney (KT), or human skeletal muscle (SMT) RNA (Invitrogen) was used. Each RNA-seq library contains distinct sequencing primer regions as well as universal primer sequences for clustering on Singular Genomics flow cells. Libraries were pooled and clustered simultaneously in one F2 flow cell. Library molecules went through clonal amplification on the patterned flow cell, allowing multiple templates to amplify within the same nanowell. The flow cell was then subjected to eight sequential rounds of primer binding and 50bp single-end sequencing on the G4™ Sequencing Platform, until all libraries were sequenced. **Results** The Max Read™ configuration with 8 independent sets of 50 bp single-end reads resulted in an average of 32 million reads per library, with an average of 84% of bases \geq Q30. This corresponds to >250 million reads per individually addressable lane, or >1 billion reads per F2 flow cell using the G4™ sequencer. Gene expression analysis demonstrates strong correlations between the Max Read™ format and NextSeq™ 2000 results using the same RNA input, while gene body coverage uniformity of the G4 Max Read™ and NextSeq™ 2000 results show similar mean distributions of coverage depth across the length of mapped transcripts. **Conclusions** The Max Read™ format enables up to eight independent single-end 50 bp reads on the same patterned flow cell by leveraging multiple unique sequencing primer binding sites on distinct library molecules. This study demonstrates the ability of achieving >1 billion 50 bp RNA-Seq reads with an average of 84% of base calls \geq Q30. The G4™ sequencer, which can run 4 flow cells in parallel, can produce up to 4 billion short reads using the Max Read™ format, providing a new, powerful method for high-throughput short read sequencing.

Omics Technologies Posters - Thursday

PB3073. Metabolomic analysis of vitreous humor from the eyes with different uveal melanoma prognostic subtypes defined by tumor gene expression profiling.

Authors:

F. Demirci¹, L. Tang¹, B. Beser², M. Kachman², H. Demirci²; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Uveal melanoma (UM) is the most common primary intraocular cancer observed in adults and exhibits distinctive genetic/molecular features. UM is an aggressive cancer, and despite advances in primary tumor management, survival rate remains low due to high rate of metastasis. A prognostic test currently available for UM involves the gene expression profiling (GEP) of tumor samples, often obtained by fine-needle aspiration biopsy (FNAB) prior to eye-preserving brachytherapy. GEP identifies two major prognostic UM subtypes (class 1 with low and class 2 with high metastatic risk). UM primarily arises from the choroid at the posterior eye segment. Posterior eye cavity is filled with vitreous, composition of which is influenced by microenvironment of posterior eye layers. Molecular analysis of vitreous may therefore provide important insights into the pathogenesis of posterior eye diseases and inform future management strategies. Changes in tumor behavior are believed to correlate with metabolic alterations upstream/downstream of altered gene expression and protein function. Since the metabolome can provide a snapshot of all biochemical processes occurring in a biological system, metabolomic profiling of vitreous may advance our knowledge of UM by providing an overall readout of tumor state/behavior. To examine whether metastasis-prone UMs develop distinct metabolic features that can be captured by vitreous profiling, we performed untargeted metabolomic analysis of vitreous samples obtained from UM-affected eyes during tumor FNAB procedure for GEP. We initially focused on ~80 samples analyzed together (both in positive and negative ion mode) on the same platform (Agilent 1290 Infinity II / 6545 Q-TOF LC/MS system with the JetStream Ionization (ESI) source). Data analysis for this platform follows a hybrid targeted/non-targeted approach using Profinder and Agilent's Find by Molecular Feature algorithm. A combined feature set was generated, followed by data QC/filtering, transformation/imputation, and association testing by controlling for FDR. While further analyses using more complex modeling frameworks are still underway to evaluate the prognostic value of vitreous profiling, our initial testing of individual metabolite associations with UM prognostic subtypes using ORIOGEN software has revealed multiple vitreous metabolites (most being unknown compounds) significantly differing between the eyes with GEP class 1 vs 2 tumors. Our results suggest the presence of vitreous metabolites that may differentiate tumor GEP-based UM subtypes and function as biofluid-based markers following further characterization.

Omics Technologies Posters - Wednesday

PB3074. Metagenomic analyses coupled with metabolic and deep immune profiling reveal coordinate effects on host-microbe interactions in chronic kidney disease

Authors:

S-C. Su¹, I-W. Wu¹, L-C. Chang²; ¹Chang Gung Mem. Hosp., Keelung, Taiwan, ²Florida Atlantic Univ., Boca Raton, FL

Abstract Body:

Perturbation of gut dysbiosis has been linked to chronic kidney disease (CKD), a gradual loss of renal function that was pathologically correlated with a sophisticated milieu of metabolic and immune dysregulation. However, the underlying host-microbe interaction is poorly understood. To address this, we performed multi-omics measurements, including systems-level gut microbiome, targeted serum metabolome, and high-dimensional immunotyping, in a cohort of 72 patients and 20 non-CKD controls. Our analyses on functional profiles of gut microbiome showed that loss of renal function decreased the diversity and abundance of carbohydrate-active enzyme (CAZyme) genes, whereas kidney failure increased the abundance of nitrogen cycling enzyme, virulence factor, and antibiotic resistance genes. Moreover, use of fecal metagenomic, serum metabolomic, and immune signatures resulted in distinct effects on differentiating mild and severe CKD from controls. Models generated using measurements of circulating metabolites (amino acids, bile acids, and short-chain fatty acids) or immunotypes were predictive of renal impairment but less so than many of the taxonomic or functional profiles derived from gut microbiota, with the CAZyme genes being the top performing model to accurately predict early stage of diseases. In addition, correlation analyses among systems-level microbiome, serum metabolome, and immune parameters revealed coordinated host-microbe relationships in CKD. Specifically, the highest fractions of significant correlations were identified with circulating bile acids by several taxonomic and functional profiles of gut microbiome, while immunotype features were only moderately associated with the abundance of microbiome-encoded metabolic pathways and serum levels of amino acids. Overall, our multi-omic integration revealed several signatures of systems-level gut microbiome in robust associations with host-microbe co-metabolites and renal function, which may be of etiological and diagnostic implications in CKD.

Omics Technologies Posters - Thursday

PB3075. Methods for screening candidate causal regulatory variants in primary immune cells by CRISPR ribonucleoprotein.

Authors:

M. Lorenzini^{1,2}, K. Sajeev^{1,2}, G. McVicker¹; ¹Salk Inst. for Biological Studies, La Jolla, CA, ²Univ. of California San Diego, La Jolla, CA

Abstract Body:

Non-coding variants that are causal for human traits, such as those implicated by GWAS, often have small effects on gene expression. Current methods for determining the function of candidate causal variants sacrifice statistical power and genomic context for screening throughput, resulting in the loss of biologically important information. To screen non-coding candidate sequences with power and versatility in situ, we have optimized a robust ribonucleoprotein method for CRISPR deletion in resting primary immune cells. We apply this method to determine candidate causal variants of type 1 diabetes at the *CTLA4* locus. Finally, we extend this method for direct capture of genomic edits to enable powerful, combinatorial Perturb-seq screening of regulatory sequences.

Omics Technologies Posters - Wednesday

PB3076. MHConstruct creates personalized full-length MHC locus haplotype sequences by finding optimal paths through a population variation graph.

Authors:

M. Mumphrey, M. Cieslik; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

BACKGROUND: The immune system is able to identify and eliminate malignant cells via presentation of neoantigens by the major histocompatibility complex (MHC). Loss of MHC function is a known mechanism allowing cancers to escape this immunosurveillance, however molecular characterization of MHC loss is hindered by the extreme polymorphism of the region. Classical genetic, transcriptomic, and epigenetic methods rely on the alignment of molecular sequences to a standard reference, which fails for the MHC locus where the average person's genome is highly divergent from the standard reference. While methods exist for personalized analysis of individual MHC genes, there is a lack of tools that allow for personalized analysis of the entire 5 Mb MHC locus, which contains many accessory genes and intergenic regulatory regions. Population variation graphs have been used to allow for DNA-seq analysis of large polymorphic regions, however there are few algorithms that allow for the analysis of other data types directly on population graphs.

METHODS: We have developed MHConstruct, a tool that allows for reconstruction of the linear diploid MHC locus on an individual basis. MHConstruct leverages the VG toolkit to create a population variation graph of the MHC locus using a database of 95 full length MHC locus sequences. A SNP array is then constructed containing the genotype for each SNP in each reference sequence used to construct the graph. To generate a personalized linear diploid reference from the graph, DNA-seq reads from an individual are aligned to the graph and all SNPs are genotyped. Genotypes are phased, and the viterbi algorithm is used to find the optimal path through the SNP array for each set of phased genotypes. The viterbi paths are then used to pull segments of the corresponding reference sequences to construct the optimal linear sequence given the available set of reference sequences.

RESULTS: MHConstruct was able to construct diploid reference sequences for the MHC locus that had a rate of variation less than ~1 SNP/kb, which is the expected level of variation across the standard reference genome in non-polymorphic regions. Phasing and switch error rates were further characterized using linked-read sequencing. Finally, we demonstrate the use of MHConstruct references with standard tools for bisulfite sequencing and CHIP-seq data.

CONCLUSIONS: MHConstruct creates personalized diploid reference sequences for the MHC locus that are as high quality as the standard reference is for other less-polymorphic regions of the genome. This allows for personalized molecular analyses of the entire MHC locus using standard approaches that require a linear reference.

Omics Technologies Posters - Thursday

PB3077. MiRNA regulates potential genetic pathways involved in diabetic retinopathy: Implication for an early disease risk prediction

Authors:

S. Vishwakarma¹, P. Goel¹, R. Papparu¹, M. Tyagi¹, R. Narayanan¹, S. Chakrabarti¹, I. Kaur²; ¹LVPEI, Hyderabad, India, ²Prof Brien Holden Eye Res. Ctr., L V Prasad Eye Inst., Hyderabad, India

Abstract Body:

Background: Diabetic Retinopathy (DR) is a common microvascular retinal disease and can be classified into non-proliferative (NPDR) and proliferative diabetic retinopathy (PDR). Hypoxia and hyperglycemia along with dysregulated genetic signaling and altered metabolic pathways, affect the microvasculature of retina in DR leading to neuroinflammation, neurodegeneration, and neovascularization. However, neurodegeneration as an early event in DR development has not been explored much. Herein, we assessed the regulation of genes involved in both early diabetic retinal change and late stages of DR for their role in early risk prediction. **Methodology:** Cadaveric human retinal tissues with and without a history of diabetes (n=15) were collected from Ramayamma International Eye bank, INDIA. Retinal tissues were stained with H&E and PAS stain for detecting early diabetes-induced microvascular changes in the retina. Blood samples were collected from NPDR, PDR, and age-matched control patients with prior informed consent. RNA was isolated from tissue and blood using Purelink Kit. Quality and quantity assessment was done by bioanalyzer and nanodrop respectively. Tissues and blood RNA samples (RIN \geq 7) proceeded for mRNA and miRNA global profiling respectively on the Agilent platform. mRNA data were compared with published data (GSE160130). MiRNA targets were predicted by *in-silico* tools. qRT-PCR (SYBR green) was done for validation. **Result:** HandE and PAS staining showed ganglion cell layer thinning and significant increase in the number of blood vessels and their thickness in diabetic retina compared to control. A total of 1217 differentially expressed genes including ncRNA were found in diabetic retina compared to control. Their comparison with published data of NPDR and PDR retina tissue, identified 702 mRNA, 62.04% mRNA were present in DM, NPDR and PDR. Approximately 26.7% of these genes were found significantly up-regulated in DM, NPDR, and PDR and are involved in apoptosis, chemokine signaling, neurodegeneration, AGE-RAGE signaling, Rho-GTPases pathway, cell cycle, and MAPK signaling, etc. Next, we identified regulatory miRNA from our data including hsa-miR-142-3p targets genes involved in cell migration, proliferation, and hematopoiesis, hsa-miR-19a-3p, and hsa-miR-18a-5p that regulate angiogenesis via endothelial cell regulation, apoptosis, and neurodegeneration. Besides these, many other miRNAs were found to regulate the activation of pathological pathways in early DR that needs further functional validations. **Conclusion:** This study underscores the potential role of miRNA in regulating neurodegeneration and early DR risk prediction.

Omics Technologies Posters - Wednesday

PB3078. Mitigating challenges of large-scale single cell data management, querying, and analysis with REVEAL SingleCell

Authors:

K. Sharma, C. Bragdon, U. Mudgal, Z. Pitluk, S. Sarangi; Paradigm4, Waltham, MA

Abstract Body:

Advancements in single cell technologies have allowed sustained exponential growth in the rate of published studies and datasets over the last decade. Individual studies have also increased in scale over time - during the first half of 2020, approximately 1,400,000 cells were added to the pool of public data every month - and this number continues to grow. New 'atlases' that try and standardize data across tissues are being generated frequently, and large multi-organ datasets from multiple patients such as Tabula Sapiens are being released as an effort to get closer to doing population-scale analyses. Complicating the growth in data generation is the proliferation of algorithms to help clean, integrate, harmonize, cluster, and analyze single cell datasets. While exciting and vital for addressing problems such as identifying sub-populations of cancer cells, these advances have created unique challenges. These include efficient storage and retrieval due to variable and inherent sparsity of data, rapid querying across hundreds of samples for co-expression of genes, running multiple algorithms at scale in one workflow due to the variability in results (e.g., for clustering), and handling multi-omics data within one platform. We had previously described REVEAL SingleCell, a computational platform for working with multi-omics single cell data (*BMC Genomics*, 2021). Here, we present new and improved features of REVEAL SingleCell developed to address the above challenges. We demonstrate scalability and performance by presenting results from two design of experiments spaces: 1) measuring time to load data using two inputs: % sparsity (1% - 100%) and number of cells (10,000 - 1 million); 2) measuring time for queries (e.g., co-expression of genes) using three inputs: % sparsity (1% - 100%), number of cells (5 - 100 million), and number of genes (1 - 300). We also present benchmarking performance numbers for cross-sample queries using metadata fields (e.g., cell type = epithelial), and computations (e.g, correlation between genes across thousands of cells, calculation of z-scores at scale). Finally, we show the use of iSets to enable the creation of custom atlases with samples spanning many projects that facilitate data reusability. Taken together, we provide useful benchmarks for software developers as the scale of single cell datasets increases.

Omics Technologies Posters - Wednesday

PB3079. Mitochondrial Genomic ratio (mitoscore) is an efficient biomarker to predict the implantation efficacy of human embryos

Authors:

S. Chettiar¹, J. Dattani², A. Patel³, D. Jhala⁴, F. Parikh⁵, R. Uppala⁶; ¹Gujarat Univ., Ahmedabad, India, ²Ministry of Hlth.and Family Welfare, Ahmedabad, India, ³GeneXplore Diagnostics and Res. Ctr. Pvt. Ltd., Ahmedabad, India, ⁴Sch. of Sci., Gujarat Univ., Ahmedabad, India, ⁵Jaslok Hosp., Mumbai, India, ⁶Genexplore Diagnostics and Res. Ctr., Ahmedabad, India

Abstract Body:

Objective: Single embryo transfer is gaining momentum in the clinical field of assisted reproductive technology (ART). However, selecting a single embryo for transfer is difficult as there is no reliable biomarker to predict the success of implantation and development. Here we discuss mitoscore as predictive markers by genomic rather than, visual, and metabolomic markers. This score is calculated using a ratio of the whole to mitochondrial genome, identified using next generation sequencing (NGS). In turn, this data educates an AI-based algorithm for the selection of embryos for enhanced implantation success through large dataset. **Design:** Couples undergoing infertility treatment and preimplantation genetic testing (*PGT*) were enrolled for the study with written consent. All grade I and II embryos were biopsied and subjected to NGS. The chromosome aneuploidy data and nuclear to mitochondrial genome ratio were analyzed using R-based machine learning and deep learning. To investigate the predictive power of the mitoscore and its correlation with clinical outcomes, observed data and predicted data were compared and statistically analyzed. **Material and methods:** Trophoctoderm biopsy leading to 3-6 cell source was subjected to NGS for 3000 embryos. The data of these 3000 embryos were processed via deep learning and machine learning algorithms like random forest, support vector machine, general linear model, and linear discriminant analysis. An additional neuronal network scope was added to the pipeline, giving a stimulus enhancement to the dataset and thus finetuning the data outcome. Opensource software package like the h2o R package was employed to perform R analysis and Box-Behnken design (BBD) using response surface methodology (RSM). **Results:** NGS data revealed the euploid status of the embryo and the ratio of the whole genome sequence to the mitochondrial genome. Ratios were plotted on a know graph ratio and values were converted to percentages. Embryos with more than 40% and less than 25% ratio were found to either have aneuploidy or were not selected for transfer. R^2 analysis also predicted with a confidence interval of 0.9921 that mitoscores above 37 and less than 28 were predictive of aneuploidy in any one/multiple chromosomes. With increasing dataset size, we intend to improve the sensitivity and specificity of the technology adapted with a positive prediction by machine DOE and R^2 analysis with superior confidence. **Conclusions:** From the experimental data, R package prediction through deep learning, and machine learning, it can be concluded that mitoscore values can be reliably employed as a biomarker for implantation efficacy of embryo.

Omics Technologies Posters - Thursday

PB3080. More bang for your buck! Universal Sequence Detection: A novel full spectrum variant genotyper for NGS data

Authors:

J. Van Den Akker, K. Wong, A. Rostamianfar, Z. Langley, M. Cusack, A. Ondov, A. Zimmer, A. Zhou, S. Topper; Color Hlth., Burlingame, CA

Abstract Body:

Reduced cost and interest in novel variant detection has led the field of clinical genetic testing to adopt NGS technologies over once widely-used genotyping arrays. However, some advantageous aspects of genotyping can now be applied to NGS technologies to improve known variant detection. Here, we describe a novel algorithm we developed, called Unique Sequence Detection (USD), which improves the sensitivity and efficiency of known variant detection in short-read sequencing data.

USD is a genotyping-based variant caller that utilizes a library of predefined unique sequences to accurately, efficiently, and repeatably genotype known variants from short-reads. USD is able to genotype the entire spectrum of variant types, from single nucleotide variants (SNVs) to complex structural variants (SVs).

USD operates in two phases: 1) a one-time unique sequence generation step where unique sequences are constructed based on previously well-characterized variants and 2), a pattern matching step where an alignment file is scanned to identify reads supporting the constructed unique sequences. Each sequence has three parts: a “core” sequence that spans the variant and needs to be matched exactly and left and right “flanking” sequences where some variation is tolerated to provide resilience against co-variants and sequencing errors.

Leveraging our existing library of clinically reported pathogenic variants, we found that USD can genotype challenging SNVs, small insertions and deletions (indels), and SVs that are often missed, called incorrectly, or require a special implementation of variant callers. These include known challenging variants with high GC content or near homopolymers, including *MSH2* c.942+3A>T. Our USD library also contains unique sequences to identify known breakpoints for complex SVs, including the hard-to-call Boland inversion (*MSH2* inversion exons 1-7). In a set of 1531 unique variants including SNVs, indels, MNVs, and complex SVs, USD achieved a sensitivity of 99.61% with a specificity >99.9%.

In conclusion, USD is a comprehensive genotyping algorithm that enables accurate detection of known variants of all sizes and complexities. It boosts detection of known pathogenic variants using either targeted or whole genome short-read sequencing, thereby reducing the need for manual review and special pipeline implementations for known challenging variant calls as well as improving overall pipeline efficiency. Importantly, subsequent observations of validated variants can be added regularly to the USD library, thereby continuously improving the overall sensitivity and efficiency of the clinical genetic test.

Omics Technologies Posters - Wednesday

PB3081. Multi-Omics Integration for Osteoporosis Prediction through Mixture-of-Experts Framework

Authors:

Y. Gong¹, J. Greenbaum¹, L. Jiang², A. Liu³, X. Zhang¹, H. Shen⁴, H-W. Deng¹; ¹Tulane Univ., New Orleans, LA, ²Tulane Univ. Sch. of Med., New Orleans, LA, ³Tulane Univ., Metairie, LA, ⁴Tulane Univ, New Orleans, LA

Abstract Body:

Abstract With the development of high-throughput omics measurement technologies, it is essential for researchers to integrate heterogeneous genomics and molecular information from multiple omics to develop more accurate disease-relevant prediction models. We will apply a mixture-of-experts deep generative model to integrate multi-omics profiles for normal/osteoporosis classification and novel osteoporosis-related biomarkers identification. We will use the miRNA sequencing, DNA methylation, and gene expression data in peripheral blood monocytes (PBMS) from Caucasian (n=558) and African American (n=359) men sampled from the Louisiana Osteoporosis Study (LOS). Subjects will be classified into normal/osteopenia groups based on their hip BMD (T-score <-1.0). Mixture-of-experts deep generative framework, a novel deep generative model-based framework with better performance than other conventional methods for dimension reduction, will be used for multi-omics profiles integration and osteoporosis diagnosis. Through this method, variational autoencoder (VAE) will be used to learn high-level representations in each omics data (gene expression, miRNA expression, and DNA methylation) and we will integrate all learned representations through the mixture of experts, which weight the learned representations to form the integrated ones. The integrated representations from VAE and phenotypes of the samples will be used to train the flexible neural forest model, which is a multi-layer feed-forward neural network for classification. We will then apply 5-fold cross-validation to assess the performance of the model. To reveal the potential biomarkers related to osteoporosis, we will select the features with the highest weight in the VAE framework. Gene Ontology (GO) analysis will then be used to identify the biological functions and related pathways of these features. Through the mixture-of-experts deep generative model, we will learn the representations of the complex multi-omics datasets and construct a classification model for osteoporosis prediction based on the learned representations. Further, based on this model, we will identify the significant biomarkers related to osteoporosis, which will provide novel insights into the pathological mechanisms of the disease. **Keywords:** osteoporosis, bone mineral density, deep learning, disease prediction

Omics Technologies Posters - Thursday

PB3082. Multi-omics integration via similarity network fusion to detect subtypes of aging

Authors:

M. Yang¹, S. Matan-Lithwick², Y. Wang³, P. De Jager⁴, D. Bennett⁵, D. Felsky⁶; ¹Univ. of Toronto and CAMH, Toronto, ON, Canada, ²Krembil Ctr. for Neuroinformatics, Ctr. for Addiction and Mental Hlth., York, ON, Canada, ³Rush Univ. Med. Ctr., Chicago, IL, ⁴Columbia Univ Med Ctr, New York, NY, ⁵Dept. of Neurological Sci., Rush Med. Coll., Chicago, IL, ⁶The Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada

Abstract Body:

Background: Molecular subtyping of brain tissue provides insights into the heterogeneity of common neurodegenerative conditions, such as Alzheimer's disease (AD). However, existing subtyping studies have mostly focused on single data modalities, such as RNA sequencing (RNAseq), which provide incomplete neurobiological information on pathological processes. To remedy this, we applied similarity Network Fusion (SNF), a method capable of integrating multiple high-dimensional multi-'omics data modalities simultaneously. **Methods:** We analyzed human frontal cortex brain tissue samples characterized by five 'omic modalities: bulk RNAseq (18,629 genes), DNA methylation (53,932 cpg sites), histone H3K9 acetylation (26,384 peaks), tandem mass tag proteomics (7,737 proteins), and metabolomics (654 metabolites). SNF followed by spectral clustering was used for subtype detection, with subtype numbers determined by eigen-gaps and dip-test statistics. Normalized Mutual Information (NMI) was calculated to determine the contribution of each modality and feature to the fused network. Resulting subtypes were characterized by associations with 12 age-related neuropathologies and cognitive performance. **Results:** Fusion of all five data modalities (overlapping n=111) yielded four molecular subtypes (nS1=32, nS2=26, nS3=31, nS4=22); S1 exhibited lower episodic memory performance than other subtypes proximal to death ($t=4.7$, $p=1.5 \times 10^{-4}$). Histone acetylation (NMI=0.32) and RNAseq (NMI=0.16) contributed most strongly to this fused network; the top individual features were an acetylation peak in the promoter of CD200 and RNA abundance of PARP4. Secondary analysis fusing only RNAseq and histone acetylation (n=520) yielded five subtypes which were correlated with the fully integrated subtypes (Fisher's $p=5.0 \times 10^{-4}$) and strongly associated with AD neuropathology and episodic and semantic memory. Sensitivity analyses of all modality combinations and sample subsets found substantial influences of sample size and subtype number, but reinforced the importance of histone acetylation, RNAseq, and DNA methylation. **Conclusion:** We identified highly integrative molecular subtypes of aging derived from up to five multi-'omics data modalities simultaneously. These subtypes recapitulate some features of previous subtyping work in AD using single modalities, but also provide new molecular targets and shed light on the benefits and challenges of multi-omic integration and individual subtyping in this field.

Omics Technologies Posters - Thursday

PB3083*. Multiset correlation and factor analysis enables exploration of multi-omic data

Authors:

B. Brown¹, C. Wang², S. Kasela¹, F. Aguet³, D. Knowles¹, T. Lappalainen⁴; ¹New York Genome Ctr., New York, NY, ²Columbia Univ., New York, NY, ³Broad Inst., Cambridge, MA, ⁴KTH Royal Inst. of Technology & NY Genome Cntr, New York, NY

Abstract Body:

Multi-omics data promise to revolutionize biomedical research, providing insight into the coordination of molecular processes that cannot be gleaned from any lone 'omics mode. However, joint inference methods are lacking in either the number or type of modes that can be used, or in flexibility and efficiency. We developed Multi-set Correlation and Factor Analysis (MCFA), an unsupervised data integration method that enables fast inference of shared and private factors in multimodal data. MCFA is designed to overcome challenges that are common with 'omics data such as the large number of features relative to the sample size, disparate data types, and the unknown contributions of mode-specific technical factors. Moreover, MCFA requires no user-specified tuning parameters, is efficient enough to apply to thousands of samples, and has a strong connection to canonical correlation analysis. We have applied MCFA to 614 ancestry-diverse individuals, each profiled with 5 different genomics technologies as part of the TOPMED-MESA multi-omic pilot. We integrated RNA, methylation, protein, and metabolite data using MCFA, and further integrated genetic data by analyzing genetic associations with the inferred factors. We found extensive sharing across data modes: 30% of the total variance in protein expression is explained by the shared space; 17% for RNA expression and metabolites; and 8% for methylation. Samples cluster strongly by race and sex in the shared space, even though this is not true for the top PCs of individual modes and the shared space was inferred without genetic or sex chromosome features. Next, we evaluated the total phenotypic variance explained by each of our inferred spaces. The shared space captures 80% of the variance in race, 95% of the variance in sex, and 60% of the variance in age, while additionally capturing effects of sample cell-type composition, air quality measures, and inflammatory biomarkers. Private spaces often captured technical features such as batch but also captured variation in cell-type composition not explained by the shared space and, in the metabolite data, the month of sample collection. Our genetic analysis revealed association enrichment at known disease GWAS SNPs and *trans*-eQTLs for several factors. We further investigate one factor showing strong GWAS enrichment that is correlated with cell-type composition, inflammatory biomarkers and BMI, revealing a connection between global methylation patterns, lipid metabolism and related diseases. Our study provides a framework for further integrative analysis of multi-omic data, with inferred axes of molecular variation a promising future trait for GWAS and pheWAS studies.

Omics Technologies Posters - Wednesday

PB3084. Neuroimmune insights through single-cell transcriptomics of paired brain and blood from living human subjects

Authors:

D. Liu, J. Fullard, L. E. Liharska, E. Cheng, L. Wilkins, N. Beckmann, B. H. Kopell, P. Roussos, A. Charney; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Background: Genetic and clinical studies point to roles for the immune system in neuropsychiatric diseases, yet the mechanistic explanation for these observations remains unclear. Brain has a resident immune system that interacts with peripheral immunity, and under psychological and pathological conditions, cells moving dynamically between the central nervous system (CNS) and the peripheral circulation have been described. The exact nature of these interactions has not been characterized in the living human brain. This study aims to characterize the cellular and molecular basis of the CNS-periphery interaction, by analyzing the single-cell transcriptomes profiled on matched brain-blood specimens from living human subjects. **Methods:** The prefrontal cortex and peripheral blood were sampled simultaneously from donors during neurosurgical procedures, and were processed together on the same day for single-cell RNA sequencing. Following quality control, 105,833 immune cells from 13 brain-blood pairs were analyzed. Cell populations in the brain and the blood were compared with respect to cell type composition and cell transcriptional states, and were correlated for gene expression across the 13 pairs. **Results:** Both resident (microglia, macrophages) and peripheral immune cells (PICs, including T, B, natural killer (NK) cells, monocytes) were detected in brain samples. PICs detected in the two tissues displayed different cell type proportions, and the most notable difference was seen for the relative abundance of T cells and their subpopulations. Comparing gene expressions in brain PICs and their blood counterparts, brain PICs showed an up-regulation in pathways including immune signaling, stress response, apoptosis, and metabolism, whereas blood PICs expressed higher level of ribosomal genes and genes involved in defense response to antigens. Clustering brain and blood cells together identified transcriptional states that are broadly shared between tissues, with an exception of a brain-specific CD8⁺ T population. Across 13 brain-blood pairs, the cell types that showed the highest average gene-wise correlation with microglia were blood NK and CD4⁺ T cells. Data on repeatedly sampled donors showed that gene expression changes over time were only modestly correlated between tissues and were on average greater in brain as compared to blood. **Conclusions:** There are widespread differences and correlations between the CNS and the peripheral immunity at the cellular and molecular level. These insights highlight the potential to extract molecular information about the brain via a routine blood draw.

Omics Technologies Posters - Thursday

PB3085. New generative deep models to discover novel disease-gene associations in large-scale genomic cohorts

Authors:

R. Orenbuch, J. Frazer, M. Dias, K. Brock, D. Marks; Harvard Med. Sch., Boston, MA

Abstract Body:

Accurate and unbiased variant effect prediction can increase power to discover true gene-disease connections in both rare and common variant association studies, by weeding out mutations that are unlikely to cause disease. However, the majority of variants have no known consequence, as the rate of discovery of novel variants is greatly outpacing experimental and clinical validation of the contribution of individual variants to disease phenotypes. Building on our recent success in providing an orthogonal source of evidence for benign/pathogenic classifications for human variants with EVE, we have developed a model which combines deep-variation, across millions of years of evolution, with shallow-variation, across the human population, to address this. Here, we use our scores to increase our power to discover novel gene-phenotype associations from large-scale datasets, focusing on diseases of aging and related phenotypes in the UK Biobank. We investigate if our scores increase the amount of explained heritability in existing burden-testing methods and develop new methods that leverage the continuous nature of our scores and their level of uncertainty to increase our power to detect gene-disease associations with varying levels of penetrance. We both recapitulate known and discover novel genotype-phenotype associations and ultimately fill in the gap of unexplained heritability in these diseases.

Omics Technologies Posters - Wednesday

PB3086. New genotyping and sequencing methods for high throughput molecular blood group typing

Authors:

A. Franke¹, T. Steiert¹, H. El Abd¹, M. Wittig¹, C. Gassner²; ¹Christian-Albrechts-Univ. of Kiel, Kiel, Germany, ²Private Univ. im Fürstentum Liechtenstein, Triesen, Liechtenstein

Abstract Body:

Blood group antigens play an important role in transfusion medicine and can cause unwanted dangers for the recipient through immune reactions. For this reason, high throughput typing of donor blood is carried out in blood banks to allow for the most suitable donor-recipient-match possible. The standardized blood group antigen nomenclature is stored in the ISBT (International Society of Blood Transfusion) database in the form of alleles with their associated genetic variations. Thus, it is technically possible to determine blood group alleles/antigens, based on targeted genotyping and also targeted sequencing. Firstly, we present a SNP array-based approach that enables this process in high throughput at low per-sample costs. We validated the SNP array results using data from 57,531 antigens of an ethnically diverse reference sample set including the 1000-Genomes samples and provide the blood calls and underlying data as a resource and reference to the community. To demonstrate the utility of the blood group array in disease association studies, we genotyped a COVID-19 case-control panel to validate the previous ABO association and to test for an association with other blood group phenotypes. Secondly, we present another software tool and benchmarking study to call blood group alleles from targeted NGS or WGS data. Proof-of-principle WGS data for 500 clinically interesting samples was generated and analysed for novel blood group alleles.

Omics Technologies Posters - Thursday

PB3087. Novel Principal Component Analysis reveals rich gene expression contexts from snRNA-seq within and across cell-types.

Authors:

S. Carver¹, A. Gusev²; ¹Harvard Univ., Cambridge, MA, ²Dana Farber Cancer Inst., Longwood, MA

Abstract Body:

Background:Single cell genomics has advanced our ability to profile individual cells especially in terms of their cell type. Precisely subdividing cells by type instead of analyzing bulk sequencing data has allowed researchers to identify key cell types driving disease as well as unique cell type functionality. To move beyond canonical cell types, work has been done to identify cell states (sub-groupings of cells within cell type) or cell programs (cell subsets based on common functionality across cell types). However computational methods for identifying cell states are often not sufficiently justified and validating biologically meaningful cell states rather than states derived from technical noise is often difficult due to a limited ground truth data for novel states.

Methods:We propose novel methods based on hierarchical and conditional Principal Component Analysis (PCA) to identify biologically meaningful cell states and programs from single-cell transcriptomics and epigenomics. We benchmarked our approach against conventional methodologies such as consensus Non-negative Matrix Factorization (cNMF) and t-distributed Stochastic Neighbor Embedding (t-SNE). Method performance was evaluated using simulated data, followed by application in a multi-omic single-nucleus study from control and late-stage Alzheimer's Disease (AD) prefrontal cortex. The identified cell contexts were validated by consistency with snATAC-seq, association with AD pathology, and consistency with known cell types from larger external atlases.

Results:In simulated data, the PCA-based approaches performed best at identifying both states and programs, particularly for rare states and continuous programs. PCA could reliably identify contexts in as few as 10 cells, compared to greater than 600 cells for cNMF and t-SNE, respectively. Applied to real AD data, PCA substantially outperformed cNMF for identifying rare cell types and the two approaches were comparable for common cell types. Validating the identified cell programs using snATAC-seq from the same donors revealed >100 contexts with statistically significant biological structure ($p < 0.05$ by permutation test).

Conclusion:We demonstrate that snRNA-seq data is rich with biologically relevant gene expression contexts which are often missed by conventional approaches. We present a powerful framework for identifying and validating novel contexts using PCA and implicate new cellular processes in AD pathology. Further integration of novel contexts into disease data from Genome-Wide Association Studies is ongoing.

Omics Technologies Posters - Wednesday

PB3088*. Novel sequencing platform using an open fluidics architecture and mostly natural chemistry for cost-efficient whole genome sequencing.

Authors:

D. Lipson, O. Barad, D. Shem-Tov, F. Oberstrass, M. Pratt, G. Almogy; Ultima Genomics, Newark, CA

Abstract Body:

Advances in next generation sequencing over the last two decades have enabled cutting edge life sciences research as well as a rapidly growing set of clinical applications, ranging from carrier screening and prenatal testing to tumor profiling and early cancer detection. However, sequencing cost reduction has stalled in recent years limiting the widespread adoption of genomic-based diagnostics into standard of care.

We have recently introduced a novel sequencing platform that enables whole-genome sequencing at an initial cost of \$100 per genome (\$1/Gb), by employing a novel open fluidics architecture in combination with a mostly-natural sequencing by synthesis (mnSBS) chemistry that together enable the longer read advantages of non-terminating nucleotides and the throughput and scalability of optical endpoint scanning at a significantly lower consumable cost. Beyond the initial implementation, this novel architecture provides multiple dimensions for continuous reduction sequencing cost even further in the not-distant future.

We evaluated the performance of the new platform by generating whole-genome sequence data for the seven Genome-in-a-Bottle reference samples HG001-7 at 40X, demonstrating read length of ~300bp, good coverage uniformity across the genome, and high base quality (Q30 >85%). Variant calling was performed by an updated version of DeepVariant that was optimized to use base quality information that is unique to the system, and demonstrates 99.7% concordance on single nucleotide variants and >97% concordance on InDel variants in non-repetitive genomic regions, with further improvements expected from enzyme evolution and additional optimization of the sequencing workflow.

We additionally demonstrated the applicability of the platform for other high-throughput applications including single-cell RNA-seq, deep (~100X) WGS of cell-free DNA for residual disease monitoring, whole-genome methylation analysis and agnostic pathogen profiling. The continuous drive for sequencing affordability and ubiquity will undoubtedly lead to deeper genomic insights and further integration of sequencing assays into clinical practice.

Omics Technologies Posters - Thursday

PB3089. NTSM: Fast sample swapping detection and ancestry estimation on unprocessed heterogeneous raw sequencing data

Authors:

J. Chu^{1,2,3}, J. Rong^{2,3}, X. Feng^{2,3}, H. Li^{2,3}; ¹Boston, Boston, MA, ²Dana Farber Cancer Inst., Boston, MA, ³Harvard Med. Sch., Boston, MA

Abstract Body:

Due to human error, sample swapping in large cohort studies with heterogeneous data types (e.g. mix of Oxford Nanopore, Pacific Bioscience, Illumina data, etc) remains a common issue plaguing large-scale studies. At present, all sample swapping detection methods require costly and unnecessary (e.g. if data is only used for genome assembly) alignment, positional sorting, and indexing of the data in order to compare similarly.

Instead of alignment, the similarity between raw data samples can be determined using a set of carefully curated SNP sequences indexed into a time and memory-efficient k-mer counting algorithm. These counts can be compared via a likelihood ratio-based test very quickly, with an accuracy determined by the coverage of the data and the error rate of the data type. This first pass QC method, independent of downstream analysis, can save an immense degree of computational time and is robust enough to be used on many data types combinations.

In addition, as a crude form of variant calling, the method can also be used to determine population-level ancestry analysis quickly and generically find approximate sample ancestry via a population-based PCA without requiring extensive variant calling. Population-level PCA analysis of this type also usually requires the same data types to be used in this kind of analysis, due to the requirements of carefully tuned variant calling algorithms, which is unnecessary in our streamlined, robust, k-mer counting-based methodology. Other QC metrics, such as error rate estimation can be determined from this information, providing useful utility in downstream analysis.

Omics Technologies Posters - Wednesday

PB3090. ODER: Optimising the Definition of Expressed Regions, a publicly available R package to improve annotation of RNAseq data and investigate novel transcription using short-read RNA-sequencing data

Authors:

J. Brenton¹, E. Olagbaju¹, K. Montgomery¹, S. Sethi², L. Collado-Torres³, S. Guelfi⁴, D. Zhang¹, M. Ryten⁵; ¹Univ. Coll. London, London, United Kingdom, ²Astex Pharmaceuticals, Inc., Cambridge, United Kingdom, ³Lieber Inst. for Brain Dev., Baltimore, MD, ⁴Verge Genomics, San Francisco, CA, ⁵UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom

Abstract Body:

Incomplete annotation of the human transcriptome remains a problem facing modern genetic research with the choice of gene annotation source or version affecting variant interpretation, quantification of RNAseq expression and downstream differential gene, splicing and quantitative trait analyses. We developed Optimising the Definition of Expressed Regions (ODER), a publicly available R package (<http://bioconductor.org/packages/release/bioc/html/ODER.html>) which addresses this problem by using short-read RNAseq data to re-annotate known genes. ODER leverages the derfinder R package (<http://bioconductor.org/packages/release/bioc/html/derfinder.html>) to assess base-level coverage, identify contiguous blocks of transcription, termed expressed regions (ERs) and connect these expressed regions to known genes. ODER requires two input files per sample, a BigWig and junction file. A typical ODER pipeline consists of four functions. First, using the BigWig coverage data, ODER improves the identification of ERs by varying two parameters, the mean coverage cutoff (MCC), defined as the minimum read depth that an individual base pair requires to qualify as being expressed, and the maximum region gap (MRG), defined as the maximum number of base pairs between two ERs below which they are merged. These are varied to find the minimum difference between ERs and the user-provided “gold standard” exon definitions. Second, using junction files, ODER finds overlaps between optimised ERs and junction reads, short reads mapping to the reference annotation with a gapped alignment, to connect ERs to known genes. Third, ERs are filtered for those that overlap a single or two non-intersecting junctions. ER boundaries are then trimmed to match the exon boundaries of the overlapping junction/s. Finally, ODER generates a count matrix, which provides the mean coverage over each ER. Testing of ODER’s accuracy with small RNA-seq data sets demonstrated ERs could be defined accurately with only ten polyA-selected RNA-seq samples, increasing its value. ODER has already been applied to short-read RNA-seq data generated by the Genotype-Tissue Expression Consortium data to identify potential novel coding exons and 3’UTRs (<https://rytenlab.com/browser/app/vizER>; <https://astx.shinyapps.io/F3UTER/>). Thus, ODER provides users with an efficient means of leveraging user-generated and public short-read RNA-sequencing data for the identification of novel exons which can then be studied using targeted and more expensive long-read RNA-sequencing approaches.

Omics Technologies Posters - Thursday

PB3091. On statistical modeling of single-cell RNA-seq data accounting for technological and biological dropouts.

Authors:

I. Mukhopadhyay¹, P. Mondal²; ¹Indian Statistical Inst., Kolkata, India, ²Michigan State Univ., East Lansing, MI

Abstract Body:

Analyzing single-cell RNA-seq data is challenging due to proper mathematical modeling of biological factors that influence the observed expression levels in the data. The transcription level is a transient process affected by spatial effects, cell cycle, cell types, and other environmental factors. Moreover, a low amount of initial cDNA before amplification adds to considerable noise in the data making it difficult to analyze. A typical dataset contains a large number of zero expressions, also known as dropouts. Moreover, the bimodal pattern of the expression level adds additional complexity to analyze the data. We propose a statistical model that simultaneously takes care of all these factors to fit a distribution on the expression levels of individual genes. Our model is based on a gene-wise unimodal or bimodal Gaussian distribution coupled with a generalized linear model with a probit link. We identify the gene-specific and cell-specific effects on overall dropout probability and overall expression levels. We aim to distinguish the zeros arising due to technical errors from zeros appearing due to biological factors. We have also developed a method for testing differential expression between two groups taking care of all intricacies of single-cell RNA-seq data and evaluated asymptotic properties. Our model can be readily used for other downstream analysis that can be performed using single-cell RNA-seq data. Extensive simulation and real data analysis validate the performance of our methods.

Omics Technologies Posters - Wednesday

PB3092. Optimizing microbiome extraction methods for human intestinal resections.

Authors:

E. P. Ryu¹, W. A. Koltun², G. S. Yochum², E. R. Davenport¹; ¹Pennsylvania State Univ., University Park, PA, ²Pennsylvania State Coll. of Med., Hershey, PA

Abstract Body:

Fecal samples are typically used as a proxy for the human gut microbiome. While fecal sampling is non-invasive and convenient, microbiome diversity varies along the intestinal tract and the fecal microbiome does not fully capture this variation. This issue may be particularly impactful for host genetic-microbiome studies, as host genetic variants are more likely to interact with taxa residing on the mucosal layer of the intestinal lining than those in the lumen. As a result, the microbiome collected from intestinal tissue resections may provide greater insight into the associations between host genes and microbes compared to fecal samples. The caveat is that there are no standardized protocols for characterizing the microbiome from tissue. This is problematic for two reasons: 1) tissue samples are low microbial biomass environments, so the microbiome is prone to kit contamination, and 2) the vast amount of host DNA can wash out microbial DNA. This second issue can make the PCR amplification standard in 16S rRNA gene sequencing studies particularly challenging, as primers can bind to off-target regions in the host DNA. To address this gap, I am performing a comprehensive benchmarking study to identify the ideal i) DNA extraction kit, ii) tissue sample type, and iii) 16S rRNA gene primer set for generating microbiome data from human tissue samples. First, I am comparing three kits that have been previously used to extract the microbiome from intestinal biopsies: Zymo Quick-DNA, DNeasy Blood & Tissue, and DNeasy PowerLyzer PowerSoil. Second, I am using two sample types: full-thickness terminal ileum resections and corresponding mucosal scrapings. Samples have been acquired from the Inflammatory Bowel Disease Tissue Biobank at the Penn State Milton S. Hershey Medical Center (n = 4 full-thickness and n = 2 mucosal lining per kit, for a total n = 18). Finally, I am testing two 16S primer sets: standard V4 Earth Microbiome Project primers, and V1 primers, which have been reported to have superior performance in samples with heavy host DNA contamination. Evaluation across methods will include identifying the extent of kit contamination for each DNA extraction kit, total material extracted, and amount of host DNA off-target amplification in the 16S rRNA gene sequencing results. This study will be one of the first to compare these three DNA extraction kits, different sample types, and different primer sets for intestinal tissue microbiome extraction. Establishing tissue microbiome extraction protocols will facilitate investigations of host genetic-microbiome associations and push the field towards examining the host genes most closely involved with the microbiome.

Omics Technologies Posters - Thursday

PB3093. OrphaID: a new platform for rare intellectual disabilities in Orphanet in partnership with ERN-ITHACA

Authors:

M. Amin¹, A. Olry¹, V. Serriere-Lanneau¹, C. Zweier², A. Hugon³, B. Popp⁴, C. Rodwell¹, C. Lucano¹, H. Ali¹, A. Rath¹, A. Verloes³; ¹INSERM, US14-Orphanet, Paris, France, ²Univ Erlangen-Nürnberg, Erlangen, Germany, ³Robert DEBRE Univ. Hosp., Paris, France, ⁴Univ. of Leipzig Hosp. and Clinics, Leipzig, Germany, Leipzig, Germany

Abstract Body:

Background: Intellectual disability (ID) is a group of heterogeneous disorders affecting up to 3% of worldwide population and characterized by significant impairment in cognition and behavior which can be associated with other syndromic or dysmorphic features. The heterogeneity of intellectual disabilities and rarity of most forms render the genetic diagnosis challenging in most cases. Hence, stemmed a need for developing a comprehensive list of curated intellectual disability related genes and associated phenotypes. **Methods:** Orphanet, which is the largest portal for rare diseases in the world has partnered with the European Reference Network on Rare Congenital Malformations, Autism and Rare Intellectual Disability (ERN-ITHACA) in order to develop a platform (OrphaID) for curated intellectual disability related genes and phenotypes. This list of intellectual disability-related genes is curated in partnership with ERN-ITHACA and SysID/SysNDD (www.sysid.dbmr.unibe.ch). **Results:** OrphaID currently (March 2022) contains 803 Orpha-coded ID phenotypes linked to 1390 curated ID-genes. This will allow systematic and comprehensive access to and retrieval of ID specific information within Orphanet. The list of OrphaID entries is linked to both internal (e.g. genes information, classification, HPO signs) and external resources (e.g. SysNDD, OMIM). It can be searched or filtered by gene, disease or OMIM number and the results can be downloaded in different file formats (excel, csv and txt). **Conclusion:** OrphaID is a valuable tool for all users (patients, researchers and clinicians) with interest in rare intellectual disabilities. The platform is freely accessible at <https://id-genes.orphanet.app/ithaca/>.

Omics Technologies Posters - Wednesday

PB3094. Outcomes from the Care4Rare Canada unsolved research pipeline for 950 patients with non-diagnostic exome sequencing data

Authors:

K. Boycott¹, T. Hartley¹, K. Kernohan¹, D. Dymant¹, E. Soubry¹, C. Marshall², M. Couse², R. Mendoza², J. Dowling², M. Acker², F. Bernier³, M. Innes³, B. McInnes³, J. Parboosingh³, R. Lamont³, P. Au³, A. Lehman⁴, M. Waqas⁴, A. Mohajeri⁴, Care4Rare Canada Consortium; ¹Univ. of Ottawa, Ottawa, ON, Canada, ²Univ. of Toronto, Toronto, ON, Canada, ³Univ. of Calgary, Calgary, AB, Canada, ⁴Univ. of British Columbia, Vancouver, BC, Canada

Abstract Body:

Progress toward the discovery of the genetic basis of every rare disease (RGD) has been substantial over the past decade secondary to the introduction of exome sequencing, with the identification of 250-300 new disease-gene associations every year. However, despite our increased understanding of the mechanisms of RGDs, one-third of patients remain undiagnosed because the molecular mechanism underlying their rare disease is either not yet recognized within, or is outside of, the coding genome. We recruited 950 unsolved patients with a suspected RGD who had non-diagnostic exome sequencing data from either the clinical or research setting. Reanalysis of the existing exome data from the proband, and in most cases additional family members, provided a diagnosis in a known disease gene for 82 families (9% of the cohort) and identified novel candidate genes in 195 families (21%), of which 79 have been validated as disease-causing (8%) using data-sharing and laboratory investigations. Participants remaining without a diagnosis or compelling candidate were considered (based on phenotype and sample availability) for an unsolved multi-omics research pipeline. The outcome of the first 133 families has shown a total gain in diagnostic yield of 13.5% (18/133) (11% for short-read genome sequencing (15 of 133 analyzed thus far), 8% for transcriptome sequencing (2 of 26), and 8% for long-read genome sequencing (1 of 14)) above and beyond what could be detected in the exome sequencing data, suggesting that disease mechanisms beyond the reach of the exome are important contributors to the unsolved cohort. Additional analysis of these data will likely further increase their contribution to the diagnostic yield. Leveraging opportunities to promote international sharing of deeper levels of clinical and 'omic data and utilizing new technologies to understand RGDs is an important path forward to complete the morbid anatomy of the human genome and enable diagnoses for all patients with a RGD in the coming decade.

Omics Technologies Posters - Thursday

PB3095. Overcoming FFPE hurdles to enable high quality hybrid capture libraries and somatic mutation detection in matched tumor-normal patient samples.

Authors:

S. Chavadi; NEB, Burlington, MA

Abstract Body:

Overcoming FFPE hurdles to enable high quality hybrid capture libraries and somatic mutation detection in matched tumor-normal patient samples.

In cancer genomics, a common source of DNA is formalin-fixed, paraffin-embedded (FFPE) tissue from patient surgical samples, where in most cases high quality fresh or frozen tissue samples are not available. FFPE DNA poses many notable challenges for preparing NGS libraries, including low input amounts and highly variable damage from fixation, storage, and extraction methods. Due to the high cost of sequencing and variability of coverage, regions of interest are often specifically enriched using hybrid capture-based approaches, but these methods require a high input of diverse, uniform DNA library to achieve the coverage required for somatic mutation identification in tumor samples.

We applied a multi-faceted approach to improving the yield, quality, and coverage from libraries generated from cancer patient sample sets. Using tumor-normal pairs, we evaluated the impact of our method on the sensitivity and accuracy of somatic variant detection. Matched frozen tissue provided a truth set of somatic variants for a particular tumor as well as a gold standard by which we could assess the quality of our FFPE hybrid capture libraries. Combining DNA damage repair and a novel enzymatic fragmentation mix upstream of library preparation not only reduced the false positive rate in somatic variant detection by repairing damage-derived mutations but also improved the library yield, complexity, coverage uniformity, and hybrid capture library quality metrics. The use of unique molecular identifier (UMI)-containing adaptors ensured accurate duplicate marking. Finally, a new PCR master mix boosts the library yield without compromising library quality in FFPE samples ranging from very poor quality to high quality. This combinatorial approach allows even the poorest quality FFPE samples to achieve high quality libraries with sufficient input for hybrid capture in our study.

Omics Technologies Posters - Wednesday

PB3096. PacBio HiFi sequencing provides highly accurate CpG methylation calls without bisulfite treatment.

Authors:

C. Saunders, D. Portik, R. J. Hall, D. Seifert, D. Diaz, A. Töpfer, K. Nyquist, A. M. Wenger; PacBio, Menlo Park, CA

Abstract Body:

PacBio long-read sequencing observes a polymerase in real time as it incorporates fluorescently labeled nucleotides to synthesize a DNA strand. Kinetic signatures from the sequencing polymerase, including pulse width and interpulse duration, correlate with chemical modifications to the canonical DNA bases. These kinetic signatures enable the accurate detection of 5-methylcytosine (5mC) at CpG sites from standard PacBio HiFi reads, without requiring bisulfite treatment or other modifications to sample preparation or sequencing. HiFi CpG methylation calls are consistent with orthogonal technologies, including a correlation of 0.95 with methylation results from short read whole genome bisulfite sequencing. Unlike short read bisulfite sequencing, HiFi sequencing has no observed depth bias against CpG islands, and the methylation calls are made in highly-accurate long (15-20kb) reads, facilitating the identification of cis genetic variants in phase with the methylation signal. Long HiFi reads can map to more CpG sites in the genome; when aligned to CHM13, methylation can be called on 11% more reference CpG sites compared to short-read sequencing. We demonstrate that methylation calls from HiFi reads can detect and characterize various clinically relevant conditions, including repeat expansions and imprinting disorders, highlighting their potential for rare disease and cancer research applications.

Omics Technologies Posters - Thursday

PB3097*. Paired whole genome and whole transcriptome sequencing of a large cohort of undiagnosed pediatric trios with neurological phenotypes

Authors:

R. Wang¹, Q. Zhang¹, K. Ramsey², K. Bluske¹, A. Chandrasekhar¹, Y. Shin¹, J. albay¹, S. Batalov³, D. hernandez¹, A. Crawford¹, M. Bainbridge⁴, M. Huentelman², R. Taft⁵; ¹Illumina, Inc, San Diego, CA, ²TGen, Phoenix, AZ, ³Rady Children's Hosp., San Diego, CA, ⁴Rady Children's Hosp., San Diego, CA, ⁵Illumina Inc, San Diego, CA

Abstract Body:

Whole transcriptome sequencing (WTS) for undiagnosed genetic disease provides an attractive addition to whole genome sequencing (WGS) by providing functional annotation of variants whose effects are difficult to determine otherwise. While the diagnostic rate for WGS ranges from ~30-60% depending on the clinical indications for testing, recent efforts to incorporate orthogonal data types such as RNA-seq have shown an increase in diagnostic yield particularly for variants of unknown significance, deep intronic variants, and splice-site adjacent variants.

We performed Illumina based WGS and WTS sequencing on 250 trios, where the pediatric proband showed signs of a neurological disorder, in collaboration with Rady Children's Hospital and Translational Genomics Research Institute (TGEN). We performed exhaustive analysis of multiple sample preparation methodologies, including side by side comparisons of ribodepletion and poly-adenylation library construction methods, and assessed fragment lengths, the impact of sample input type and amount, and the efficacy of targeted depletion of high abundance small RNAs.

Two strategies were applied to improve diagnostic yield and gauge their relative effectiveness. In an "RNA first" approach, we identified genes with outlier expression, splice junction usage, or exon utilization using blood based WTS. WGS data was used to identify nearby variants associated with retained introns and mis-spliced transcripts. In a "DNA first approach", we focused on variants of unknown significance and splice site adjacent variants outside of the typical +2 boundaries and investigated their relative effect on transcript abundance and splicing in the RNA-seq data. Using WTS data, variants of unknown significance (VUS) could be both ruled out or upgraded in their diagnostic classification. Our trio-based design allowed for confirmation of pathogenic splicing through segregation analyses within family members.

In this work we also assessed the need for standardized reporting of pathogenic splice variants for potential use in a clinical setting. Lastly, we share our experiences with quality control, annotation, and software implementation to integrate WGS and WTS data.

Omics Technologies Posters - Wednesday

PB3098. Papillary renal cell carcinoma (pRCC) functional heterogeneity unraveled by single nucleus sequencing (Sn-Seq) technologies

Authors:

R. Xin, Y. Zhu, Z. Peralta, M. Gibbons, V. Giangarra, S. Taylor; 10xgenomics, Pleasanton, CA

Abstract Body:

Renal cell carcinoma (RCC) is the most common type of kidney cancer worldwide with an approximate 20% mortality rate. Papillary renal cell carcinoma (pRCC), the second most common RCC, accounts for 10-15% of all RCCs. Compared to other subtypes, developing treatment for pRCC is challenging due to the inherent tumor heterogeneity, diverse molecular mechanisms and acquired drug resistance. To reveal the underlying mechanisms and identify specific biomarkers, we examined pRCC frozen tissues via single nucleus RNA sequencing (snRNA-seq) and single nucleus ATAC sequencing (snATAC-seq), using frozen tissue nuclei prepared by the 10x Genomics Chromium Nuclei Isolation kit.

Single cell sequencing technologies are powerful toolkits for characterizing tumor cell populations and the tumor microenvironment. However, accessing and working with tissues can be challenging as these require careful coordination with clinical sites. Enzymatic digestion-based single cell dissociation also introduces technical biases and biological artifacts. Single nucleus sequencing (Sn-Seq) provides an alternative strategy for profiling both gene expression profile and chromatin accessibility from frozen and fresh tissues.

Here, we profiled the transcriptome of more than 30,000 nuclei from two independent pRCC donors. To identify potential malignant cell populations, we compared our data with the single cell RNAseq datasets from three independent healthy donors (Human Cell Atlas: <https://data.humancellatlas.org/>) and characterized 9 distinct cell populations. Integration analysis identified 5 unique clusters of epithelial origin in the pRCC samples. Interestingly, we found SH3RF3 and TBC1D8, genes usually highly expressed in endothelial cells in these unique tumor clusters. Both type1 (such as MET) and type 2 (such as NFE2L2 and SETD2) pRCC markers were also overexpressed in these clusters, suggesting these cells are potentially malignant. Next, multiomic integration of snRNA-seq with snATAC-seq data revealed consistent gene expression profile and chromatin accessibility for the major cell types identified. We also performed CNV analysis based on snATAC-seq data and detected potential CNV events located in 41 distinct genomic regions, including copy gains of vascular growth factors (VEGFs) and genes in MAPK signaling pathway.

Our multiomic study revealed diverse molecular mechanisms underlying pRCC heterogeneity, spanning genomic aberrations, and dysregulation of signaling pathways. These findings highlight that adoption of a more standardized Sn-Seq workflow may facilitate the clinical research and the development of more effective pRCC therapies.

Omics Technologies Posters - Thursday

PB3099. Peripheral leukocyte transcriptome dynamics following ischemic stroke of large vessel and cardioembolic etiologies

Authors:

P. Carmona-Mora¹, B. Knepp¹, G. C. Jickling², X. Zhan¹, M. Hakoupiian¹, H. Hull¹, N. Alomar¹, H. Amini¹, F. R. Sharp¹, B. S. Stamova¹, B. P. Ander¹; ¹Univ. of California-Davis, Sacramento, CA, ²Univ. of Alberta, Edmonton, AB, Canada

Abstract Body:

Peripheral blood leukocytes mediate the response following Ischemic Stroke (IS) and display distinctive transcriptomic signatures that allow IS patients to be distinguished from controls. Dissecting the temporal dynamics of gene expression in peripheral leukocytes after IS can guide the refinement of diagnostic biomarker candidates and identify drug targets. RNA-seq was performed on peripheral monocytes, neutrophils, and whole blood samples (38 IS and 18 controls). A whole genome co-expression network analysis (WGCNA) was used to identify genes whose expression is associated with time after IS. Differential expression (DE) analyses were performed by binning patients into time point groups and per stroke etiology. Neural network analyses (SOM), were performed on DE genes to enable clustering into smaller groups based on their trajectory over time. WGCNA identified dozens of genes associated with time after stroke, including cell-specific markers and monocyte subtype and polarization markers. Some highly interconnected time-associated hub genes correlated with stroke severity. These include *TIA1*, a TNF- α repressor in monocytes, and constant and variable immunoglobulin genes in whole blood. DE analyses showed distinctive signatures for all sample types and at all time bins and IS etiologies, revealing many DE genes (DEGs) at only one time bin and IS cause. Unique patterns of temporal gene expression and pathways were distinguished for monocytes, neutrophils and whole blood with enrichment of interleukin signaling pathways for different timepoints and IS etiologies. SOM clustered DEGs as a function of time post-IS, revealing those that increase or decrease expression over time, or those peaking only at a specific time bin. SOM profiles identified important attributes of DEGs from opposite expression trajectories. In neutrophils, the toll-like receptor pathway is associated with a profile that decreases on day 1 and increases after 24 h in cardioembolic stroke. Whereas, for large vessel stroke, this pathway is enriched in a DE profile that decreases at all times compared to controls. Toll-like receptor signaling is a promising target for treating cardiovascular disease, impacting downstream pro- or anti-inflammatory molecules, like TNF- α , interleukins, and interferons. Consideration of time after stroke as a key variable enables the refinement of biomarker genes for diagnosis. This study identified potential time- and cell-specific biomarkers and drug targets. Our approach represents a model for assessing key genes in acute disease in a more strategic manner considering dynamic factors that influence pathophysiology and treatment selection.

Omics Technologies Posters - Wednesday

PB3100. Perturb-seq analysis reveals key mediators of TNF α -induced transcriptional response.

Authors:

D. Tedesco, T. Liu, D. Hu, N. Isachenko, D. Deng, N. Dolganov, A. Chenchik; Cellesta, Inc., Mountain View, CA

Abstract Body:

To identify genes involved in TNF α induced transcriptional response, a 10X single-cell Perturb-Seq screen was run on HEK293-Cas9 cells with an 88-sgRNA pooled library. As a benchmark, a parallel screen was run with cells transduced in arrayed format with a subset of the sgRNAs of the pooled library. The 10X Perturb-Seq screen was carried out in triplicate with increasing amounts of cells/10X reaction, in order to determine: (a) the minimum number of cells/sgRNA required to cover the full complexity of the pooled library, (b) the maximum number of cells that can be loaded in the 10X reaction without losing single-cell resolution due to encapsulation of cell doublets. Transcriptional profiling identified TNFRSF1A, IKBKG, RELA, and CHUK as key mediators of TNF α transcriptional response in both the arrayed and pooled library screens. Increasing the number of cells per 10X reaction had a positive effect on the sgRNA library coverage, without adversely affecting the magnitude of the observed effect of TNFRSF1A, IKBKG, RELA, and CHUK knockout (KO) on TNF α mediated transcriptional response. Perturb-seq was confirmed as a promising technology for the efficient and scalable interrogation of many individual gene in one single experiment, enabling the identification of transcriptional profiles linked to specific gene-KOs.

Omics Technologies Posters - Thursday

PB3101. Peruvian Hospital<Urban Antibiotic Resistance found in Local Wastewater

Authors:

L. Jaramillo Valverde¹, N. Pablo-Ramirez¹, V. Roa- Linares¹, C. Martinez-Jaramillo¹, S. Alvites-Arrieta¹, M. Ubillis¹, D. Palma-Lozano¹, S. Davison², A. Gomez², H. Guio¹; ¹UNIVERSIDAD DE HUANUCO, HUANUCO, Peru, ²Univ. of Minnesota, MINNESOTA, MN

Abstract Body:

BACKGROUND: Metagenomics studies all the genetic material of ecosystems, which can provide information about the presence of bacterial species, pathogens and virulence genes. Most studies of bacterial resistance in Peru have been carried out at the hospital level with limitations in culture methods and not at the community level (resistomes). Considering that many pathogens are not cultivable or bacterial culture media are not available, the objective of this project is to evaluate antimicrobial resistance genes (ARGs) due to the diversity and abundance of antimicrobial resistance genes based on metagenomic analyzes in urban wastewater from hospitals and community in a location of 2021. The present investigation is of an explanatory and experimental type, where wastewater samples will be taken from two health systems (MINSA and EsSALUD) and from the community. **METHODS:** In this study, a high-throughput sequencing-based metagenomic approach was applied to investigate the community composition of bacteria and ARGs in untreated wastewater from two different types of hospitals (MINSA - public and EsSalud - private hospital) and one form community. Here, we utilized metagenomic approaches to comprehensively reveal the diversity, abundance and hosts of ARGs in wastewater from three location. **RESULTS:** In total, 13 species and other less abundant taxa were identified from the microbiota of the wastewaters, with some different bacterial community compositions among the three locations. Total ARG analysis using the Antibiotic Resistance Genes Database (ARDB) and Comprehensive Antibiotic Resistance Database (CARD) revealed that the microbiota in the wastewaters from the three hospitals harbored different types and percentage of ARGs. Essalud - private reported the least bacterial diversity but the greatest diversity in resistance genes. *Prevotella copri* was the most abundant species is Essalud and probably influencing the pathogenesis of Rheumatoid Arthritis. Valdizan - public reported TEM beta- lactamase which as most prevalent ARG. More than 90% of resistance to ampicillin in *Escherichia coli* is due to the production of TEM. Finally, in urban waterlines we observed major number of pathway gene. **CONCLUSION:** In summary, our findings demonstrated a widespread occurrence of ARGs and ARG-harboring microbiota in untreated wastewaters of different hospital, suggesting that protection measures should be applied to prevent human infections. Concurrently, hospital wastewater should be treated more specifically for the removal of pathogens before its discharge into the urban sewage system.

Omics Technologies Posters - Wednesday

PB3102. PhaseDancer: Targeted assembly of the complex syntenic regions in non-human primates of the HSA2 fusion site.

Authors:

B. Poszewiecka¹, K. Gogolewski¹, P. Stankiewicz², A. Gambin¹; ¹Univ. of Warsaw, Warsaw, Poland, ²Baylor Coll. of Med., Houston, TX

Abstract Body:

We have developed a genome assembler for long error-prone read sequencing data targeted at regions enriched with segmental duplications (SDs). PhaseDancer follows a bottom-up paradigm by taking as an anchor short sequence to extend iteratively using greedy strategy. Reads mapped to the anchor sequence are first clustered and then one cluster is selected and subjected to assemble the anchor sequence for the next iteration.

This innovative approach enables the local de novo assembly in a significantly reduced time with an increased continuity when compared to the existing whole-genome de novo assemblers. Additionally, PhaseDancer is accompanied with the web application for visualizing every iteration of the algorithm, providing constant insight into the complexity of the processed sequences and enabling the incorporation of the expert decisions if needed. PhaseDancer has proven to be used for the local assembly of the highly variable genomic regions, improving the understanding of the structure variability of the mosaic genomic regions at the nucleotide level.

To validate the functionality of our tool, we have assembled the subtelomeric regions of chromosomes 2A and 2B of the chimpanzee, bonobo and gorilla genomes, significantly extending the reference sequences. Finally, we demonstrate PhaseDancer superiority over the existing tools in the analysis of evolutionary complex SDs.

Omics Technologies Posters - Thursday

PB3103. Phenome-wide gene prioritisation leveraging Knowledge Graphs, Graph Convolutional Networks and UK Biobank PheWAS.

Authors:

D. Vitsios¹, I. Melas¹, L. Middleton¹, B. Rozemberczki¹, R. Dhindsa², A. Harper¹, G. Edwards¹, S. Petrovski¹; ¹AstraZeneca, Cambridge, United Kingdom, ²Baylor Coll. of Med., Houston, TX

Abstract Body:

The growth of genomic datasets continues to fuel advances in human-validated target identification and validation, encompassing multiple layers of biological annotations. Best leveraging the wealth of available gene annotations to infer gene-disease associations is far from trivial. Here, we present Mantis-ML v2, building onto our previous work (Vitsios et al, AJHG 2020), but with two key novel advances. Firstly, we now leverage knowledge graphs to complement the existing structured gene features from dozens of annotation resources, including features selected via natural language processing. Secondly, we adopt Graph Convolutional Networks (GCNs), employing Stargraph, to leverage the combined information from the knowledge graph and the structured features. Knowledge graphs capture relationships between genes such as protein-protein interactions, co-expression, signalling reactions etc., augmenting our structured knowledge-base of features. The underlying hypothesis is that genes closely interacting with each other are likely to have similar properties and potentially play a similar role in disease. The knowledge graph we constructed consists of 8.7M edges between genes and related entities and is a subset of AstraZeneca's Biological Insights Knowledge Graph (BIKG), combining data from 52 data sources. Graph-based gene features are also calculated on top of BIKG, capturing structural information such as gene importance, influence and clusters. Our GCN classifier is trained across hundreds of balanced datasets within Mantis-ML's robust semi-supervised learning framework, providing out-of-fold gene-disease associations across the entirety of the human exome. Our revised gene prioritisation model achieves significantly increased performance, with an overall 11.2% boost in classification power (median AUC=0.95 across 5,000 diseases from Human Phenotype Ontology and OpenTargets compared to AUC of ~0.85 from our original Mantis-ML). Notably, we provide refined gene-disease associations by statistically evaluating the overlap of top Mantis-ML v2 predictions against an extensive set of orthogonal genetic signals, derived from a phenome-wide association study (PheWAS) on ~450k UK Biobank participants. Accompanying all the results is an interactive web-resource with the scores and rankings specific to either a disease or gene. We expose the overlap of high-ranking Mantis-ML genes with those significantly associated within the UKB 450k exomes PheWAS, serving as both a means to cross-validate the new associations and to introduce a stronger form of gene prioritisation through the intersection of the two orthogonal statistics.

Omics Technologies Posters - Wednesday

PB3104. PIPseq, a novel and highly scalable technology based upon Pre-templated Instant Partitions (PIPs), is powering single-cell RNA sequencing into the million cell era

Authors:

K. Fontanez, A. Osman, R. Meltzer, C. Hayford; Fluent BioSci., Watertown, MA

Abstract Body:

Single-cell transcriptomics have revolutionized modern biology by providing unprecedented insight into the mechanisms of cellular differentiation and function in complex tissues and organisms. As scRNA-seq transitions from a discovery tool into translational applications requiring complex experimental design, multiple sample comparisons, and multivariable sample perturbations, there is a growing need for simple, highly scalable, cost-effective, library preparation methods. Current scRNA-seq methods have limited scalability. Droplet based applications generate one drop at a time, limiting throughput, and rely on complex, expensive instrumentation. Nano-well array based methods are limited by surface area, and therefore locked to defined manufactured consumables. Combinatorial barcoding enables high throughput transcriptomics but is limited by complex, labor intensive workflows. Fluent BioSciences has developed new methods for scRNA-seq enabled by Pre-templated Instant Partitions (PIPseq), that enables sensitive, efficient, and highly scalable scRNA-seq improving the accessibility and application of high-cell input single cell transcriptomics. Here we demonstrate single-tube PIPseq configurations accommodating 200,000 - 1M cell input and evaluate key performance metrics using model cell lines, peripheral blood mononuclear cells (PBMCs), and complex tissues. The PIPseq platform enables novel experimental paradigms not addressable with current platforms such as highly multiplex sample batching, targeted sequencing of large numbers of cell input, and identification of very rare cell populations from complex mixtures.

Omics Technologies Posters - Thursday

PB3105. Population diversity and selection of recent gene duplications detected using a complete human genome sequence

Authors:

D. Soto¹, A. Sekar¹, G. Kaya¹, M. Mastoras¹, E. Green², A. M. Andres³, M. Y. Dennis⁴; ¹Univ. of California Davis, Davis, CA, ²NHGRI (NIH), Bethesda, MD, ³Univ. Coll., London, United Kingdom, ⁴Univ. of California, Davis, Davis, CA

Abstract Body:

Human-specific duplicated genes (HSDs) are strong candidates for neurodevelopmental traits and diseases unique to our species. Assessment of the recently published T2T-CHM13 complete genome identified 417 genes embedded in recent segmental duplications (>98% sequence identity; SDs-98) with near fixed copy-number (CN=2) in the 1000 Genomes Project (1KGP), with 231 having evidence of expression in the fetal neocortex. This list includes genes with known roles in neurodevelopment, such as ARHGAP11B, as well as many other uncharacterized genes. Comparing CN across 1KGP populations, we also identified 174 genes that show stratification (VST>95th percentile), including those with previous evidence (e.g., KANSL1) and interesting new candidates (e.g., NPY4R, previously implicated in body-mass index). Examining our ability to detect SNVs and indels across SDs-98, we compared short- and long-read sequence data from eight individuals from the T2T Diversity Panel and observed a recall of <0.1 resulting in a depletion of total variants identified from 1KGP across these regions (12 variant/kbp) versus the unduplicated genome (38 variant/kbp). Using 1KGP variants in accessible regions (representing ~10% of SDs-98), we assessed signatures of natural selection by calculating Tajima's D and identified 22 protein-encoding genes showing consistent outlier values, including the SPDYE3 and PMS2P1 locus, which is CN fixed. To enable detection of variation across a set of 34 fixed HSD genes, we performed targeted capture PacBio HiFi sequencing in ~200 1KGP individuals from five populations. The approach yielded ~3 kbp length reads with average coverage >80×, from which we identified 20,043 biallelic SNVs (73.4% novel). Using a combination of dN/dS and Ka/Ks, we show that ancestral paralogs universally exhibit signatures of purifying selection while duplicate paralogs exhibit relaxed constraint, suggestive of neutrality or perhaps positive selection. We also identified ten regions with high levels of population differentiation (measured by FST), overlapping genes HYDIN2 and SRGAP2C, putatively due to local adaptation. Our approach highlights potential evolutionarily relevant human gene duplications, which will become priority candidates for future functional studies.

Omics Technologies Posters - Wednesday

PB3106. Population Scale Genetic Interpretation Software for Reporting Pathogenic and Likely Pathogenic Variants Impacting the ACMG59 Genes.

Authors:

R. Olson, J. Arroyo, J. Morales-Rosado, M. Arioti, P. Duffy, S. Safgren, K. Kolbert, N. Larson, H. Sicotte, K. Lazaridis, J. Olson, E. Klee; Mayo Clinic, Rochester, MN

Abstract Body:

Pathogenic (P) and Likely Pathogenic (LP) variants within the ACMG59 genes are medically actionable variants associated with successful intervention strategies capable of reducing the public health burden of common genetically inherited diseases, but manual population level screening is ultimately time consuming. In this work, we developed, validated, and utilized a Scalable Automated Variant Interpretation (SAVI) software to prioritize potentially reportable variants in the ACMG59 (+*LDLRAP1*) genes from 983 individuals having genome sequencing.

SAVI is Python based software to annotate variants within a sample VCF to prioritize all P and LP variants using information from CAVA sequence ontology, catalogs (HGMD, ClinVar, and gnomAD), and in silico tools such as REVEL and SpliceAI. SAVI compiles this information to generate two distinct metrics, an internally developed InDepth Score (IDS) and a Range of Pathogenicity Classification (RPClass), based on ACMG guidelines.

Performance testing of the IDS (1-6 likely P/LP, 29-59 unlikely P/LP) in 782 variants previously curated resulted in 99.5% variants with an IDS of 1-6 classified as P/LP, while 3.4% were reported as P/LP with an IDS of 29-59. In comparison, 82.6% of variants with an IDS of 1-6 were classified as P/LP and 8.6% of variants with an IDS of 29-59 were classified as P/LP by InterVar.

Using SAVI software, we successfully processed 983 population samples with genome sequencing enrolled in the Biobank study at the Mayo Clinic and completed manual curation of prioritized variants. ~7.3 million variants across the ACMG59 (+*LDLRAP1*) were identified after bioinformatic genome alignment and quality variant filtering. 258 discrete variants represented by 433 total variants (0.006%) within 335 samples (34%) were prioritized by SAVI. After manual curation, 42 discrete variants, found in 63 samples (6.4%) were classified as P/LP and therefore reportable to the patient following clinical variant confirmation. Three of the reportable variants had an IDS of 1-6 (100% P/LP), 39 had an IDS of 7-21 (45% P/LP), and all 169 variants with an IDS score 22-59 were classified as VUS/LB/B.

In summary, we have developed and implemented a tool that has a high degree of accuracy for minimizing the prioritization of non-reportable variants while maintaining sensitivity. Future goals are to support the ACMG73 gene set and begin processing the Mayo Clinic Project Generation study (55k samples). Additionally, we will continue to refine the prioritization scoring algorithm using machine learning to minimize the false positivity rate to further reduce the burden of preventative screening on a population level.

Omics Technologies Posters - Thursday

PB3107. Predicting the transcriptional activity and mechanism of action of small molecules using deep learning

Authors:

C. Barnhill, A. Petukhov, J. Chang, O. Botvinnik, J-H. Ju, S-G. Ji; BridgeBio Pharma, Inc., Palo Alto, CA

Abstract Body:

Predicting the effect of a small molecule on gene expression in silico offers a unique approach to discover therapeutic compounds and identify unintended off-target effects. Additionally, predicting the mechanism of action (MOA) of molecules that are transcriptionally active may yield a more specific understanding of which genes are affected. We present a deep learning approach to predict the transcriptional activity of small molecules based solely on the structural information contained in the graph representation of molecules using a graph neural network (GNN). Data from the Connectivity Map, a resource consisting of experimental data that profiles the effects of perturbagens (small molecules as well as gene overexpression and knockdown reagents) on transcriptional activity in human cell lines, was used to train two separate models. The first model is a regression task to predict the transcriptional activity score (TAS) of a compound. TAS is an aggregate measure that represents an upper bound of a compound's transcriptional activity across multiple cell lines. We applied a quality control (QC) process to the TAS dataset to remove outliers for any of six QC metrics specific to the dataset, yielding a final dataset of 22,032 compounds. We trained a GNN using Bayesian hyperparameter optimization with RMSE as the validation metric. The second model is a multilabel classification task to predict the MOA of a compound. Compounds in the dataset are annotated with MOAs based on literature review and the similarity of their transcriptional signatures. A single compound can have multiple MOA labels. We applied a separate QC process to the MOA dataset to remove duplicate compounds and restrict to MOAs with at least 10 unique compounds, yielding a final dataset of 28,204 compounds and 89 MOAs. We trained a separate GNN using Bayesian hyperparameter optimization with weighted average precision as the validation metric. We applied both trained models to compounds in the ZINC database to validate how this method might be used for discovery. ZINC is a dataset of over 997 million compounds for virtual screening of small molecules. Compounds in ZINC are labeled by their molecular weight (MW) and LogP (solubility) values. We restricted our analysis to a sample of one million lead-like compounds with MW between 250 and 350 Daltons and LogP between 1 and 3. We identified 3,184 molecules predicted to be transcriptionally active and with predicted class probability > 0.9 for at least one MOA. Experimental validation of these molecules is pending. This approach provides a new method to identify novel compounds that modulate gene expression in known MOAs.

Omics Technologies Posters - Wednesday

PB3108. Prokaryotic and viral genomes recovered from 787 Japanese gut metagenomes revealed microbial features associated with diets, populations, and diseases

Authors:

Y. Tomofuji¹, T. Kishikawa¹, Y. Maeda¹, K. Ogawa¹, Y. Otake¹, S. Kawabata¹, T. Nii¹, T. Okuno¹, E. Oguro-Igashira¹, M. Kinoshita¹, M. Takagaki¹, N. Oyama², K. Todo¹, K. Yamamoto¹, K. Sonehara¹, M. Yagita¹, A. Hosokawa³, D. Motooka⁴, Y. Matsumoto⁴, H. Matsuoka⁵, M. Yoshimura⁵, S. Ohshima⁵, S. Shinzaki¹, S. Nakamura⁴, H. Iijima¹, H. Inohara¹, H. Kishima¹, T. Takehara¹, H. Mochizuki¹, K. Takeda¹, A. Kumanogoh¹, Y. Okada^{1,6,7}; ¹Osaka Univ. Graduate Sch. of Med., Suita, Japan, ²Kawasaki Med. Sch., Kurashiki, Japan, ³Suita Municipal Hosp., Suita, Japan, ⁴Res. Inst. for Microbial Diseases, Osaka Univ., Suita, Japan, ⁵NHO Osaka Minami Med. Ctr., Kawachinagano, Japan, ⁶Graduate Sch. of Med., The Univ. of Tokyo, Hongo, Japan, ⁷RIKEN Ctr. for Integrative Med. Sci., Tsurumi, Japan

Abstract Body:

Microbial genomes recovered from the gut metagenome sequencing reads are important resources for studying the gut microbiome. However, the current populational diversity of the prokaryotic genomes is still limited because the number of microbial genomes recovered from populations other than European, North American, and Chinese is relatively low. The Japanese have unique dietary culture and habits, which resulted in the unique features of the gut microbiome. Recovering microbial genomes from the Japanese gut metagenome is necessary for obtaining deep insights into the Japanese gut microbiome and increasing the populational diversity of the databases. Here, we reconstructed 19,084 prokaryotic and 31,395 viral genomes from the 787 Japanese gut metagenome shotgun sequencing data as Japanese Metagenome Assembled Genomes (JMAG) and Japanese Virus Database (JVD) and made them publicly available. The JMAG and JVD are one of the largest microbial genome datasets for a single population and contribute to increasing the populational diversity of the microbial genomes. Utilizing these unique microbial genome databases, we explored the population-specific features of the gut microbiome. Population-specific enrichment of the *Bacillus subtilis* and β -porphyranase among the JMAG could derive from the Japanese traditional food natto (fermented soybeans) and nori (laver seaweed), respectively. Five food-associated bacterial species (four dairy-related species and natto-related *Bacillus subtilis*) were shared among the Japanese at the strain level and two dairy-related species were nominally associated with the East Asian-specific missense variant rs671:G>A in ALDH2 which was associated with dairy consumption. As for the viral genomes, 62.9% of the species-level clusters in the JVD were novel. The composition of the β crAss-like phages was low among Japanese but relatively high among the Africans and Oceanians. Several clades of the crAss-like phages decreased in rheumatoid arthritis, systemic lupus erythematosus, and inflammatory bowel diseases but increased in colorectal cancer. Analysis on the CRISPR sequences identified the viruses-prokaryotes interaction network and revealed that abundances of the viruses and their hosts tended to be positively correlated. In summary, we recovered the MAGs and viral genomes from the Japanese gut MSS data. We revealed the features of the Japanese gut metagenome, novel associations of the crAss-like phages to populations and diseases, and virus-prokaryote interaction. Our dataset which includes MAGs, viral genomes, and CRISPR spacers are publicly available (https://github.com/ytomofuji/JMAG_JVD).

Omics Technologies Posters - Thursday

PB3109. Proteoform inference using proteogenomic approach in non-small cell lung cancer and healthy control plasma proteomes reveals disease-associated protein isoforms

Authors:

M. Donovan, Y. Huang, J. Wang, A. Stukalov, T. Wang, A. Siddiqui, S. Batzoglou; Seer Bio, Redwood City, CA

Abstract Body:

Complexity of human biology far exceeds what is encoded in 20,000 protein-coding genes, however, this complexity may be largely explained by functionally distinct protein forms, or proteoforms, arising from alternative splicing, allelic variation, and post-translational modifications. Recent advances in proteomics are making it possible to analyze the plasma proteome at the level of individual isoforms at a scale not possible a few years ago. However, inference of proteoforms from detected peptides using bottom-up proteomic approaches remains challenging. Here we use the Proteograph™ Product Suite to perform quantitative proteomic experiments across a large cohort can improve proteoform inference by using peptide quantitative profiles. We performed a proteogenomic analysis of 80 healthy controls and 61 early-stage non-small-cell lung cancer (NSCLC) plasma samples to systematically infer proteoforms arising from alternative gene splicing and allelic variation. First, we hypothesized disease-associated proteoforms arising from alternative splicing would display differential abundance patterns. We performed a proteome-wide differential abundance analysis and identified four proteins with peptides with opposite patterns of abundance in NSCLC and healthy controls. Among these include BMP1, where inspection of the peptide locations relative to the known protein coding isoform models indicates the short protein isoform is associated with NSCLC samples and longer isoforms are associated with healthy samples. To identify other proteoforms arising from alternative splicing, we examined similarly abundant (i.e., covarying) peptides that also were sequentially clustered by implemented COrrelation-based functional ProteoForm (COPF) assessment with minor adjustments and additional filtering steps. We applied a post-translational cleavage detection strategy to 200 protein candidates with covarying peptides to test if covarying peptide clusters are disproportionately located on one protein terminus and identified over 20 proteoforms. Some identified proteoforms displayed opposite abundance patterns between the NSCLC and healthy samples, implying disease-associated function, including endostatin, a naturally occurring 20-kDa C-terminal short proteoform from type XVIII collagen (P39060) that has been shown to serve as an anti-angiogenic agent with evidence in treating NSCLC. These results demonstrate the identification of proteoforms offers increased opportunities to identify potential novel biomarkers for disease.

Omics Technologies Posters - Thursday

PB3110. Proteogenomic approaches improve genetic discovery across diverse human diseases.

Authors:

M. Koprulu, J. Carrasco-Zanini, E. Wheeler, N. Kerrison, N. Wareham, M. Pietzner, C. Langenberg; Univ. of Cambridge (MRC Epidemiology Unit), Cambridge, United Kingdom

Abstract Body:

Studying the plasma proteome as the intermediate layer between the genome and the phenome has the potential to identify disease causing genes and variants and improve our understanding of the underlying mechanisms. Here, we conducted a cis-focused proteogenomic analysis of 2,923 plasma proteins measured in 1,180 individuals using a novel antibody-based assay to identify disease causing genes across the human phenome and systematically refine causal genes at previously reported GWAS loci. We identify 1,553 distinct credible sets of protein quantitative trait loci (pQTL), a third of which (n=531) contained cis-pQTLs not previously reported. Of these, 182 signals were seen for 117 proteins never studied before, and 349 were detected despite proteins having been targeted in earlier, larger studies. We identified 224 proteins with robust evidence to contribute to the aetiology of 578 unique health outcomes using statistical colocalization, including proteins with potential for therapeutic interventions for type 2 diabetes (fibroblast growth factor 4 [FGFR4] and gastrin releasing peptide [GRP]). We demonstrated convergence between pQTL colocalisation and rare loss of function gene-burden evidence for disease associations of 25 proteins, including a role of TIMD4 in macrophage mediated LDL-cholesterol lowering. Proteogenomic evidence improved causal gene assignment; 480 of the credible sets overlapped with reported GWAS loci and highlighted different causal genes or refined longer lists of candidate genes for 40% of these loci.

Our findings demonstrate the ability of broad capture, high-throughput proteomic technologies to robustly identify new gene-protein-disease links, provide mechanistic insight, and add value to existing GWASs by enabling and refining causal gene assignment.

Omics Technologies Posters - Thursday

PB3112*. Pseudogenes limit the identification of common, functionally important transcripts generated by their parent genes

Authors:

E. Gustavsson¹, S. Sethi², Y. Gao², D. Zhang¹, J. Brenton¹, S. Garcia-Ruiz¹, R. Reynolds¹, A. Wernick^{1,3}, C. Arber¹, J. Evans^{1,3}, S. Wray¹, S. Gandhi^{1,3}, H. Houlden¹, C. Bento², H. Saini², J. Hardy¹, M. Ryten¹; ¹Univ. Coll. London, London, United Kingdom, ²Astex Pharmaceuticals, Cambridge, United Kingdom, ³Francis Crick Inst., London, United Kingdom

Abstract Body:

Genomic and transcriptomic studies are critically dependent on accurate gene annotations. However, this process can be complicated by highly homologous sequences which cause short RNA-sequencing (RNA-seq) reads to align to two or more genomic regions (multi-mapping). Given that the human genome contains 14,709 defective gene copies, termed pseudogenes, belonging to 3,491 parent genes of which 734 are known to cause Mendelian disease, this is a significant problem. The aim of this study was to investigate the impact of pseudogenes on transcriptomic analyses by focusing on the disease-relevant example of *GBA* and its expressed pseudogene *GBAPI*. By analyzing short-read RNA-seq data from the anterior cingulate cortex, we found that only $41.7 \pm 11.2\%$ of all reads mapping to *GBA* were uniquely mapping with $96.0 \pm 2.0\%$ of the multi-mapping reads being assigned to *GBAPI*. This led us to hypothesize that there are significant inaccuracies in the annotation of both genes. To address this, we generated targeted *GBA* and *GBAPI* PacBio SMRT isoform sequencing data from 12 regions of the human central nervous system (CNS) because of *GBA*'s importance in brain diseases. We identified 18 *GBA* transcripts with a novel open reading frame (ORF) and 7 novel *GBAPI* transcripts predicted to encode a protein. Focusing on *GBA*, we found that transcripts with a novel ORF collectively accounted for 15.8% - 31.7% of transcription, depending on tissue, and were all predicted to impact on catalytic function of *GBA* encoded beta-glucocerebrosidase. Transfection of novel transcripts into *GBA* knockout cells resulted in clear translation with no observed catalytic activity. Next, we used Oxford nanopore direct cDNA long-read sequencing data derived from the human CNS to analyse the transcription of other disease-relevant parent-pseudogene pairs, including *PTEN* and *PTENPI*, albeit in less detail. Again, we discovered inaccuracies in annotation, exemplified by novel transcripts of both genes. To generalise these findings further we performed annotation-independent re-analyses of public short-read RNA-seq data. Using data provided by the Genotype-Tissue Expression Consortium, we found that the proportion of parent genes with evidence of novel unannotated expression was significantly higher than that of protein coding genes without pseudogenes (Wilcoxon $p = 3.3e^{-12}$) and that this was a consistent finding across all human tissues. In summary, our results suggest systematic inaccuracies in the annotation of parent genes with significant implications for our understanding of disease-relevant genes.

Omics Technologies Posters - Wednesday

PB3113. PubCaseFinder: Implementation of the new function designing Mondo-based virtual gene panels to increase diagnostic efficiency for rare genetic diseases.

Authors:

T. Fujiwara¹, J-M. Shin¹, H. Saitsu², A. Yamaguchi³; ¹Database Ctr. for Life Sci., Kashiwa, Chiba, Japan, ²Hamamatsu Univ. Sch. of Med., Hamamatsu, Shizuoka, Japan, ³Tokyo City Univ., Setagaya, Tokyo, Japan

Abstract Body:

Next-generation sequencing-based diagnostic tests have been attempted for the diagnosis of affected individuals with suspected rare genetic diseases. Recently, virtual gene panels (VGPs), sets of causal genes associated with rare genetic diseases, have been used for the interpretation of the results of whole-exome sequencing (WES) and whole-genome sequencing (WGS). In the UK 100,000 Genomes Project pilot study on rare disease diagnosis, PanelApp which is a database containing VGPs has been applied for automated variant filtering to simplify and accelerate the time-consuming interpretation process of WGS results and improve diagnostic rates. PanelApp includes 332 VGPs of various granularities, such as the "Paediatric disorders" VGP containing 7019 causal genes and the "Amyloidosis" VGP containing 11 causal genes. However, the design of these VGPs of various granularities is a laborious task because they are curated manually. In this study, we implemented the new function in PubCaseFinder (<https://pubcasefinder.dbcls.jp>), a clinical decision support system for rare genetic diseases, to automatically design VGPs by using curated disease-gene associations (GPAs) and the Mondo disease ontology (Mondo). The MedGen, GenCC, and Orphadata databases provide publicly available curated GPAs for genetic diseases in OMIM and rare diseases in Orphanet. Mondo is an ontology, a standard and controlled vocabulary to represent knowledge in the domain, that integrates existing sources of disease definitions, such as OMIM, Orphanet, and Disease Ontology. It provides a hierarchical structure that can be used for classification or "rolling up" diseases to higher-level groupings. Using this feature, any VGPs can be created by grouping diseases at any hierarchy of about 25,000 diseases included in Mondo and grouping the causal genes related to those diseases. For example, PanelApp provides "Rare anaemia" VGP (99 genes), "Cytopenias and congenital anaemias" VGP (221 genes), and Severe Paediatric Disorders" VGP (2691 genes) related to anemia, while PubCaseFinder can provide "Hemolytic anemia" VGP (55 genes) and "Hematologic disease" VGP (395 genes) that PanelApp does not have. In this presentation, we report the details of the VGP design function implemented on PubCaseFinder and the evaluation of the diagnostic efficiency of the function using a set of affected individuals with suspected rare genetic diseases. As a result of the evaluation, we showed that our proposed Mondo-based VGPs contributed to improving the time-consuming interpretation process of WGS results and ultimately improving diagnostic efficiency.

Omics Technologies Posters - Thursday

PB3114*. Pyro-Velocity: Probabilistic and scalable RNA velocity inference from single-cell data

Authors:

L. Pinello¹, Q. Qin¹, E. Bingham¹, G. La Manno², D. Langenau³; ¹MGH/Harvard/BROAD, Charlestown, MA, ²EPFL, Lausanne, Switzerland, ³MGH, Charlestown, MA

Abstract Body:

Single cell RNA-seq assays have dramatically advanced our ability to study and model cellular differentiation and cell fate decision. RNA velocity is an analysis framework that has become fundamental in the toolbox of the single-cell research community. RNA velocity — the time derivative of the gene expression state — is a powerful technique used to predict cell fate based on spliced and unspliced reads from single-cell RNA-seq and RNA-processing kinetics models. However current RNA velocity models have several shortfalls: they do not provide any uncertainty estimation of the recovered velocity vectors and cell fate; they assume independent transcriptional processes for different genes and the required preprocessing choices (e.g. dimensionality reduction, KNN-smoothing) can dramatically influence their predictions and lead to misinterpretation of developmental order and cell fate, especially for non-expert users. To address these challenges, we propose Pyro-Velocity a probabilistic and end-to-end inference framework for RNA velocity based on variational inference that is scalable to millions of cells, based on unsmoothed data, and that naturally provides uncertainty estimation of cell fate based on a joint learning of transcriptional processes across genes. Our probabilistic framework also provides a natural estimation of uncertainty for kinetics parameters and latent time that can reflect sequencing noise or recapitulate biological processes like bifurcation points. In addition, Pyro-Velocity can be used to learn cell fate decisions from one dataset and predict RNA velocity to other datasets with similar biological contexts. We have tested and quantitatively evaluated Pyro-Velocity on several single-cell RNA-seq and lineage tracing datasets from developmental tissues and pediatric tumors. Our method outperforms available methods to infer RNA velocity based on observed fate supported by lineage information and/or latent time recovered with orthogonal methods. In addition, our method provides new visualizations to highlight uncertainty of the recovered vector fields or to show the parallel dynamic of key driver genes describing cell fate decision and commitment. In summary, Pyro-Velocity is a new probabilistic RNA velocity framework and a user-friendly and end-to-end software package to study cell fate decision in tumor and normal cells from single-cell RNA-seq data.

Omics Technologies Posters - Wednesday

PB3115. pyTCR: a comprehensive and scalable platform for TCR-Seq data analysis to facilitate reproducibility and rigor of immunogenomics research

Authors:

K. Peng¹, J. Moore², J. Brito¹, G. Kao¹, A. M. Burkhardt¹, H. Alachkar¹, S. Mangul¹; ¹Univ. of Southern California, Los Angeles, CA, ²Orange Coast Coll., Costa Mesa, CA

Abstract Body:

T cells are crucial components of the adaptive immune system as they are activated after being exposed to antigens. During the activation, V (variable), D (diversity), J (joining) segments in the T cells receptor loci undergo VDJ recombination to create diverse repertoires for recognizing and binding to the epitopes of the antigens presented by major histocompatibility complex (MHC). With the development of high throughput sequencing, TCR-seq provides the opportunities to understand adaptive immune responses, and further helps with diagnosis, prognosis prediction, and treatment outcome prediction in a variety of diseases including cancer, autoimmune disease, infectious disease, and allergies.

Due to the diversity and complicity of the TCR repertoire, computational methods are needed are important in understanding the features. Existing tools have promoted the advancement in TCR analysis. However, the existing tools fail to provide easy to use interface for biomedical researchers with no or limited background. They don't offer integrative analysis as they provide disjointed commands instead. Moreover, the analysis is not comprehensive as other tools are usually needed in order to finish the analysis. Furthermore, existing tools have limited options to customize the analysis and visualization. An alternative solution is urgently needed in this field.

pyTCR is a comprehensive platform with a rich set of functionalities of TCR repertoire analysis for biomedical researchers. Our cloud-based easy-to-use platform is based on the interactive notebook with the enhancement of reproducibility and transparency, by providing comprehensive and integrative functions, and customizable manipulations. The platform that pyTCR utilizes is interactive notebooks which code and results are all available to the users. pyTCR provides basic sample statistics such as number of reads, number of clonotypes, and convergence, clonality analysis, overlap analysis, segment usage analysis, diversity analysis, and motif analysis. In each analysis type, metrics, visualization, and statistical analysis are provided, which offers a comprehensive solution to TCR analysis.

The existing gap between traditional biomedical research and bioinformatics provides a substantial barrier for biomedical researchers to utilize computational tools to analyze high throughput data. Our tool will illustrate the capacities of cloud-based notebooks as the solution to bridge the gap, where users with no to limited bioinformatics background or experience would be able to use notebooks to analyze the data with transparent analysis and reproducible results.

Omics Technologies Posters - Thursday

PB3116. Quantifying regional DNA methylation improves detection of biologically relevant associations in Alzheimer's Disease

Authors:

T. Eulalio¹, D. Nachun², S. Montgomery¹; ¹Stanford Univ., Stanford, CA, ²Stanford Univ., Redwood City, CA

Abstract Body:

DNA methylation is an epigenetic modification linked to numerous neurological diseases, such as Alzheimer's Disease (AD), however, a mechanistic understanding of methylation's role in these diseases remains elusive. DNA methylation at individual CpG sites is often aggregated across functional genomic regions, such as promoters or enhancers, to improve interpretability. Oftentimes, the average methylation value is used for downstream analyses such as detecting differential methylation or unsupervised learning. While this can provide valuable insights, averaging regional methylation to a single value is noisy and further fails to capture more complex methylation patterns, particularly in regions with high CpG density or a mixture of hypo- and hypermethylated CpGs. We hypothesized that these patterns could be better captured by using principal components analysis (PCA) to summarize regional methylation while retaining orthogonal and interpretable methylation signals for downstream analyses. We demonstrated the utility of this approach in the Religious Order Study/Memory and Aging Project (ROSMAP) dataset, a large dataset with methylation collected from the dorsolateral prefrontal cortex (DLPFC) of AD patients and controls. In addition to analyzing the bulk tissue data, we used cell type deconvolution to estimate the cell type-specific methylation signal for four major brain cell types; astrocytes, endothelial cells, neurons, and oligodendrocytes/oligodendrocyte progenitor cells. We compared the performance of the regional aggregation of CpGs with principal components to averaging in identifying differentially methylated genes (DMGs) and pathway enrichment of DMGs for AD-relevant biology. Regional PCA identified over ten times more differentially methylated genes than averaging in the bulk data and up to four times more in the cell type-specific datasets. The genes identified by regional PCA included over 159 AD-relevant genes found in the Open Targets database (overall association score > 0.11, 95th percentile) that were not identified using averages; including the highly-evidenced AD genes SORL1, ABCA7, and TREM2. In gene set enrichment analysis of DMGs in bulk tissues and cell type-specific data, differentially methylated genes identified by regional PCA were significantly enriched in up to nine times more Alzheimer's Disease terms than genes identified by averaging. In conclusion, we show that summarizing DNA methylation within genomic regions using PCA captured more biologically relevant signals than averaging and improves our understanding of the role of methylation in disease.

Omics Technologies Posters - Wednesday

PB3117*. Rapid whole genome and whole exome variant detection using a novel fluorescently labeled reversible terminated nucleotide sequencing system and GPU-based accelerated analysis

Authors:

T. Looney¹, K. Gouin¹, A. Tong¹, M. Fabani¹, A. Sethia², R. Shultzaberger¹, M. Samadi³, T. Zhu³, E. Glezer⁴; ¹Singular Genomics, San Diego, CA, ²NVIDIA, Santa Clara, CA, ³NVIDIA, San Jose, CA, ⁴La Jolla, CA

Abstract Body:

Background Next generation sequencing (NGS) has become an indispensable tool for the diagnosis of genetic disease, though there remains a need to reduce turnaround for time sensitive applications such as newborn genetic screening. Reducing turnaround requires faster sequencing and accelerated data analysis. We recently introduced the Singular Genomics G4 platform for rapid sequencing-by-synthesis (SBS), which can deliver four human whole genomes at ~30x coverage in under 19 hours. Here we present accelerated pipelines for whole genome and whole exome germline variant detection on the G4 that leverage the NVIDIA Clara Parabricks platform and custom DeepVariant models.

Methods The G4 sequencer employs a novel fluorescently labeled reversible terminated nucleotide chemistry in combination with a rapid acquisition camera to deliver a cycle time of ~2.5 minutes. To maximize speed and performance for germline variant detection, we trained custom DeepVariant models for exome and whole genome analysis using a warm start from the Illumina WGS or WES models, respectively, and data from genome in a bottle samples HG002-6 (2x150bp reads; multiple library preparation kits used in generation of exome data). We iteratively explored the DeepVariant parameters `alt_align_pileup`, `vsc_min_fraction_snps` and `vsc_min_fraction_indels` before validating performance on HG001. Finally, the validated models were adapted to run on the GPU-based NVIDIA Clara Parabricks platform for accelerated analysis.

Results The baseline DeepVariant Illumina whole exome model delivered a precision and recall of 99.23% and 98.36% for SNPs, and 97.73% and 92.02% for indels, respectively for HG001 at 100x coverage. The trained model showed improved performance for both SNPs and indels, with a precision and recall of 99.53% and 98.53% for SNPs, and 97.76% and 93.09% for indels, respectively from the same library. A similar improvement in performance was observed for the trained whole genome model. The models were adapted to Parabricks to deliver a fastq-to-vcf turnaround of 29 and 4 minutes for 30x whole genome and 100x whole exome analysis, respectively, on an 8 GPU AWS p4d.24xlarge instance.

Conclusions We have successfully trained and implemented accelerated DeepVariant whole genome and whole exome models for the G4, resulting in improved performance for both SNP and indel detection and a speed matching that of the fastest analysis solutions to date. We anticipate that the combination of rapid-SBS and GPU-based acceleration will significantly reduce turnaround for the most time sensitive NGS applications.

Omics Technologies Posters - Thursday

PB3118. RedSpa: a method for Reference-based domain detection in multiple-sample Spatial transcriptomics studies

Authors:

Y. Ma¹, X. Zhou²; ¹Univ. of Michigan, Ann Arbor, MI, ²Univ MICHIGAN, Ann Arbor, MI

Abstract Body:

Various spatially resolved transcriptomic technologies have enabled the characterization of tissue-specific transcriptomic landscapes, i.e., spatial domain detection. However, existing methods cannot simultaneously analyze multiple tissue slices to disentangle both the cell type compositions and the complex tissue architectures. Moreover, all of the standard domain detection methods perform dimension reduction on the gene expression matrix before domain detection, in which the low-dimensional embeddings may not be optimal. None of the current methods utilizes the information from well-established single-cell RNA-seq technologies to facilitate domain detection. Finally, the published methods can't deal with high-resolution large-scale spatial transcriptomics datasets with more than 100k measured spatial locations. To fill these gaps, we developed RedSpa, which leverages cell type specific gene expression information from single-cell RNA-seq (scRNA-seq) to detect spatial domains in multiple tissue slices. RedSpa iteratively optimizes the cell type composition matrix and spatial domain labels across tissue slices, while accounting for the within-slice and between-slice compositional similarities to ensure optimal clustering performance. We demonstrate the advantages of RedSpa through in-depth analysis of five spatial transcriptomics datasets generated from different technologies, species, and tissues. In real data applications, RedSpa achieves up to 78% clustering accuracy gain over existing domain detection methods. The accuracy gain brought by RedSpa allows us to reveal the transcriptomic landscape in complex tissues including the human prefrontal cortex, squamous cell carcinoma, spermatogenesis, mouse brain, and olfactory bulb.

Omics Technologies Posters - Wednesday

PB3119. Regulome-wide association study identifies chromatin accessibility associated with pancreatic cancer risk

Authors:

S. Liu¹, H. Zhong¹, J. Zhu¹, Y. Wu², L. Wu¹; ¹Cancer Epidemiology Div., Population Sci. in the Pacific Program, Univ. of Hawaii Cancer Ctr., Univ. of Hawaii at Manoa, Honolulu, HI, ²Div. of Cancer Res. and Training, Dept. of Internal Med., Charles Drew Univ. of Med. and Sci., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract Body:

Background: Genome-wide association studies (GWAS) have reported dozens of genetic variants associated with risk of pancreatic cancer (PC), a deadly malignancy. However, a large majority of the identified variants lie in non-coding genomic regions and do not directly impact the coding sequence of genes, which retard the transformation of GWAS findings into clinical application. The assay for transposase-accessible chromatin using sequencing (ATAC-seq) could quantify DNA accessibility with the use of hyperactive mutant Tn5 Transposase that inserts sequencing adapters at the accessible chromatin regions, which theoretically enables investigation of transcription factor (TF) binding. Allele-specific Regulome-wide association study (RWAS) is a newly developed design to identify chromatin accessibility that is potentially associated with diseases. However, their potential link with PC remains elusive. **Methods:** The pan-cancer chromatin accessibility genetic prediction models were built using LASSO penalized regression (for variants in the cis-locus) and top1 (single best variant in the cis-locus) methods, by leveraging allelic, total, or combined signal. We evaluated associations between genetically predicted levels of chromatin accessibility and PC risk by investigating 8,280 cases and 6,728 controls included in the PanScan (I, II, and III) and PanC4 consortia. The correlations between ATAC-seq accessibility and gene expression across 373 samples were assessed to predict the links between the accessibility features and regulated genes. **Results:** After analyzing 561,208 developed pan-cancer RWAS models, 50 accessibility features were identified to be associated with PC risk at a false discovery rate (FDR) of <0.05. These include 24 at seven novel loci not reported in PC GWAS. Ten accessibility features (located in promoter, intron, exon, and distal genomic regions) were identified to be linked with regulated genes at FDR <0.01. Four linked genes were protein-coding genes, including *GSDMB*, *CHST5*, *LDHD*, and *PGAP3*. Of them, *GSDMB* and *PGAP3* had been reported in our previous transcriptomic-wide association study. Five other linked genes are categorized into lincRNA, miscRNA, antisense, or TEC protein-coding genes. **Conclusion:** Our study identified accessibility features in non-coding genomic regions associated with PC risk and predicted linked genes regulated by these features. Our findings provide new insights into the regulatory mechanisms of non-coding regions for pancreatic tumorigenesis.

Omics Technologies Posters - Thursday

PB3120. Revealing the sequence of neutralizing antibodies produced during infection using VyCAP B cell screening equipment coupled to an optimized single cell RNA sequencing protocol.

Authors:

K. Rubben, A-S. Vander Plaetsen, D. Deforce, F. Van Nieuwerburgh; Ghent Univ. - Lab. of Pharmaceutical Biotechnology, Ghent, Belgium

Abstract Body:

Since the sudden emergence and worldwide spread of the highly contagious SARS-CoV-2 almost three years ago, the rapid availability of new therapies has shown to be of utmost importance to lower morbidity and mortality levels. Besides the use of vaccines as a preventive measure, monoclonal antibodies (mAb) constitute a possible therapeutic approach against infections. Several methods for producing monoclonal antibodies already exist and have been used in the past. However, these methods all suffer from certain drawbacks inherent to the specific workflow that is used. Therefore, we are optimizing a novel workflow that aims to reveal the humoral B cell immune response during an infection in a fast and efficient way. The proposed workflow includes the use of a specifically designed filter unit to obtain single B cells in a microwell chip containing 6400 wells, a clamp unit to analyse their antibody production capacity on an antigen-coated membrane, and a puncher system to isolate the specific antibody-secreting B cells (ASC) of interest. These ASC are then lysed and the antibody mRNA sequence is determined by performing optimized reverse transcription (RT) and nested PCR protocols using human B cell primer mixes. Finally, these sequences can easily be determined by Sanger sequencing and cloned into a vector, allowing subsequent transfection of a suitable cell line. First, the VyCAP single B cell screening and isolation workflow was optimized to work with ASC from whole blood of patients, as the use of these cells provides challenges for cell viability and retainment of antibody production. Preliminary experimental testing in 24-well plates revealed that enriching CD38⁺ ASC using a negative and subsequent positive MACS selection step, incubating this cell fraction for 68 hours in a medium containing TLR 7/8-L and IFN- α , and detecting the produced antibodies by using both anti-IgA and anti-IgG secondary antibodies, maximized the chances of detecting viable SARS-CoV-2-specific ASC. Second, the cell lysis, RT and nested PCR protocols needed to be optimized to work with single cell input material. This was successfully done using an antibody-producing hybridoma B cell line. Future research will focus on determining the effect of the optimized conditions on ASC incubated in the more stressful microwell chip environment and applying the optimized RT and nested PCR protocols on single SARS-CoV-2-specific ASC. When all these optimizations are carried out successfully, this workflow could be used to reveal the B cell response during any infection and thereby contribute to the search for a suitable therapy against any infectious agent in the future.

Omics Technologies Posters - Wednesday

PB3121. Rigorous benchmarking of HLA callers for RNA sequencing data.

Authors:

R. Ayyala¹, D. Yu¹, S. Sadek², T. Tao³, R. Alomair⁴, S. Knyazev⁵, S. Mangul³; ¹Univ. of Southern California, Los Angeles, CA, ²California State Univ., Fullerton, Fullerton, CA, ³Univ. of Southern California, Sch. of Pharmacy, Los Angeles, CA, ⁴Univ. of California, Los Angeles, Los Angeles, CA, ⁵Univ. of California, Los Angeles, David Geffen Sch. of Med., Los Angeles, CA

Abstract Body:

Precise identification of alleles in the human leukocyte antigen (HLA) region of the genome is crucial for various clinical and research applications. HLA typing in the clinical laboratory setting is challenging due to its resource intensiveness and the high polymorphism of HLA residues within the population. To address this limitation, many computational tools have been developed to use widely accessible RNA-seq data to predict HLA types.

However, existing RNA-Seq-based benchmarking studies are not up to date, lacking to investigate newly released HLA callers. Previous studies also do not explore the effects of the IPD-IMGT/HLA reference versions, sequencing parameters (i.e., read length and coverage), and the ancestral diversity of the HLA region on prediction quality, while using large-scale and realistic gold standards.

As such, we rigorously compared the performance of 12 HLA callers on 445 RNA-seq samples across 6 cohorts in reference to their respective gold standard datasets. In each case, we produced evaluation metrics of accuracy, that is the percentage of correctly predicted alleles to determine the best performing tool. At the clinically relevant resolution (Four Digits), OptiType had the highest accuracy at 94% with seq2HLA following at a 92% accuracy. Moreover, all HLA caller outputs for Class 1 Genes were separated by HLA allele to determine the number of miscalled alleles. HLA Allele B was miscalled the most while HLA Allele A had the least amount of miscalls across all callers. In addition, CPU and memory metrics were collected for each tool, in which seq2HLA used the least amount of CPU and memory relative to all other tools. This study offers crucial information for researchers regarding appropriate choices of methods for an HLA analysis.

Omics Technologies Posters - Thursday

PB3122. rMATS-turbo: An efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data.

Authors:

J. Adams^{1,2}, Y. Wang^{1,3}, Z. Xie¹, E. Kutschera¹, K. Kadash-Edmondson¹, Y. Xing^{1,2}; ¹The Children's Hosp. of Philadelphia, Philadelphia, PA, ²Univ. of Pennsylvania, Philadelphia, PA, ³Univ. of California, Los Angeles, Los Angeles, CA

Abstract Body:

Pre-mRNA alternative splicing is a prevalent mechanism for diversifying eukaryotic transcriptomes and proteomes. Regulated alternative splicing plays a role in many biological processes, and dysregulated alternative splicing is a feature of many human diseases. Short-read RNA sequencing (RNA-seq) is now the standard approach for transcriptome-wide analysis of alternative splicing. Since 2011, our lab has developed and maintained rMATS, a computational tool for discovering and quantifying alternative splicing events from RNA-seq data. The rMATS software has been widely used by the research community. Here we describe the contemporary version of rMATS - called rMATS-turbo - a fast and scalable re-implementation that maintains the statistical framework and user interface of the original rMATS software, while incorporating a revamped computational workflow with substantial improvements in speed and data storage efficiency. The rMATS-turbo software scales up to massive RNA-seq datasets with tens of thousands of samples. To illustrate the utility of rMATS-turbo, we describe two representative application scenarios. Firstly, we describe a broadly applicable two-group comparison to identify differential alternative splicing events between two sample groups, including both annotated and novel alternative splicing events. Secondly, we describe a quantitative analysis of alternative splicing in a large-scale RNA-seq dataset (~1,000 samples), including the discovery of alternative splicing events associated with distinct cell states. We detail the workflow and features of rMATS-turbo that enable efficient parallel processing and analysis of large-scale RNA-seq datasets on a compute cluster. We anticipate that rMATS-turbo will be useful for studying alternative splicing in diverse biological systems.

Omics Technologies Posters - Wednesday

PB3123. RNA monsters generated by human mitochondrial RNA polymerase

Authors:

E. Greenwald¹, J. Park², N. Jain¹, I. Zheludev¹, K. Artiles¹, D-E. Jeong¹, W. Yin², A. Fire¹; ¹Stanford Univ., Palo Alto, CA, ²Univ. of Texas Med. Branch at Galveston, Galveston, TX

Abstract Body:

Human hepatitis delta virus and plant viroids are RNA pathogens that are copied by host DNA-dependent RNA polymerases (RNAPs) (Cunha et al., 2015. World J Virol.). RNA replication by transcription polymerases, classically perceived as DNA-dependent, could provide a means to copy and transmit RNA-based genetic material. Former studies found that the DNA-dependent RNA polymerase from bacteriophage T7 can regenerate certain RNAs, providing a model system for RNA replication by transcription polymerases (Jain et al., 2020. Science). Do similar transcription RNA polymerases also have RNA-templated replication activity? We performed in vitro reactions using human and yeast mitochondrial RNAP, which have structural similarity to T7 RNAP. Thus far, we have shown that human and yeast mitochondrial RNAP do yield RNA in in vitro no-template-added reactions. Sequencing of these reaction products shows strings of As interrupted by other bases, and these resulting sequences are not homologous to any known sequences. Further experiments are underway with a mix of DNAs as possible templates, with less As in the reaction, and with different sequencing methods to further understand the RNAs that mitochondrial RNAPs tend to make. The RNA-templated RNA replication and untemplated transcription activities by transcription polymerases challenge the textbook view of transcription. In the case of mitochondrial polymerase, it is intriguing that transcription by this enzyme could serve as a flexible sensor of both template sequence and biochemical environment. This provides a fundamental insight into the evolution in an RNA world that could indicate how RNA-based biology could respond to environmental conditions.

Omics Technologies Posters - Thursday

PB3124*. Robust mapping of cell states from multiomics data using implicit feature selection.

Authors:

H. Hu¹, Y. Choi¹, G. Quon²; ¹Univ. of California Davis, Davis, CA, ²Univ California Davis, Davis, CA

Abstract Body:

scRNAseq and scATACseq assays measure distinct aspects of gene regulation, and both are routinely used to infer cell types and states of individual cells. When the same cell types are profiled using both technologies, some cell types distinguishable in RNA space are not distinguishable in ATAC space, and vice versa. The introduction of multi-omics assays such as SHARE-seq that measure both data modalities (RNA, ATAC) in the same cell provide an opportunity to directly measure the extent to which the RNA and ATAC spaces can be mapped to each other, and therefore provide insight into the extent to which variation in open chromatin regions can explain variation in gene expression.

A critical step in mapping RNA and ATAC state spaces is feature selection: both the genes and the chromatin regions are typically subsetted based on variability before cell state mapping is performed. This is problematic because the high variable features may not be correlated across modalities and modality-specific cell type identifications can be nonaligned, leading to poor mapping.

Here we show that by training a deep neural network that simultaneously selects features and maps between RNA and ATAC state spaces, we substantially improve our ability to map cell state spaces. To do so, we present a novel supervised framework designed for jointly capturing the associated regions and genes by leveraging paired multi-omics data without pre feature selection. Our framework leads to improved prediction of gene expression from chromatin accessibility in 70-78% of the total genes, across data from diverse assays (SNARE-seq, SHARE-seq, 10x Multiome), and also when predicting chromatin accessibility from gene expression. We directly attribute this performance gain to the lack of explicit feature selection, compared to existing approaches. On the SNARE-seq cortex data for example, among the top 3000 genes that best predict ATAC state space, 753 genes are not selected by feature selection strategies employed by other approaches, illustrating how feature selection can throw away informative genes. Our framework also enables us to augment training data with larger uni-modal datasets available on the same cell types, leading to an improvement in accuracy of over 55% of all chromatin regions compared to when only multi-omic datasets are available.

Finally, we demonstrate the generalizability of our approach to studying multi-modal datasets involving data modalities such as RNA and single neuron electrophysiology. Our approach is therefore useful for establishing the extent to which mappings between e.g. RNA and ATAC cell state spaces can be established.

Omics Technologies Posters - Wednesday

PB3125. SARS-CoV-2 genomic surveillance in Rwanda: Introductions and local transmission of the B.1.617.2/Delta variant of concern.

Authors:

Y. Butera¹, M. Semakula², S. Hong³, N. Bollen⁴, S. Dellicour⁴, M. Artesi⁵, V. Bours⁶, K. Durkin⁵, L. Mutesa⁷, G. Baele⁴; ¹Ctr. for Human Genetics, Univ. of Rwanda, Kigali, Rwanda, ²Ctr. for Statistics, Hasselt Biostatistics and Statistical Bioinformatics Ctr., Diepenbeek, Limburg, Belgium, ³Dept. of Microbiol., Immunology and Transplantation, Rega Inst., KU Leuven, Leuven, Belgium, ⁴Dept. of Microbiol., Immunology and Transplantation, Rega Inst., KU Leuven, Leuven, Belgium, ⁵Lab. of Human Genetics, GIGA Res. Inst., Liège, Belgium, ⁶Lab. of Human Genetics, GIGA Res. Inst., Liège, Belgium, ⁷Univ. of Rwanda, Kigali, Kigali City, Rwanda

Abstract Body:

The emergence of the SARS-CoV-2 Delta variant of concern (lineage B.1.617.2) in late 2020 resulted in a new wave of infections in many countries across the world, where it often became the dominant lineage in a relatively short amount of time. We here report on a novel genomic surveillance effort in Rwanda in the time period from June to September 2021, leading to 201 SARS-CoV-2 genomes being generated, the majority of which were identified as the Delta variant of concern. We show that in Rwanda, the Delta variant almost completely replaced the previously dominant A.23.1 and B.1.351 (Beta) lineages in a matter of weeks, and led to a tripling of the total number of COVID-19 infections and COVID-19-related fatalities over the course of only three months. We estimate that Delta in Rwanda had an average growth rate advantage of 0.034 (95% CI 0.025-0.045) per day over A.23.1, and of 0.022 (95% CI 0.012-0.032) over B.1.351. Phylogenetic analysis reveals the presence of at least seven local Delta transmission clusters, with two of these clusters occurring close to the border with the Democratic Republic of the Congo, and another cluster close to the border with Tanzania. A smaller Delta cluster of infections also appeared close to the border with Uganda, illustrating the importance of monitoring cross-border traffic to limit the spread between Rwanda and its neighboring countries. We discuss our findings against a background of increased vaccination efforts in Rwanda, and also discuss a number of breakthrough infections identified during our study. Concluding, our study has added an important collection of data to the available genomes for the Eastern Africa region, with the number of Delta infections close to the border with neighboring countries highlighting the need to further strengthen genomic surveillance in the region to obtain a better understanding of the impact of border crossings on lowering the epidemic curve in Rwanda.

Omics Technologies Posters - Thursday

PB3126. scNanoGPS enables high throughput single cell nanopore sequencing of same cell mega-omics in human tumors

Authors:

C-K. Shiau; Northwestern Univ., Chicago, IL

Abstract Body:

Long read single cell nanopore RNA sequencing (scNanoRNAseq) is emerging as a powerful technology to simultaneously profile phenotypes and genotypes of same cells, which however are challenged by lacking robust computational tools and requiring parallel short reads to curate sequencing errors. To address these challenges, we developed a computational tool, scNanoGPS to deconvolute barcoded long reads into single cells and single molecules without short reads curation and calculate both phenotypes (gene expression, isoform) and genotypes (mutation, fusion) of same cells. We applied scNanoGPS onto 24,000 single cells of 4 tumors and 3 cell lines. Our results showed that scNanoGPS can robustly detect single cell mega-omics to delineate cell types and cell states as well as mutation clonality. In tumor samples, scNanoGPS dissected tumor microenvironment and aligned tumor developmental hierarchies onto evolutionary lineages. In summary, scNanoGPS enabled high-throughput single cell mega-omics analysis with single cell third generation sequencing technologies.

Omics Technologies Posters - Wednesday

PB3127. Searching for Somatic Variation in Rapid-onset Obesity with Hypothalamic dysfunction, Hypoventilation and Autonomic Dysregulation (ROHHAD)

Authors:

S. Barclay^{1,2}, C. M. Rand³, M. Khanbabaei^{1,2}, N. T. Bech-Hansen^{1,2}, D. E. Weese-Mayer^{3,4}, K. C. Kurek^{1,2}; ¹Univ. of Calgary, Calgary, AB, Canada, ²Alberta Children's Hosp. Res. Inst., Calgary, AB, Canada, ³Ann & Robert H. Lurie Children's Hosp. of Chicago, Dept. of Pediatrics, Div. of Autonomic Med. and Stanley Manne Children's Res. Inst., Chicago, IL, ⁴Northwestern Univ. Feinberg Sch. of Med., Chicago, IL

Abstract Body:

Classically, the syndromes associated with somatic genetic variation have been those in which the tissues harbouring the pathogenic change are visibly altered from the tissues around them. More recently, this group has grown to include less 'visible' diseases, such as epilepsy and immune disorders. ROHHAD is a rare pediatric disorder in which healthy children between the ages of 1.5-7 years present with signs of hypothalamic dysfunction and dysregulation of the autonomic nervous system, that include rapid-onset obesity and potentially fatal central alveolar hypoventilation. ROHHAD is part of a group of disorders of respiratory control and autonomic nervous system dysregulation, some of which have genetic causes (i.e., CCHS, caused by *PHOX2B* variants). Extensive genetic investigations, including candidate gene, exome, and genome sequencing studies, have failed to identify a recurrent germline genetic explanation in any ROHHAD cohort. This fact, combined with the observations that a) about 40-50% of individuals with ROHHAD have tumours of neural crest origin, and b) our cohort contains two pairs of monozygotic twins discordant for ROHHAD, have led us to consider the possibility that ROHHAD pathogenesis may involve somatic changes. To investigate this hypothesis, we undertook whole genome sequencing, at an average of 100X depth, using blood, of the discordant monozygotic twin pairs. We identified only one discordant protein-altering variant in one of the twin pairs: an *FMR1* variant that is synonymous in the major isoforms of FMRP, but results in a missense change in the nuclear-localized isoforms ISO6 (c.1319C>T, p.(T440M)) and ISO12 (c.1256C>T, p.(T419M)). The variant was detected on 5/103 reads in the proband; 0/118 in the unaffected twin; 0/130 in the mother; and 0/50 in the father. We validated this finding using digital PCR, a variant detection method sensitive to 0.001%, which showed that the variant was present at 3% and 5% in a blood and tumour sample from the proband, respectively, and 0% in blood samples from the unaffected twin and parents. Whether this finding in a single ROHHAD proband has any relevance to the pathogenesis of ROHHAD is yet unknown. It has illustrated the potential of high-throughput sequencing to detect low level somatic variation, but points to the fact that increased depth of sequencing (even higher than 100X) would be required to have confidence that all variants at a frequency of 5% or lower will be detectable. We have recently had the privilege of receiving autopsy tissue donated by a family who lost a child to ROHHAD, and are in the process of investigating directly the tissues thought to be *affected* - the hypothalamus and brainstem.

Omics Technologies Posters - Thursday

PB3128. SeeNV: a pipeline ready tool for the integration and visualization of CNVs and their callers

Authors:

M. Bradshaw; Univ. of Colorado Boulder, Boulder, CO

Abstract Body:

Copy number variants (CNVs) can affect health and physiology by altering gene expression, manipulating epigenetic factors, and/or gene expression. CNV identification, or calling, is now common practice in genetic diagnostic laboratories. To maximize sensitivity, it is now an established best practice to use an ensemble of callers. However, the various CNV calling programs utilize different algorithms, parameters, and inputs. These differences can result in inconsistencies making it difficult to discern whether or not a call is a true positive, which necessitates manual curation by a clinician to determine which calls are both real and relevant.

To aid in this task of manual curation we present SeeNV, a tool to help balance the clinical need for high diagnostic yield without overwhelming clinicians with dozens of false positives. SeeNV provides means to simultaneously visualize and integrate multiple algorithms, parameter sets, lab specific reference cohorts and large public datasets. By comparing calls to those from the same lab, it can be determined if a call is real or an artifact of lab specific sample preparation and processing. Integration of large public datasets such as 1000 Genomes Project and varDB, make it possible to establish the population frequency, indicative of its relevance or deleterious effect. SeeNV provides a series of local and chromosomal visualizations. Local plots depict statistics such as read coverage and population frequency in the area immediately surrounding a call. Chromosomal scale plots help visualize the amount of noise present within a sample and how much noise is in the region containing the call.

SeeNV is already in use as part of a clinical pipeline at Children's Hospital Colorado; we have also generalized it with default parameters based on publicly available data so it can be used as a drop-in tool for labs with limited technical resources. The robust features comprising SeeNV are bundled into a simple command line tool with single step installation and a single command to run it with built in parallelization. SeeNV is freely available for download on GitHub:

<https://github.com/MSBradshaw/SeeNV>

Omics Technologies Posters - Wednesday

PB3129. Sentieon DNAscope LongRead: a highly accurate, fast, and efficient pipeline for germline variant calling from PacBio HiFi reads.

Authors:

D. Freed¹, W. J. Rowell², A. M. Wenger², Z. Li¹; ¹Sentieon Inc, San Jose, CA, ²Pacific BioSci.s, Menlo Park, CA

Abstract Body:

PacBio HiFi sequencing leverages the random nature of errors in the underlying technology to generate highly accurate consensus reads that are both long and highly accurate, with read lengths greater than 10kb and average base accuracy exceeding 99.8%. Sentieon develops fast, efficient and accurate software for processing genomic data, including improved implementations of the GATK best practices pipelines for germline and somatic variant calling. To further improve germline variant calling accuracy beyond the GATK, Sentieon has developed DNAscope and has used DNAscope to successfully participate in and win awards in multiple precisionFDA (pFDA) challenges, including the original pFDA Truth Challenge and the pFDA Truth Challenge V2.

Here, we present DNAscope LongRead, a highly accurate pipeline for germline variant calling from PacBio HiFi reads that is an update and extension of our earlier DNAscope tool for germline variant calling from short reads. We benchmark DNAscope LongRead using samples from the pFDA Truth Challenge V2 and evaluated variant calling performance using the Genome in a Bottle v4.2.1 benchmark dataset. Compared to the pFDA Truth Challenge V2 winning submission, DNAscope LongRead reduces overall sequencing errors by 15%, with an average of 9,130 errors (mean F1-score of 0.9988) in the pFDA samples relative to the v4.2.1 benchmark.

DNAscope LongRead uses a machine learning model that was trained using samples sequenced using PacBio's chemistry 2.0 or earlier and evaluated against the Genome in a Bottle v4.2.1 benchmarks. To test the ability of DNAscope LongRead to generalize, we performed variant calling on a sample sequenced with PacBio's chemistry 2.2 on a Sequel II system. With this sample, DNAscope LongRead made 9,231 errors, similar to the accuracy obtained with earlier sequencing chemistries. Variant calling accuracy is also high across the Challenging Medically Relevant Genes benchmark, demonstrating the tool's ability to extend to more difficult benchmark regions. DNAscope LongRead was able to perform variant calling in under 5 hours on an Intel Xeon machine with 32 virtual CPUs for each sample tested and used less than 16GB of memory. These results demonstrate that DNAscope LongRead is a fast and accurate tool for calling germline variants from PacBio HiFi reads.

Omics Technologies Posters - Thursday

PB3130. Sentieon DNAscope: high accuracy small variant calling using machine learning.

Authors:

B. Gallagher; Sentieon Inc, San Jose, CA

Abstract Body:

Sentieon develops fast, efficient, and accurate software for processing genomic data, including improved implementations of the GATK best practices pipelines for germline and somatic variant calling. To further improve germline variant calling accuracy beyond the GATK, Sentieon has developed DNAscope and has used DNAscope to successfully participate in and win awards in multiple precisionFDA (pFDA) challenges, including the original pFDA Truth Challenge and the precisionFDA Truth Challenge V2. We present DNAscope, an accurate and efficient short-read germline small-variant caller. DNAscope combines the robust and well-established preprocessing and assembly mathematics of the GATK's HaplotypeCaller with a machine-learned genotyping model. Benchmarks of DNAscope and DNaseq (Sentieon's GATK-matching germline variant calling pipeline) demonstrate that DNAscope achieves superior SNP and insertion/deletion accuracy with reduced computational cost. DNAscope's machine-learning genotyping model is tuned for established and newly developed short read sequencing platforms including Illumina, MGI/BGI, Element Biosciences, and Ultima Genomics.

For a 30x HG002 sample sequenced on an Illumina platform DNAscope achieves an F1-score of 99.57% for SNPs and 99.46% for INDELS on the Genome in a Bottle v4.2.1 benchmark dataset. Notably, DNAscope reduces total errors by more than two-fold (from 87,607 to 34,634) relative to GATK. The high accuracy of DNAscope across multiple samples suggests that DNAscope's model is not overfit to the HG001 and HG005 samples used during model training. DNAscope models have been successfully trained for the Illumina, MGI, Element Bioscience and Ultima Genomics platforms, and produce state-of-the-art variant calling accuracy on each platform. DNAscope software can be run on x86 or ARM based architecture and can complete alignment and variant calling for a 30x WGS in less than 1 hour for less than \$2 on AWS.

Omics Technologies Posters - Wednesday

PB3131. Sequence-to-expression models of compact promoters for cell-type-specific promoter design

Authors:

A. Reddy, M. Herschl, A. Lu, A. Kumar, X. Geng, P. Hsu, S. Levine, N. Ioannidis; Univ. of California, Berkeley, Berkeley, CA

Abstract Body:

Advances in gene delivery technologies are enabling rapid progress in molecular medicines. Achieving the full potential of gene and cell therapies requires the precise expression of genetic cargo in desired cell types. However, current approaches are limited by the small number of experimentally demonstrated cell type-specific gene regulatory elements. It would be very useful to have promoter sequences of minimal length to enable facile delivery, capable of promoting expression in particular target cell types while silencing expression in other cell types.

Designing these promoters is difficult as it is infeasible to experimentally validate every possible sequence. Thus, we propose a data-driven machine learning-based approach to design promoters. Deep learning (DL) models can accurately predict gene expression using only the surrounding sequence by learning sequence features such as the presence and strength of transcription factor (TF) binding motifs. We can use these sequence-to-expression models to test sequences in-silico. However, DL models are prone to overfitting and produce unreliable results when tested using inputs that are very different from the training inputs. Model-based optimization (MBO) algorithms can combat these problems and reliably utilize these models to generate desirable promoters. Therefore, we are developing DL models of cell-type-specific expression from short promoter sequences that will be used together with state-of-the-art MBO algorithms to design promoters for gene therapy with improved cell-type specificity.

Most existing sequence-to-expression models predict expression using a large sequence input context to capture the effects of nearby cis-regulatory elements. However, it is challenging for compact gene therapy designs to incorporate distal regulatory elements, instead relying on TF binding within a short promoter sequence. Here, we train DL models to predict gene expression from short input sequences in a variety of cell types, using both MPRA and RNA-seq data for training. We evaluate multiple architectures and training strategies using the MPRA promoter sequence or short windows of endogenous gene promoters for the RNA-seq data. To demonstrate the use of these models in promoter optimization, we focus on three leukemia cell lines with the aim of designing promoters to target high expression in each cell line individually while minimizing expression in the other two. We are generating experimental data to validate the performance of our models and to train a second round of models optimized for these cell lines.

Omics Technologies Posters - Thursday

PB3132. Sequencing By Binding (SBB) demonstrates superior performance in low-pass whole-human-sequencing applications

Authors:

K-Y. Chen¹, A. Torkamani²; ¹The Scripps Res. Inst., La Jolla, CA, ²Scripps Res. Translational Inst., La Jolla, CA

Abstract Body:

Recently, PacBio introduced its sequencing by binding (SBB) chemistry, which provides an order-of-magnitude higher accuracy than the Illumina sequencing by synthesis (SBS) technology. In this study, the performance of Illumina SBS and PacBio SBB is compared on HG002 relative to the Genome in a Bottle benchmark for both high-coverage and low-coverage sequencing data. A HG002 WGS library was sequenced by PacBio SBB to 40x coverage. An SBS dataset for comparison was obtained from the Genome in a Bottle website. Deduplication and base quality score recalibration were conducted on SBS data. We then performed variant calling for single nucleotide variants (SNVs) and insertions and deletions (indels). Genotype imputation and variant evaluation are exercised for performance comparison. PacBio SBB achieved higher true-positive rate and lower false-negative rates in both SNVs and indels compared to SBS. Additionally, PacBio SBB provided a higher genotype concordance rate than SBS compared to benchmark SNVs and indels. Considering transition/transversion ratio (ti/tv) and heterozygous/non-reference homozygous ratio (het/hom ratio) as measurement of call quality, ti/tv and het/hom ratio generated by PacBio SBB were closer to the ones generated from benchmark calls. To access call quality, we evaluated variant calls in both platforms without imputation process across low-coverage sequencing data. For SNV calls, the results showed that SBB achieved a higher genotype concordance rate compared to SBS. In addition, after the imputation process, we found that SBB also offers higher imputation accuracy outcome across low-coverage sequencing data. PacBio SBB delivered high-quality outcomes in both high-coverage and low-coverage sequencing data compared to Illumina. These outcomes suggest that the performance of PacBio SBB exceeds well-established Illumina SBS platform and demonstrates its potential application in the field of genomic studies.

Omics Technologies Posters - Wednesday

PB3133. Sequencing By Binding (SBB) enables a lower limit of detection for Tuberculosis resistance genes *gyrA* and *katG*.

Authors:

C. Wike¹, C. Allender², N. Pezeshkian³, D. Nasko⁴, S. Pond¹, E. Engelthaler²; ¹Translational Genomics Res. Inst., Phoenix, AZ, ²Translational Genomics Res. Inst., Flagstaff, AZ, ³Pacific BioSci.s, Menlo Park, CA, ⁴Pacific, Menlo Park, CA

Abstract Body:

Tuberculosis (TB) remains a leading infectious disease due to many challenges associated with treating *Mycobacterium tuberculosis* (Mtb), including its ability to develop drug resistant mutations. Identifying the correct antibiotics to treat individuals with TB is critical and difficult to do without the ability to screen for certain drug resistance mutations in Mtb. Next generation sequencing (NGS) has enabled the detection of these genetic variants, specifically in genes *katG* and *gyrA*, which have been associated with reduced sensitivity to first-line and second-line treatments, respectively. However, these genes are encoded in GC-rich regions (~60-80%) that are challenging for current Sequencing By Synthesis (SBS) NGS techniques to sequence. We demonstrate that revolutionary sequencing-by-binding (SBB) chemistry can reliably sequence Mtb drug resistance mutations in *gyrA* and *katG* genes, despite their high GC content. Additionally, the improved accuracy of SBB enables a new lower limit of detection (0.001%) that is two orders of magnitude lower than current SBS sequencing platforms. Lower limits of detection along with rapid screening of resistance genes will enable earlier and more precise treatment of TB.

Omics Technologies Posters - Thursday

PB3134. *SigAlign* (Similarity-guided Align): A new, high-performance generalized pairwise alignment algorithm using simplified and intuitive regularization criteria

Authors:

K. Bahk, J. Sung; Genome Epidemiology & Hlth.Data Lab., Graduate Sch. of Publ. Hlth., Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract Body:

BACKGROUND: Quantifying similarities between two nucleotide sequences, *a.k.a.* "pairwise alignment algorithm (PAA)," is a core but computationally expensive task in bioinformatics. A method to reasonably reduce the pairwise comparison is the key to this algorithm. Current PAA methods generally rely on the "affine-penalty function", which inevitably introduce multiple criteria that lack biological connections such as the "size of k-mers" or "X-drop value". Here, we propose a new alignment algorithm, "*SigAlign*" (Similarity-guided Align). We attempted to demonstrate that the new align algorithm is more reliable and faster than current standard methods, allowing flexible adjustment for the characteristics of data.**METHODS:** Multiple new methods constitute the *SigAlign* algorithm: first, automatically determine the optimal size of *k-mers*; second, search the exact matches or "seeds" using automatic and optimized methods in the reference; third, start outward PAA on both ends of the seeds until violating the new criteria. *SigAlign* uses two intuitive and simplified criteria to determine the extension/termination of PAA - 1) minimum alignment length (MAL) and 2) maximum penalty per length (MPpL). We compared the performance of our method with BWA-mem, Hisat2, and Bowtie2. We compared the throughput, alignment score and length using *Mycobacterium tuberculosis* and the human genome samples. We also performed a "recall test" to locate the exact match in the reference mixed with highly similar sequences (UHGG, unified human gastrointestinal genome), using simulated short-reads from the Illumina platforms.**RESULTS AND INTERPRETATIONS:** The proposed algorithm worked faster with higher alignment scores than all BWA-mem, Hisat2, and Bowtie2 at the same default penalty schemes with respective tools; up to 3.4, 1.2, and 9 times faster. The recall rate is near complete (~100%) for the *SigAlign*, whereas the rate of the other three tools ranged from 20 to 50%. We believe a new alignment algorithm based on simplified but adjustable criteria that bear biological connotations (MAL and MPpL). We expect the *SigAlign* would increase the alignment performance with the economy. The new method may need further validation on diverse, real-life tasks.

Omics Technologies Posters - Wednesday

PB3135. Simplified RNA-Seq Library Prep: Improved RNA sequencing results for FFPE and blood samples

Authors:

E. Putnam, S. Panja, S. Chung, B. Komorous, M. Ait Ichou; Quantabio, Beverly, MA

Abstract Body:

RNA-seq studies carried out using high-throughput sequencing of cDNA have provided tremendous insight into cellular transcript studies on a large and comprehensive scale. However, technical challenges such as laborious and lengthy workflows, affordability, compromised accuracy, read coverage biases and limited transcript diversity have impeded implementation of the technology in many labs. There are additional challenges when working with RNA derived from FFPE and blood, which are commonly used in precision medicine research. FFPE RNA is usually degraded and consists of small sized fragments. Blood RNA has high abundance of globin mRNA that could potentially mask the reads from regions of interest by saturating the sequencing runs.

Here we present a simple, affordable, high-performance solution for directional RNA-seq library preparation: the sparQ RNA-Seq HMR Kit. The kit integrates depletion of ribosomal and globin transcripts (human, mouse, and rat) and RNA fragmentation into a single step and tube. The proprietary, highly optimized enzymes and streamlined workflow generates high quality, directional, transcriptome NGS libraries from either intact or degraded RNA samples, with key improvements for FFPE and blood samples.

Utilizing RNA derived from FFPE and blood, we evaluated libraries prepared using sparQ RNA-Seq HMR Kit. The data gathered from the FFPE samples were compared to other RNA-seq kits. Comparison with 20 million sampled reads from each library demonstrated superior performance, even with limiting quantity: increased unique transcript identification, uniform 5' to 3' transcript coverage, efficient rRNA and globin mRNA removal, higher yield, and excellent data concordance.

The sparQ RNA-Seq HMR Kit with integrated ribo-globin depletion technology, enables faster time to result (5 hours), minimal hands-on time, and fewer pipetting steps, while generating high quality transcriptomic data for FFPE and blood samples.

Omics Technologies Posters - Thursday

PB3136. Simulating pathogenic and likely pathogenic variants for bioinformatics pipeline testing

Authors:

Q. Zeng, B. Huang, P. Hu, N. Russell, S. Letovsky, N. Leach, A. Kenyon; Labcorp, Westborough, MA

Abstract Body:

Next Generation Sequencing (NGS) is widely used to detect clinically significant variants in clinical samples. The analytical performance (sensitivity and specificity) of NGS tests is dependent on many factors, including the selection of an analysis pipeline, targeted genomic regions, variant types and length, library quality, sequencing technology, read length and quality, and coverage. It is critical to understand the impact of these factors in order to optimize testing performance.

The performance of NGS analysis pipelines can be assessed using different reference materials. First, a high-quality reference data set (e.g., Genome in a Bottle Consortium) provides a set of “gold standard” variants in a small number of well-characterized genomes. Second, some public repositories (e.g., Coriell Institute) have a large number of samples with well-characterized individual pathogenic variants. A third source is previously characterized clinical samples with reported pathogenic and likely pathogenic variants (PLP). All of these reference materials can be utilized to assess the sensitivity and specificity of the analysis pipeline, but they are limited in the number of PLP, and therefore, none provides a comprehensive picture of PLP detection across the genome.

Another approach is to supplement clinical samples and gold standard references with simulated samples. However, previous simulation studies have been limited in the number of simulated variants, mostly focusing on specific variant type(s) and variant length range. In this study, simulated samples are generated for all currently available ClinVar PLP, using a streamlined variant simulation framework developed in-house. We analyzed simulated samples with Illumina paired-end reads for over 110,000 PLP variants to assess the performance of our NGS testing analysis pipeline. Our analysis confirmed that the NGS pipeline has high sensitivity for the vast majority of the ClinVar PLP across the different variant types and length range throughout the genome. We also identified a number of challenging ClinVar PLP requiring customized pipeline settings, which further enhanced the performance of our NGS testing analysis pipeline. Our variant simulation framework for all current ClinVar PLP could be readily used by others to evaluate the performance of the NGS testing pipelines.

Omics Technologies Posters - Wednesday

PB3137*. Simultaneous dimensionality reduction and cell-type annotation of single-cell RNA-seq data using marker enriched uniform manifold approximation and projection.

Authors:

A. Khan^{1,2}, A. Sathe³, J. T. Wu³, R. J. Huang^{4,5}, M. R. Amieva^{6,7}, R. Tibshirani^{8,9}, M. H. Gephart^{2,5}, H. Ji^{3,5,10}, S. Han^{1,2,5}; ¹Quantitative Sci. Unit, Dept. of Med., Stanford Univ. Sch. of Med., Stanford, CA, ²Dept. of Neurosurgery, Stanford Univ. Sch. of Med., Stanford, CA, ³Div. of Oncology, Dept. of Med., Stanford Univ. Sch. of Med., Stanford, CA, ⁴Div. of Gastroenterology and Hepatology, Stanford Univ. Sch. of Med., Stanford, CA, ⁵Stanford Cancer Inst., Stanford Univ. Sch. of Med., Stanford, CA, ⁶Dept. of Microbiol. and Immunology, Stanford Univ. Sch. of Med., Stanford, CA, ⁷Dept. of Pediatrics, Stanford Univ. Sch. of Med., Stanford, CA, ⁸Dept. of Statistics, Stanford Univ., Stanford, CA, ⁹Dept. of BioMed. Data Sci., Stanford Univ., Stanford, CA, ¹⁰Stanford Genome Technology Ctr., Stanford Univ., Stanford, CA

Abstract Body:

Background

Cell type identification is a key step in analyzing single-cell RNA-seq data. Existing methods typically take the cluster-then-annotate approach, wherein clustering is performed across cells using principal components that capture the dimensionality reduction of highly variable genes; and cell clusters are visualized through Uniform Manifold Approximation and Projection (UMAP). Cell type annotation is then performed for each cluster utilizing external marker gene information that represents highly expressed genes for each cell type. Thus, cell type annotation vs. dimensionality reduction are considered as independent problems and solved separately. This separation could potentially lead to poor annotation due to discrepancy between clustering and visualization subspaces.

Methods

We propose a novel method that unifies the problems of cell type annotation and dimensionality reduction/projection using marker enriched UMAP. Conventionally, UMAP constructs a cell-cell similarity graph and projects it to a lower dimensional space. In our approach, we directly incorporate marker gene information into UMAP by constructing and combining the following two graphs through weighting. One graph captures cell-cell similarity based on correlation patterns of marker gene expression, while the other graph encodes proximity of cells based on correlation patterns in the rest of highly variable genes. Cell type assignment is computed based on the UMAP projection using the weighted graph. We evaluate the performance of the proposed versus two existing methods, scType and Garnett, using public data for human colorectal cancer (CRC) and melanoma from NCBI Gene Expression Omnibus.

Results

The melanoma data consisted of 2887 cells from 19 melanoma tumors, curated into 6 cell types. The proposed method achieved high overall annotation accuracy (AC) of 0.90 (95% confidence interval [0.89-0.91]) outperforming both scType (AC 0.82; [0.80-0.83]) and Garnett (AC 0.35; [0.33-0.37]). In the largest group of T cells (2068 cells), the main source of misclassification by scType and Garnett occurred from mis-annotating 400 and 496 cells as 'unknown' due to conflicting clusters obtained independently from the annotation step; this reduced cell-type specific sensitivity to 0.76 and 0.15 (vs. 0.87 of the proposed method). Using the CRC data consisting of 364 malignant cells, the proposed method (AC 0.90; [0.86-0.93]) outperformed both scType (AC 0.59; [0.53-0.64]) and Garnett (AC 0.49; [0.44-0.55]).

Conclusion

We proposed a weighted-graph approach to incorporate marker genes into UMAP and demonstrated the potential for improving cell type annotation.

Omics Technologies Posters - Thursday

PB3138*. Single cell dissection of Alzheimer's disease and related dementias.

Authors:

A. Berg¹, B. James¹, N. Sun¹, K. Galani¹, D. Bennett², L-H. Tsai¹, M. Kellis¹; ¹MIT, Cambridge, MA, ²Rush Alzheimer's Disease Ctr., Chicago, IL

Abstract Body:

Alzheimer's Disease and Related Disorders (ADRD) are debilitating neurodegenerative disorders afflicting 47 million individuals globally. They include Alzheimer's Disease (AD), Frontotemporal Dementia (FTD), Lewy Body Disease (LBD), and Vascular Contributions to Cognitive Impairment and Dementia (VCID), which are highly heterogeneous, frequently co-occur, and share numerous clinical, pathological, and molecular manifestations, raising the question as to whether more etiologically-salient subdivisions exist. Previous attempts at cell-resolution molecular profiling and de novo subtyping were limited by the lack of single-cell, multi-omic datasets encompassing multiple ADRD subtypes. To address this challenge, we generated scRNA-seq and scATAC-seq data for 540,000 cells from the prefrontal cortex of 102 patients with AD, FTD, LBD, VCID, and neurotypical diagnoses. We used these data to derive cell type-specific functional descriptions of ADRD pathologies at the transcriptional, epigenomic, and TF-regulatory levels, including TF-regulon activation scores derived by SCENIC. Of all these levels, TF-regulons were particularly distinct between ADRDs and highly interpretable, revealing multiple shared and distinct GO-enrichments between the diseases. We trained a cell-type-specific logistic regression model for disease classification, showing consistently superior performance than gene expression or epigenomic information alone. We used the model coefficients to weight target genes in further biological enrichment analyses, and to deconvolve and identify shared and distinct cell-level regulatory patterns across the diseases, and as a basis for de novo subtyping across the ADRD spectrum. We also used flow, centrality, and community analyses of our TF-regulatory networks to reveal network-level cell-type-specific and disease-specific regulatory disruption. Lastly, we validated our biological enrichments and de novo ADRD subtypes using transcriptional, epigenomic, and phenotypic data. Overall, our cell-resolution profiling across ADRDs demonstrates a novel, algorithmically-amenable and biologically-informed dimensionality reduction technique for single-cell data analysis in complex traits, and reveals etiologically salient de novo subtypes in ADRD.

Omics Technologies Posters - Thursday

PB3139. Single cell RNA-seq analysis of Bone Marrow-derived Osteoblasts.

Authors:

L. Dillard¹, W. Rosenow¹, G. Calabrese¹, I. mesner¹, B. Al-Barghouthi¹, A. Abood¹, E. Farber¹, S. Onengut-Gumuscu¹, D. Brooks², M. Bouxsein², S. Tommasini³, M. Horowitz³, C. Rosen⁴, C. Farber¹; ¹Univ. of Virginia, Charlottesville, VA, ²Harvard Univ., Cambridge, MA, ³Yale Univ., New Haven, CT, ⁴Univ. of Maine, Orono, ME

Abstract Body:

Skeletal development and maintenance are controlled by numerous cell-types. Historically, genomic studies of bone cells have been challenging due to difficulties in isolating homogenous cell-types from marrow or bone. Recently, this has begun to change with the emergence of single-cell technologies. Here, we profile the transcriptomes of individual cultured bone marrow-derived stromal cells (BMSCs), a popular model of osteoblast differentiation and activity, from five Diversity Outbred (DO) mice using single-cell RNA-seq (scRNA-seq). The goals of the study were to explore technical challenges, evaluate heterogeneity, and determine if BMSCs could serve as a model for the generation of cell-type specific profiles of osteogenic cells derived from hundreds of mice in order to inform genetic studies. We demonstrate that dissociation of BMSCs from a heavily mineralized matrix has little effect on viability or their transcriptomic signature. Furthermore, we show that BMSCs cultured under osteogenic conditions are highly heterogeneous and consist of cells with characteristics of mesenchymal progenitors, pre-adipocytes, osteoblasts, osteocytes, and non-osteogenic immune cells. Importantly, these cells were representative and similar from a transcriptomic perspective to their *in vivo* counterparts. We also demonstrate the ability to multiplex cells and subsequently assign cells to their sample-of-origin using demultiplexing approaches based on genotypes inferred from coding SNPs. Exploring BMSCs from individual mice revealed that cell composition is dependent on characteristics of bone, especially bone marrow adipose content. Finally, we show that genes with higher expression in preadipocytes and osteocytes contribute more of the genetic signal identified by bone mineral density GWAS than the other cell-types profiled. Together, these data suggest that scRNA-seq of BMSCs has the ability to provide insight into the biology of multiple cells in the mesenchymal lineage. It also suggests that BMSCs may be an excellent model for the generation of cell-type specific profiles of homogenous bone cell-types in large mouse and human populations.

Omics Technologies Posters - Wednesday

PB3140. Single neuron whole genome sequencing for somatic CNVs using three different genome amplification methods

Authors:

C. Proukakis^{1,2}, E. Kalef-Ezra^{1,2}, D. Perez-Rodriguez¹, I. Bomann¹, M. Toffoli³, S. Behera⁴, S. Scholz⁵, Z. Zaubmuktane^{1,2}, F. Sedlazeck^{6,2}; ¹Univ. Coll. London, London, England, United Kingdom, ²Aligning Sci. Across Parkinson's (ASAP) Collaborative Res. Network, Chevy Chase, MD, ³Univ. Coll. London, London, United Kingdom, ⁴Baylor Coll. of Med., Houston, TX, ⁵NIH, Bethesda, MD, ⁶Baylor Coll. Med., Houston, TX

Abstract Body:

Somatic mutations occur in healthy and diseased brain. We have detected somatic CNVs of the alpha-synuclein gene (*SNCA*) using FISH in patients with synucleinopathies (Parkinson's disease and multiple system atrophy- MSA). Large CNVs can be detected by low coverage single-cell whole genome sequencing (scWGS) after whole genome amplification (WGA). We have already performed this as proof-of-principle in two MSA brains, with CNVs in ~30% of cells. We have now compared three WGA methods: Picoplex (Takara), a hybrid method, and two modifications of isothermal multiple displacement amplification (MDA): Primary template-directed amplification (PTA, BioSkryb) which uses chain-terminating nucleotides to avoid over-amplification, and droplet MDA (dMDA) in the X-drop device (Samplix) which amplifies genomic fragments in droplets. We amplified 55 single nuclei from the cortex of one MSA and one control brain followed by Illumina scWGS ~1x. The fraction of the genome covered at least once is higher in PTA than Picoplex, and lowest in dMDA. We analysed data using Ginkgo. At 500 kb bins, most cells amplified by Picoplex or PTA were suitable for CNV detection as the median absolute deviation (MAD) was <0.3, but none of the dMDA cells were successful (mean 0.19 for Picoplex, 0.20 for PTA, 0.53 for dMDA). Possible CNVs (gains) were reported in more MSA than control cells: Picoplex- 67% MSA, 17% control; PTA: 86% MSA, 43% control. We performed further analysis using Alfred for cells sequenced together. The error rate in MSA was highest in Picoplex (0.75%) against 0.49% in PTA and 0.37% in dMDA. For PTA, the error rate was lower both in control brain (0.37%) and NA12878 cells (0.34%), and for dMDA this was also lower in control brain (0.29%). The excess errors in MSA may represent DNA damage or somatic SNVs. The fraction of read pairs properly mapped in MSA was 91.4% in Picoplex, 89.4% in dMDA and 84.4% in PTA. Notably this was higher in PTA in control brain (87.8%) and highest in NA12878 cells (89.3%). dMDA also showed a higher fraction in control brain (91.9%). These results suggest that PTA leads to most chimeras. The lower fraction of properly mapped pairs in MSA using both PTA and dMDA could be due to somatic SVs. Nanopore sequencing of 6 MSA cells amplified by dMDA is being analysed. We provide comparative data from 3 WGA methods, with additional analysis of errors and chimeras, and long read sequencing, underway. The possible excess of errors and improperly mapped read pairs in MSA merits further investigation, and correlation with Illumina and Nanopore WGS from bulk tissue will also be performed. Additional scWGS will include a second MSA brain, and cells isolated by laser capture microdissection.

Omics Technologies Posters - Thursday

PB3141*. Single-cell allele-specific expression analysis reveals dynamic and cell-type-specific regulatory effects

Authors:

G. Qi¹, B. J. Strober², J. M. Popp¹, H. Ji¹, A. Battle¹; ¹Johns Hopkins Univ., Baltimore, MD, ²Harvard T.H. Chan Sch. of Publ. Hlth., Brookline, MA

Abstract Body:

Background. Allele specific expression (ASE), which quantifies the imbalance in gene expression between two parental alleles, is a powerful tool to study cis-regulatory effects. Such effects across conditions, e.g. experimental interventions, cell types, and developmental trajectories, can reveal differential ASE (D-ASE) indicating gene by context interactions. Recently, single-cell RNA sequencing (scRNA-seq) has allowed the measurement of ASE at the resolution of individual cells. Though statistical methods for identifying D-ASE using scRNA-seq are beginning to emerge, they have not properly addressed the correlation among cells due to repeated measurements per individual, or unknown phase information in the absence of genotype data. **Methods.** We develop DAESC, a statistical method for conducting D-ASE analysis using scRNA-seq data across multiple individuals. DAESC can be used for D-ASE with respect to any condition of interest, including cell-level conditions such as cell types or pseudotime, and sample-level conditions such as disease status. DAESC is comprised of a baseline model of beta-binomial regression with random effects accounting for correlation of cells from the same individual, and an extended mixture model to conduct implicit phasing. **Results.** We demonstrate through simulations that DAESC accurately captures D-ASE effects in a wide range of scenarios. The extended mixture model leads to substantially improved power under weak linkage disequilibrium between the exonic and regulatory SNPs. We use DAESC to analyze the single-cell ASE data from an endoderm differentiation experiment of 125 human induced pluripotent stem cell lines by Cuomo et al, which was designed to identify dynamic genetic effects on gene expression. DAESC identifies 622 genes (FDR<0.05) that show dynamic ASE along the developmental trajectory, and 72% are novel findings not reported by the original paper. Gene-set enrichment analysis shows significant enrichment in pathways related to the development of bone, nervous system, circulatory system, etc. A second application to a pancreatic islet dataset identifies several genes that show cell-type-specific D-ASE between type 2 diabetes patients and controls, indicating potentially dysregulated gene expression associated with the disease. **Conclusions.** We propose a powerful method for differential ASE analysis using scRNA-seq data. The method can be applied to a wide range of scenarios to study context-specific cis-regulatory effects, ultimately illuminating regulatory effects that may underlie processes including cellular differentiation and disease.

Omics Technologies Posters - Wednesday

PB3142. Single-cell dissection of ALS and frontotemporal dementia in the human motor and prefrontal cortices.

Authors:

S. Pineda^{1,2}, H. Lee¹, B. Fitzwalter¹, R. Linville¹, E. Cook³, D. Dickson³, V. Belzil³, M. Kellis^{1,2}, M. Heiman¹; ¹Massachusetts Inst. of Technology, Cambridge, MA, ²Broad Inst. of MIT and Harvard, Cambridge, MA, ³Mayo Clinic, Jacksonville, FL

Abstract Body:

Amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD) are devastating and fatal neurodegenerative diseases that share many clinical, pathological, and genetic signatures. However, the mechanistic basis of their shared and distinct circuitry remains unknown at the molecular level. To uncover cell type-specific transcriptional changes, underlying biological pathways, and putative upstream regulators, we conducted high-resolution single-cell profiling of transcriptional alterations in the primary motor and dorsolateral prefrontal cortices of 75 sporadic and *C9orf72*+ familial ALS and FTLD donor individuals and unaffected controls, providing the most comprehensive-to-date characterization of Brodmann areas 4 and 9, and yielding insights of unprecedented resolution into both ALS and FTLD. Our analysis revealed enhanced cross-region, and cross-phenotypic vulnerability of an extratelencephalic layer Vb population that includes the ALS and FTLD-implicated Betz cells and Von Economo neurons. We identified novel and highly-specific marker genes for these previously ill-defined populations and found that the majority of these are shared across brain regions and may comprise unique vulnerability factors. Analysis of differentially expressed genes (DEGs) showed that several ALS- and FTLD-associated genes were dysregulated across phenotypes and some with regional and cell type specificity. We observed that a number of high-ranking DEGs belonged to the class of TDP-43 binding targets, and that non-canonically disease-associated cells of the same type showed similar disease signatures across both phenotypes and brain regions. We also discovered disease-associated, but regionally-dependent, changes in gene expression and protein localization of several tight junction proteins and their regulators in vascular populations. Shared across cell types, we observed enrichment of pathway terms associated with oxidative stress, protein localization, and ribosomal dysfunction, and within excitatory neurons, an enrichment for terms associated with axonal growth, maintenance, and repair consistent with reports of axonal integrity deterioration. ALS-enriched pathways showed greater specificity for microtubule maintenance- and organization-associated processes, while FTLD dysregulated genes showed generally amplified enrichment for stress response-associated pathways. Overall, our study represents the largest and most accurate molecular atlas of these two human brain regions to date, and the first cell type-specific molecular characterization of ALS and FTLD in either.

Omics Technologies Posters - Thursday

PB3143*. Single-cell genome-wide association reveals a nonsynonymous variant in *ERAP1* confers increased susceptibility to influenza virus.

Authors:

B. Schott¹, L. Wang¹, X. Zhu¹, A. T. Harding¹, E. R. Ko¹, J. S. Bourgeois¹, E. J. Washington¹, T. W. Burke¹, J. Anderson¹, E. Bergstrom², Z. Gardener², S. Paterson², R. G. Brennan¹, C. Chiu², M. T. McClain¹, C. W. Woods¹, S. Gregory¹, N. S. Heaton¹, D. C. Ko¹; ¹Duke Univ., Durham, NC, ²Imperial Coll., London, United Kingdom

Abstract Body:

Diversity in the human genome is one factor that confers resistance and susceptibility to infectious diseases. This is observed most dramatically during pandemics, where individuals exhibit large differences in risk and clinical outcomes against a pathogen infecting large portions of the world's populations. Here, we developed scHi-HOST (single cell High-throughput Human in vitro Susceptibility Testing), a method for rapidly identifying genetic variants that confer resistance and susceptibility to pathogens. scHi-HOST leverages scRNA-seq (single-cell RNA-sequencing) to simultaneously assign genetic identity to individual cells in mixed infections of cell lines of European, African, and Asian origin, reveal associated genetic variants for viral entry and replication, and identify expression quantitative trait loci (eQTLs). Applying scHi-HOST to influenza A virus (IAV), we identified eQTLs at baseline and in genes that are induced by IAV infection. Integration of scHi-HOST with a human IAV challenge study (Prometheus) revealed that a missense variant in *ERAP1* (*Endoplasmic reticulum aminopeptidase 1*; rs27895) was associated with IAV burden in cells and human volunteers. Functional studies using RNA interference, ERAP1 inhibitor, and overexpression of alternative alleles demonstrated that ERAP1 is exploited by IAV to promote infection. Specifically, the nonsynonymous substitution, which results in a glycine to aspartate substitution at ERAP1 residue 348, would disrupt the substrate binding pocket of ERAP1, likely resulting in a significantly altered preference for substrates, poorer catalytic efficiency, or both. Finally, rs27895 exhibits substantial population differentiation, with the higher frequency of the minor T allele in two African populations likely contributing to the greater permissivity of cells from these populations to IAV infection. scHi-HOST is an important resource for understanding susceptibility to influenza and is a broadly applicable method for decoding human genetics of infectious disease.

Omics Technologies Posters - Wednesday

PB3144. Single-cell multi-omics reveals gene regulation dynamics during mouse secondary palate development

Authors:

F. Yan¹, A. Suzuki^{2,3}, C. Iwaya^{2,3}, H. Yoshioka^{2,3}, G. Pei¹, X. Chen¹, M. Yu¹, L. Simon⁴, J. Iwata^{2,3}, Z. Zhao^{1,5}; ¹Ctr. for Precision Hlth., Sch. of BioMed. Informatics, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ²Dept. of Diagnostic and BioMed. Sci., Sch. of Dentistry, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ³Ctr. for Craniofacial Res., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ⁴Therapeutic Innovation Ctr., Baylor Coll. of Med., Houston, TX, ⁵Human Genetics Ctr., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract Body:

Background: Palatogenesis is a complex, dynamic, and highly orchestrated process, which relies on tight regulation of gene expression. Its disruption leads to cleft palate, the second common birth defect. **Methods:** To simultaneously resolve both gene expression and regulation dynamics and delineate palate development at the cellular level, we profiled single-cell transcriptomics and epigenomics from the same cells (n = 36,154) isolated from mouse palatal shelves at embryonic day (E) 12.5, E13.5, E14.0, and E14.5 using the 10x Chromium Single Cell Multiome platform. **Results:** Unsupervised dimension reduction and clustering analysis revealed eight major cell types defined by established marker genes, including cranial neural crest (CNC)-derived mesenchymal (*Prrx1*), epithelial (*Epcam*), endothelial (*Cldn5*), myeloid (*Lyz2*), glial (*Sox10*), neuronal (*Tubb3*), myogenic progenitors (*Myod1*), and myocytes (*Myf5*). Multiome on the same tissue allows reliable inference of critical transcription factors (TFs) by setting criteria on both modalities. Using a correlation-based approach, we identified 3,076 significant positively correlated peak-gene linkages (coefficient > 0, *adjusted p-value* < 0.05). Coupling differential gene expression analysis and motif enrichment analysis of these peaks, we pinpointed a list of TFs whose expression was enriched in the RNA measurements for a specific cell type, and whose motif accessibility was enriched in the ATAC measurements. For example, both the expression of *Foxo1* (*RNA.p-value* < 2.22×10^{-16}) and accessibility of Foxo1 binding motif MA0480.1 (*motif.p-value* = 4.91×10^{-18}) were enriched in endothelial cells. We then focused on CNC-derived mesenchymal cells. Integration of RNA velocity and pseudotime analysis revealed five trajectories, showing continuous differentiation of progenitors into different subtypes, including anterior and posterior palatal mesenchymal cells, dental mesenchymal cells, osteoblasts, and perimysial cells. The application of additional modeling showed high initial state probabilities in progenitor cells and terminal state probabilities in these subtypes, which validated inferred trajectories. We characterized a list of lineage-determining TFs that exhibited continuous progression along each trajectory, which was active at the early, middle, and late-stage of the trajectory, respectively. **Conclusion:** Collectively, our study uncovered gene-regulatory dynamics and built an atlas of gene expression and regulation of the developing mouse secondary palate. This multi-omic resource will serve as a comprehensive reference map and facilitate research on cleft palate.

Omics Technologies Posters - Thursday

PB3145. Single-cell RNA sequencing (scRNA-seq) of fresh, neuronal tissue using Pre-templated Instant Partitions (PIPseq) and comparison to alternative scRNAseq methods

Authors:

K. Kugler¹, A. May-Zhang¹, P. Frazel², S. Liddelow²; ¹Fluent BioSci., Watertown, MA, ²NYU, New York, NY

Abstract Body:

In neuroscience, scRNAseq has greatly improved our ability to identify, distinguish, and discover novel and rare cell types. However, even the largest individual scRNA-seq datasets only represent about 1% of the over 100 million cells in the adult mouse brain. PIPseq is a novel single-cell sequencing technology that makes scRNAseq accessible and affordable to all researchers and facilitates characterization and interrogation of larger and more numerous neuronal datasets. PIPseq has sample preparation improvements over other commercial systems by allowing samples to be “instantly partitioned” into single cell reaction vesicles without the use of microfluidics. PIPseq also has the advantage of convenient stable stopping points where sample collection can occur at one site and downstream processing can occur in a core lab or at Fluent BioSciences using our service model. Here, we compare the biological results and assess the performance of PIPseq versus 10x single cell RNAseq on neuronal tissue. We performed scRNAseq using both platforms using the same tissue sample (freshly dissected and dissociated postnatal day (P)30 mouse forebrain). PIPseq and 10x generated sequencing data was able to resolve all expected neuronal cell types. The proportion and variety of the common cell type markers derived from each scRNAseq method was comparable, with no drop-outs or biases observed. The differentially expressed genes from each technique were highly concordant and replicated what is commonly described in the literature and outlined in the Linnarsson Brain Atlas. These data highlight that PIPseq is a viable, reliable, and powerful method that can be used to perform scRNAseq analysis on fresh neuronal tissue in a single, simple reaction. Benefits to our simplified workflow include long shelf-life of capture cells (in PIPs), low cost, rapid generation of single cell suspensions, and scalability - with future capacity for sequencing millions of cells in a single sample.

Omics Technologies Posters - Wednesday

PB3146. Single-molecule, modified base sequencing to identify frequency and cause of rAAV vector breakpoints.

Authors:

J. Thompson¹, D. Selby¹, L. Soares¹, T. Hanscom¹, D. Browne², J. Walsh², M. Weiland², J. Korlach², D. Przybylski¹, J. Wright¹; ¹Homology Med.s, Bedford, MA, ²Pacific BioSci.s, Menlo Park, CA

Abstract Body:

There are approved gene therapies using recombinant adeno-associated virus (rAAV) as the vector for delivery of therapeutic genes, and the quality of the material is critical for safe and efficacious use. All AAV serotypes contain single-stranded DNA within each capsid that raises analytical challenges. Complementary genomic strands hybridize when capsids are lysed and standard sequencing library methods cause information from the individual genomes to be lost when inter-strand mismatches and gaps are repaired. This is especially important when gaps are repaired because that results in incomplete knowledge of whether genomes are partial or full-length. The presence of empty or partially full capsids is one factor that affects rAAV quality. We developed a novel NGS library preparation method that allows us to identify and quantitate gaps by distinguishing pre-existing DNA from DNA added during NGS library preparation. During the repair process common to NGS library methods, the standard nucleotides dCTP and dATP are replaced with the modified nucleotides 4Me-dCTP and 6Me-dATP so that any new DNA includes stretches of modified bases while the pre-existing ssDNA consists of natural, unmodified bases. Using the Sequel II system, the modified bases can be distinguished from unmodified bases, enabling breakpoint identification at high resolution. This method was used on a vector known to break during packaging, a phenomenon that can occur in AAV vectors of all serotypes. The exact location and frequencies of breaks allowed us to alter nearby sequences and reduce the frequency of breakage. The resultant new designs provide a higher quality AAV vector that is less susceptible to partial genomes. Partially filled capsids have been an ongoing regulatory agency concern and this approach provides information that standard methods do not. This use of modified bases for localizing DNA breaks enables improved vector designs and provides better characterization metrics for AAV vectors, resulting in higher quality gene therapy vectors. The same approach can be used in other systems where knowledge of pre-existing sequence and structure is important and may be lost when DNA is repaired.

Omics Technologies Posters - Thursday

PB3147. Single-nucleus profiling of the cerebellar cortex in essential tremor reveals Bergmann gliosis and oligodendrocyte myelin abnormalities as hallmarks of disease

Authors:

C-E. Castonguay¹, T. Bécrot¹, M. Medeiros¹, C. Liao¹, A. Rajput², P. Dion¹, G. Rouleau³; ¹McGill Univ., Montreal, QC, Canada, ²Univ. of Saskatchewan, Saskatoon, SK, Canada, ³Montreal Neurological Inst.-Hosp., Montreal, QC, Canada

Abstract Body:

Essential tremor (ET) is one of the most common movement disorder, affecting nearly 1% of the worldwide population. It presents itself as kinetic tremor affecting mostly the upper limbs but also other body parts such as the lower limbs and head. The disease has a significant impact on the daily lives of patients and can severely affect their autonomy. Previous genetic, transcriptomic, and histopathological studies of ET have highlighted the important contribution of the cerebellar cortex in the development of the disease. However, the molecular and cellular correlates of the disease within the cerebellar cortex are poorly understood. We sought to assess disease-related gene expression changes in all cell types of the cerebellar cortex using single-nucleus RNA sequencing of 16 ET and 16 control cerebellar samples in order to better understand the pathophysiological mechanisms underlying ET. Split-pool ligation-based transcriptome sequencing (SPLiT-seq) was used to obtain single-nucleus reads from frozen cerebellar samples. Cells were then clustered using Seurat and pseudo-bulk differential expression analysis for each cell type was done using DESeq2. We found that Bergmann glia, but not fibrous astrocytes, demonstrated differential expression of multiple genes related to astrogliosis (i.e., *ID4*, *NPL*, *SIPRI*). Pathway enrichment analysis identified terms related to glial cell differentiation (qval = 6.03E-04), astrocyte projection (qval = 5.24E-04) and the postsynaptic cell compartment (qval = 3.13E-04). In addition, differentially expressed genes in oligodendrocytes were enriched for structural constituents of myelin sheath (qval = 7.85E-03), concordant with previous associations between ET and axonal pathology. Consistent with these findings, deconvolution of bulk-RNA seq data from these 32 cerebellar cortex samples using Bisque demonstrated an increased proportion of Bergmann glia in ET samples (Z-score = 3.47, qval = 6.73E-03) as well as a significant decrease in the proportion of oligodendrocytes (Z-score = -3.05, qval = 2.93E-02). In support of previous histopathological observations, we found no difference in the proportions of Purkinje cells between cases and controls. Our study highlights potential cellular mechanisms implicating non-neuronal cells in ET, paving the way for more in-depth functional studies.

Omics Technologies Posters - Wednesday

PB3148. Single-stranded library preparation for cfDNA identifies unique fragment length signature in myelodysplastic syndrome with multi-lineage dysplasia and ring sideroblasts

Authors:

C. Schwartz¹, C. Naughton¹, J. Morgan¹, C. Troll¹, A. Ali², A. Raza², V. Rao¹, K. Harkins Kincaid¹, R. E. Green^{1,3}, C. Vaske¹; ¹Claret BioSci., Santa Cruz, CA, ²Columbia Univ. Med. Ctr., New York, NY, ³Univ. of California, Santa Cruz, Santa Cruz, CA

Abstract Body:

Fragmentomic analysis of cfDNA fragment lengths has been shown to reveal the epigenetic state of the source cells. Here we describe a novel signal based on cfDNA fragment lengths that likely reveals a novel epigenetic state in a subtype of myelodysplastic syndrome (MDS). The finding is made possible by using a single-stranded DNA sequencing preparation (*SRSLY*) that omits DNA end repair to preserve the true fragment length. We sequenced *SRSLY* libraries prepared from 1-5ng of cfDNA from healthy donors and MDS cases. We calculated the ratio of short to long cfDNA fragments across the genome, a feature that correlates with 3D nuclear compartment data derived from Hi-C experiments. A/B nuclear compartments are cell-specific and associated with open/closed chromatin, respectively. Deviations from expected open/closed chromatin states would reflect changes in the epigenomic landscape. Most MDS and normal samples show fragment length ratios concordant with the nuclear compartments from lymphoblastoid cell line GM12878. However, 75% (9/12) of the samples of subtype RS-MLD (multiple lineage dysplasia with ring sideroblasts) show consistent atypical fragment length ratios, with regions in the closed or B compartment showing a pattern resembling that of the open compartment. This signal is observable with as few as 5 million sequencing reads. We interpret this signal as an unexpected “opening” of chromatin in those genomic regions, which appear to be enriched in the B3 subcompartment. RS-MLD cfDNA also shows an atypical fragment length pattern within the open A compartment. The regions causing this deviation are enriched in the A2 subcompartment. In sum, the observed signals of genome reorganization suggest a different etiology for RS-MLD, possibly during maturation of erythrocytes, and could be a biomarker for erythrocytic disorders and treatment decisions.

Omics Technologies Posters - Thursday

PB3149*. SnapFISH: a computational pipeline to identify chromatin loops from DNA FISH data

Authors:

L. Lee¹, M. Hu¹, Y. Li²; ¹Cleveland Clinic, Cleveland, OH, ²Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract Body:

Super-resolution imaging technologies, such as chromatin tracing, DNA MERFISH, DNA seqFISH+, ORCA and OligoFISSEQ, provide powerful tools to measure chromatin spatial resolution at single cell resolution, enabling direct visualization of looping events between gene promoters and their cis-regulatory elements. However, computational methods to detect chromatin loops at high resolution from super-resolution imaging data are still lacking. Here, we describe Single-Nucleus Analysis Pipeline for DNA FISH data (SnapFISH), the first method that can identify chromatin loops from super-resolution imaging data. We first applied SnapFISH to 5Kb resolution chromatin tracing data at the 210Kb region in chromosome 3 containing the *Sox2* gene in mouse embryonic stem cells (mESCs), and found that SnapFISH can identify the loop between *Sox2* promoter and its confirmed super-enhancer ~100Kb downstream with as few as 75 cells. We further applied SnapFISH to 25Kb resolution DNA seqFISH+ data in all 19 autosomes in mESCs. SnapFISH demonstrated high sensitivity and accuracy in detecting chromatin loops, evaluated against chromatin loops detected by HiCCUPS from deeply sequenced bulk Hi-C data from mESCs. Finally, we re-analyzed the 25Kb resolution DNA seqFISH+ data collected from mouse brain cortex tissues, and detected cell-type-specific loops which are closely associated with cell-type-specific epigenetic marks and cell-type-specific gene expression. In sum, SnapFISH is the first computational method to identify chromatin loops from super-resolution imaging data, and has the potential to facilitate better characterization of cell-type-specific chromatin spatial organization, and improve our understanding of cell-type-specific gene regulation mechanisms.

Omics Technologies Posters - Thursday

PB3150*. Somatic genomic alterations in single neurons from brains with chronic traumatic encephalopathy (CTE)

Authors:

C. Ma^{1,2}, G. Dong^{1,2}, M. B. Miller^{1,2,3}, A. Y. Huang^{1,2,3}, A. C. McKee^{4,5}, E. A. Lee^{1,2,3}, C. A. Walsh^{1,2,3,6}; ¹Harvard Med. Sch., Boston, MA, ²Boston Children's Hosp., Boston, MA, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Boston Univ. Sch. of Med., Boston, MA, ⁵VA Boston Hlth.care System, Boston, MA, ⁶Howard Hughes Med. Inst., Boston, MA

Abstract Body:

Chronic traumatic encephalopathy (CTE) is a neurodegenerative disease associated with repetitive head trauma. The genetic, molecular, and cellular mechanisms behind the development of CTE are less well understood than in other neurodegenerative diseases such as Alzheimer's disease (AD). The advent of single-cell sequencing technologies allows for the study of molecular perturbations at the individual cell or cell-population level as well as the analysis of contributions of non-germline somatic mutations to disease pathogenesis. Previous studies of single-cell whole genome sequencing (scWGS) on aging and neurodegenerative brains showed that somatic single-nucleotide variants (sSNVs) increase both with aging and in disease, but present with distinct patterns of mutational signatures, suggesting that genetic, environmental, or disease states might influence this accumulation.

In this study, we applied scWGS to neurons from the prefrontal cortex of CTE brains. Using PTA (Primary Template-directed Amplification), with LiRA (Linked-Read Analysis) and SCAN-SNV (Single Cell ANalysis of SNVs) bioinformatic pipelines to distinguish sSNVs from amplification artifacts, we compared the rates of sSNV accumulation in CTE and control brains. We found a significant increase of hundreds of sSNVs in CTE as compared with age-matched controls. Additionally, we identified specific mutational signatures more abundant in CTE than in controls, distinct from the composition of mutational signatures in AD, which gave us insight into potential causative agents of CTE. Since CTE is hypothesized to be caused by exposure to repetitive head trauma, its areas of pathological overlap with other neurodegenerative diseases, such as AD, make CTE a unique model for studying the effects of molecular pathways in neurodegeneration.

Omics Technologies Posters - Wednesday

PB3151. sparQ mRNA-Seq: Consistent and high-quality mRNA library preparation from both abundant and limiting samples.

Authors:

K. Conley, M. Aitichou, R. Heller, D. Schuster; Quantabio, Frederick, MD

Abstract Body:

Next-generation sequencing (NGS) of strand-specific RNA libraries has become established as one of the most comprehensive and informative methods for transcriptome profiling for molecular analysis of disease states and biological traits, allowing for the identification of both known and novel RNA structural features and isoforms and for the accurate quantification of transcripts from both orientations. Removal of uninformative highly abundant RNAs is a critical step in achieving a focused library containing biologically relevant sequences for insights and actionable results. Ribosomal RNA depletion methods are useful for preparing libraries that contain both exonic and biologically relevant non-coding RNAs sequences and can overcome 3'-bias in degraded RNA preparations. However, these methods are often costly, time consuming, and limited by species-specific homology constraints. mRNA-seq, which exploits selection of polyadenylated RNA in addition to utilizing directional library preparation, also eliminates highly abundant and uninformative RNA sequences, and is applicable to poly(A)+ RNA from any species. However, this method is typically limited to samples capable of providing relatively large quantities of total RNA. Focused sequence analysis of less abundant mRNA quantities requires both an efficient mRNA enrichment module as well as highly optimized enzymatic reactions.

Here we present the sparQ mRNA-Seq kit that enables efficient preparation of stranded mRNA-seq libraries for Illumina NGS systems in less than 4.5 h. This workflow reduces hands-on time and delivers reliable mRNA-seq libraries from as little as 5 ng to 1 µg of total RNA with high quality reads from reference total RNA. Evaluation of library yields, mapped reads, transcript coverage uniformity, and (individual/total) gene count qualified sparQ mRNA-Seq as the most sensitive of mRNA-seq workflows compared to other leading mRNA-Seq kits. We demonstrate application of this technology to include the study of samples with limiting amounts of RNA by combining efficient isolation of mRNA with a streamlined workflow for preparing mRNA-seq libraries that delivers consistent exonic read coverage across varying RNA input amounts and reduced bias in GC-rich transcripts.

Omics Technologies Posters - Thursday

PB3152. STOC: A Simple Tool Of Compression

Authors:

R. Klein; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Modern genomic technologies are producing an avalanche of data that needs to be stored and analyzed. Efficient use of these data requires balancing the need to compress them to save on storage costs while maintaining them in a format that can easily be read as-is and randomly accessed for analyses without the need for storing the uncompressed data. The underlying data used by downstream genomics analysis tools typically is represented as a text file. Many programs that work with genetic variation data can read the files when they are compressed with the standard gzip compression program. However, gzip-compressed files do not support random access; the file must be uncompressed in-memory and scanned to identify regions of interest. To get around this limitation, various bespoke compression approaches are used, though these approaches require specialized software. Simultaneous with the growth of genomics data, numerous technologies have been developed to work with large data sets in a scalable fashion. Many of these technologies focus on data parallelism, where the same operation is performed simultaneously on each record of the data. These technologies work optimally when separate processes can work on separate parts of the compressed file at the same time, which is not possible out of the box with standard genomics file formats. Here, I propose a Simple Tool for Compression (STOC), designed to be use as part of a larger genomics data lake. The concept of this file format is simple: the first N columns of a file are stored uncompressed, as these are used for searching purposes. The remaining columns are then compressed into a single string, represented as base64 text. This then allows the data to be manipulated using standard genomics or data analytic programs, before the underlying data is uncompressed for the final computations. As STOC is especially useful in working with imputed genotype data with low minor allele frequencies, as can be found from TopMed imputation, I designed it to automatically extract and recompress the results from the TopMed imputation server. The resulting files are VCF-compliant and variant-level filtering can be performed with standard VCF-aware software. To illustrate its utility, I ran it on TopMed imputation results for a set of 3972 men. While the original bgzip-compressed TopMed data took 152GB of disk space, using STOC and bzip2 compression took slightly less space (142GB). This approach may be especially useful for datasets that will be repeatedly analyzed over subsets of the genome.

Omics Technologies Posters - Wednesday

PB3153. STRT-N: A newly optimised single-cell RNA sequencing method.

Authors:

N. Boskovic^{1,2}, G. Yazgeldi³, S. Ezer^{3,4}, S. Katayama^{1,3,4}, M. H. Tervaniemi^{3,4}, T. Skoog¹, K. Krjutskov^{5,6}, J. Kere^{1,3}; ¹Dept. of BioSci.s and Nutrition, Karolinska Inst.t, Stockholm, Sweden, ²Dept. of Obstetrics and Gynecology, Univ. of Helsinki, Helsinki, Finland, ³Folkhälsan Res. Ctr., Helsinki, Finland, ⁴Stem Cells and Metabolism Res. Program, Univ. of Helsinki, Helsinki, Finland, ⁵Competence Ctr. of Hlth.Technologies, Tartu, Estonia, ⁶Dept. of Obstetrics and Gynecology, Inst. of Clinical Med., Univ. of Tartu, Tartu, Estonia

Abstract Body:

Background:Single-cell RNA sequencing (scRNA-seq) has revolutionised the field of transcriptomics since its introduction. The protocols for scRNA-seq have progressed from one cell to the possibility of sequencing thousands of cells. However, the early methods that could use only a very limited number of samples have become obsolete with the progress of the scRNA-seq methods, creating a new need for a method that can be used for a small number of samples. Therefore, we have looked back in earlier protocols and optimised one for use with the current sequencing platforms. **Methods:** Single-cell Tagged Reverse Transcription (STRT) was previously adjusted to work on bulk RNA, globin rich tissue samples, and new sequencing technologies, but had become suboptimal for scRNA-seq. For optimisation, we tested 10 different template switching oligos (TSO), 3 different reverse transcriptase enzymes, 3 enzymes for cDNA synthesis, and additional clean-up steps. **Results:** We found that the sensitivity of the method is greatly dependent on the reverse transcriptase enzyme as well as the cDNA amplification enzyme when working with very small RNA input (10 pg). Additional clean-up step before cDNA amplification was needed to prevent by-products (primer dimers, excess TSO, and oligo dT primer) overtaking the reaction. **Conclusion:**The newly optimised STRT-N RNA-seq method is highly sensitive and can detect as low as 10 pg of RNA input. The sequencing library consists of 48 barcoded cDNA samples and is highly useful for studies with limited availability of samples such as mammalian embryos.

References: Islam S., Kjällquist U., Moliner A., Zajac P., Fan J.B., Lönnerberg P., and Linnarsson S. (2012). Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat Protoc.* Apr 5;7(5):813-28. Islam S., Zeisel A., Joost S., La Manno G., Zajac P., Kasper M., Lönnerberg P., and Linnarsson S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods.* Feb;11(2):163-6. Islam S, Kjällquist U, Moliner A, Zajac P, Fan JB, Lönnerberg P, Linnarsson S. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* 2011 Jul;21(7):1160-7. doi: 10.1101/gr.110882.110. Ezer S, Yoshihara M, Katayama S; DoGA consortium, Daub C, Lohi H, Krjutskov K, Kere J. Generation of RNA sequencing libraries for transcriptome analysis of globin-rich tissues of the domestic dog. *STAR Protoc.* 2021 Dec 12;2(4):100995. doi: 10.1016/j.xpro.2021.100995. PMID: 34950881; PMCID: PMC8672047. **Grants:**The European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813707.

Omics Technologies Posters - Thursday

PB3154. Structural and copy number variant detection, filtering, annotation, and classification by optical genome mapping in constitutional disorders.

Authors:

B. Clifford, J. Hauenstein, H. Barseghyan, A. Pang, A. Chaubey, A. Hastie; Bionano Genomics, San Diego, CA

Abstract Body:

In contrast to methods like karyotyping, chromosomal microarray (CMA), fluorescence in situ hybridization (FISH) or sequencing, optical genome mapping (OGM) is sufficiently versatile to identify all classes of structural variants from large-scale fusions to intragenic deletions in an easy to master technical and analytical workflow. OGM can comprehensively identify structural variants (SVs), copy number variants (CNVs), aneuploidies, absence of heterozygosity (AOH), and triploidy with high sensitivity and specificity. Here, we demonstrate a comprehensive and lab-ready workflow for whole genome OGM data analysis, alignment, and annotation using GRCh38, variant evaluation, and classification.

For constitutional disorders, OGM data is collected and analyzed by Bionano Access software. Maps are assembled based on direct fluorescent labeling of a six base pair sequence motif, with sample data aligned to GRCh38 to resolve variants and annotate against a common reference. SVs are further queried for overlap against a database of assay matched phenotypically healthy controls. To generate a prioritized list, we apply a workflow to filter based on size (≥ 1.5 kbp), near-absence in controls database ($\leq 1\%$ presence) and overlap with a gene. The resulting curated variant list presents an average of ~20-30 variants for further classification.

Each variant is assessed in a genome browser, showing properties such as size and putative impact to genes in a variant classifier visualization. Multiple analysts may work together to independently classify the variants. Lastly, a supervisor can reconcile analyst records to adjudicate classifications, record any large-scale genomic irregularities, and download a complete research report. Support for OGM data analysis is currently being developed for NxClinical variant classification software to enable automated ACMG scoring, additional databases and knowledgebases and ability to overlay NGS and CMA data. OGM data is well suited to identify a breadth of variants relevant to constitutional disorder research. The analytical approach annotates thousands of genome-wide variants accurately, and the filtering workflow discussed here directs focus to the most promising dozens. The subsequent classification and reporting tools provide utility for communicating their significance.

Omics Technologies Posters - Wednesday

PB3155. Structuring information via an immune-focused ontology enables the construction of a high-quality knowledge graph for the study of autoimmune diseases

Authors:

V. Q. Truong^{1,2,3,4,5}, J. D. Romano^{1,3}, A. R. Greenplate^{1,4,5,6}, S. M. Dudek^{1,3,7}, E. Wherry^{1,5,4,6}, M. D. Ritchie^{1,3,7,8}, ¹Univ. of Pennsylvania, Philadelphia, PA, ²Genomics & Computational Biology PhD Program, Philadelphia, PA, ³Inst. for BioMed. Informatics, Philadelphia, PA, ⁴Immune Hlth.Project, Philadelphia, PA, ⁵Inst. for Immunology, Philadelphia, PA, ⁶Dept. of Pharmacology & Translational Therapeutics, Philadelphia, PA, ⁷BioMed. and Translational Informatics Lab., Philadelphia, PA, ⁸Dept. of Genetics, Philadelphia, PA

Abstract Body:

Autoimmune diseases are a highly heterogeneous family of diseases which occur due to immune dysfunction causing systemic attacks against the body's own tissues. Approximately 4.5% of people worldwide are impacted by at least one autoimmune disorder, but global incidence of autoimmune disease is rising while our comprehensive understanding remains static. Recent discoveries of linked genes provide a clue for commonalities in co-occurring disorders, but tremendous work remains to uncover new links between disorders with suspected likelihoods. The study of autoimmunity is further complicated by the complexity of the immune system as a hierarchical network of interacting biomolecular components spanning multiple levels of biology (genes, proteins, pathways, immune cells, and more). The disorganization and separation of immunological knowledge across disparate repositories is a major hurdle preventing the discovery of possible biomarkers and disease mechanisms. Thus, biomedical informatics is well-suited to unify heterogeneous data into a queryable information network known as a knowledge graph (KG) where entities are stored as nodes, while edges represent the relationships connecting them. KGs offer a powerful and efficient solution for connecting heterogeneous knowledge, but recent biomedical implementations were constructed in a non-standardized process, contained outdated information, lacked structure, and/or did not capture autoimmunity well. These challenges led us to i) integrate and structure immune-related information into a queryable knowledge graph, and ii) validate an ontology-based model which turns implicit understanding into explicit reasoning. First, we identified several sources of curated immunological and biological knowledge from ImmunoGlobe, Hetionet, Pathway Commons, DisGeNet, and more. Next, we developed a specialized ontology to define the semantic meaning of entities (nodes) and relationships (edges) in our KG. Finally, we integrated information about genes, proteins, pathways, cell types, and diseases into a KG containing 8 node types, 14 relationship types, and more than 1000 immune-specific interactions. The KG quality was assessed according to established metrics: accuracy, trustworthiness, timeliness, availability, completeness, and consistency. Coupled together, our immune-focused ontology and knowledge graph enables the discovery of links between nodes, which represent pre-existing, unfilled, or new knowledge. Our work provides a path forward to explore data beyond single data-types and embrace a meta-dimensional framework for modeling strategies and applications in autoimmunity.

Omics Technologies Posters - Thursday

PB3156. SVPred: An integrated framework for Structural Variant Discovery

Authors:

V. Sarwal¹, S. Sankararaman¹, E. Eskin¹, S. Mangul²; ¹UCLA, Los Angeles, CA, ²Univ. of Southern California, LOS ANGELES, CA

Abstract Body:

SVPred: An integrated framework for structural variant discovery

Background: Structural variation (SV), refers to insertions, deletions, inversions, and duplications in human genomes. SV's plays a fundamental role in genome evolution and can underlie inherited or acquired diseases such as cancer. Comprehensive discovery of SV's from whole-genome sequencing data uses several approaches including read-pair, split-read, and read-depth. With advances in whole genome sequencing (WGS) technologies, a plethora of SV detection methods has been developed. However, dissecting SVs from WGS data presents a substantial number of challenges, with the majority of SV detection methods suffering from a high false-positive rate, and no existing method able to detect a full range of SV's present in a sample accurately.

Results: Here, we report an integrated structural variant calling framework, SVPred that combines the outputs of individual callers using a newly devised filtering and merging algorithm. Previous studies have shown the performance of SV callers to be significantly affected by the variant length. SVPred utilizes this difference by dividing the outputs of individual callers into bins based on the variant length, and combining the best-performing tools per bin. SVPred executes various combinations of Pindel, MistrVar, indelMINER, Manta, GRIDSS, BreakDancer, LUMPY, DELLY, CREST, RDXplorer, PopDel, GASV, and GenomeSTRiP to generate SV events. We evaluated the performance of SVPred on data with varying organisms and coverage. We ran SVPred on the public benchmark data for the Ashkenazi Jewish Trio son (NA24385/HG002) from the Genome-in-a-Bottle (GIAB) consortium, along with 7 strains of the mouse genome. SVPred has the highest F1 score measured across both mouse and human genomes. In addition, many of the variants predicted by SVPred were highly consistent with the experimentally validated truth set.

Conclusions: We present SVPred, a novel variant calling integrated framework to obtain a genome-wide landscape of SV's in a highly accurate manner. Using both mouse and human WGS data analysis, we show that SVPred provides an accurate SV calling framework and can serve as the gold standard for SV calling for the research community.

Omics Technologies Posters - Wednesday

PB3157. Systematic evaluation of transcriptome sequencing applied to real-world rare disease cohorts: Insights and limitations

Authors:

S. Silverstein^{1,2,3,4}, J. Fu¹, S. Donkervoort², P. Uapinyoying^{2,5}, B. N. Pusey¹, K. R. Ganapathy⁴, S. Gorokhova^{6,7,2}, T. Cassini⁸, T. DeLong², V. Ganesh⁹, B. Weisburd⁹, A. R. Foley², C. J. Tifft^{1,10}, M. Malicdan¹, Undiagnosed Diseases Network, W. Gahl¹, C. G. Bonnemann², P. Mohammadi⁴, D. R. Adams^{1,10}; ¹Undiagnosed Diseases Program, NIH, Bethesda, MD, ²Neuromuscular and Neurogenetic Disorders of Childhood Section, NINDS, Bethesda, MD, ³Rutgers New Jersey Med. Sch., Newark, NJ, ⁴Dept. of Integrative Structural and Computational Biology, The Scripps Res. Inst., La Jolla, CA, ⁵Ctr. for Genetic Med. Res., Children's Res. Inst., Children's Natl. Hosp., Washington, DC, ⁶Aix Marseille Univ, INSERM, MMG, U 1251, Marseille, France, ⁷Dept. of Med. Genetics, Timone Children's Hosp., APHM, Marseille, France, ⁸Med. Genetics & Genomic Med. Training Program, Natl. Human Genome Res. Inst., NIH, Bethesda, MD, ⁹The Broad Inst. of MIT and Harvard, Cambridge, MA, ¹⁰Office of the Clinical Director, NHGRI, NIH, Bethesda, MD

Abstract Body:

Rare diseases are estimated to affect 1 in 10 individuals. For those evaluated with exome and genome sequencing, typical diagnosis rates average 30%; incorporating transcriptomic sequencing additionally improves molecular diagnosis by 8-36% of previously undiagnosed cases. RNA studies can be used to augment genome analysis by detecting single allele transcription, interpreting potential splice modifiers leading to aberrant splicing, and identifying outlier transcription levels. Published computational tools are available to systematically evaluate these events. The DROP Pipeline, as an example, incorporates three component tools: OUTRIDER (expression outliers), FRASER (splicing outliers) and MAE (allelic imbalance).

We used DROP to analyze RNASeq data from 140 rare disease patients in two cohorts, including the NIH Undiagnosed Diseases Program (74 fibroblast samples, 75% undiagnosed) and the NINDS Neuromuscular and Neurogenetic Diseases of Childhood section (66 muscle samples, 57% undiagnosed). Each cohort contained known true positive genetic diagnoses that were shown to impair gene expression, and therefore could be detectable by the DROP analyses. We ran all DROP modules with both default and alternative parameters to assess the sensitivity and to detect new aberrant events for unsolved cases. We followed up the identified outlier genes for the undiagnosed cases by manual inspection of sequencing reads in IGV.

We found that the DROP pipeline correctly identified 3/27 (11%) known molecular diagnoses in the muscle cohort and 2/11 (18%) in the fibroblast cohort. This corresponds to an overall performance of 13% (95% CI 4.4%-28.1%), which is lower than previously reported values. Out of all candidate outliers, OUTRIDER produced one new diagnosis of *NBAS* deficiency for a patient with an atypical clinical phenotype in the fibroblast cohort. We present a detailed analysis of the results of each module, followed by a discussion of benefits and limitations of this approach.

Our analysis provides insight into optimizing the design of systematic approaches and strategies for rare disease genomics diagnostic pipelines and sets expectations of diagnostic benefit in "genome negative" cohorts. We recommend development of standardized protocols for sample and data collection when applying transcriptome analysis to clinical questions arising from individual patients.

Omics Technologies Posters - Thursday

PB3158*. Tagmented, Indexed, and Pooled followed by ChIP Sequencing (TIP-ChIP) to generate high-throughput multi-target ChIP-Seq Results

Authors:

J. Cayford, S. Traynor, **S. Ngoc Tran**, J. Poole; Active Motif, Carlsbad, CA

Abstract Body:

Epigenomic profiling approaches have become standard practice in understanding the molecular basis underlying clinically relevant disease states. The adoption of epigenetic analysis in translational medicine has been slow due to a complex workflow. Specific barriers include limited tissue availability, the laborious nature of chromatin generation, and the amount of cellular material needed for a single chromatin immunoprecipitation (ChIP) reaction. These issues are further compounded by the fact that single-plex reactions provide insufficient information to fully understand the epigenetic landscape of disease states. To address these issues, we have built upon previous concepts such as it-ChIP, mint-ChIP, ATAC-Seq, and many others which have improved the workflow and allowed for methods of probing the epigenetic landscape in cells. Here we introduce Tagmented, Indexed, and Pooled ChIP-Seq (TIP-ChIP). TIP-ChIP was developed to rapidly complete ChIP-Seq on 96 samples for multiple epigenetic and transcription factor marks. To maintain the current gold standards in the field, formaldehyde-fixed cells are tagmented with Tn5 to incorporate unique barcodes into each sample. After tagmentation and chromatin solubilization with a brief sonication, the samples are pooled and prepped for a standard ChIP-Seq reaction. The power of the technique lies in the ability to remove a fraction of the pooled sample and complete a single IP reaction with a small number of cell equivalents of all 96 starting samples which allow for numerous epigenetic marks to be in a multiplexed fashion. If four epigenetic targets were to be selected, it would be possible to complete the equivalent of 384 ChIP-Seq experiments over a three-day period, drastically improving the ability to test large cohorts of samples.

As clinical and translational research moves to incorporate epigenomics in the pursuit of personalized medicine, the innovation of rapid generation of ChIP-Seq data with fixed samples will enable many more samples to be tested in a rapid manner. Combined with a quicker workflow, higher throughput, and cheaper sequencing allows for this technique to be utilized by a wider audience than previously possible.

Omics Technologies Posters - Wednesday

PB3159. Targeted sequencing of a 1 Mb carrier screening panel using molecular inversion probes.

Authors:

J. Amaral, P. Saunders, A. Patel, J. Vieira, G. Porreca, E. Boyden; Molecular Loop BioSci.s, Inc, Woburn, MA

Abstract Body:

Common methods of target capture for next generation sequencing (NGS) include hybridization capture, which suffers from a complicated and labor-intensive workflow, and PCR capture, which is typically less onerous but prone to poor uniformity and allele dropout. We have developed and optimized an alternative method based on molecular inversion probes (MIPs) that combines robust data quality, enabled by dense bi-stranded MIP tiling that ensures high coverage uniformity and resistance to allele dropout, with a highly scalable and easily automatable protocol that may be conducted using either a single-day or overnight workflow. Historically MIP capture panels have commonly targeted relatively small regions with modest numbers of probes; to demonstrate that neither target size nor probe number is technologically limited, we developed and evaluated a full gene expanded carrier screening (ECS) panel that captures over 1 Mb of coding exons, splice sites, and 5' UTRs for 339 genes, using more than 43,000 MIPs with an average of 5 MIPs per target base. The superior performance of this panel shows that even full exome capture may be feasible, which would greatly improve on the simplicity and scalability of current methods without compromising data quality.

Omics Technologies Posters - Thursday

PB3160. Targeted sequencing of native telomeres reveals patterns of telomere length and subtelomere methylation at single chromosome resolution in human cells

Authors:

S. Hickey¹, A. Drong², P. Rughani¹, C. Tyer¹, X. Dai², J. Baulaurier¹, S. Juul²; ¹Oxford Nanopore Technologies Inc, San Francisco, CA, ²Oxford Nanopore Technologies Inc, New York, NY

Abstract Body:

Background: Telomeres are critical for genomic stability during replication, and the mechanisms for telomere maintenance are important in cancer progression and aging. The investigation of telomerase (TERT) -dependent and -independent mechanisms, such as Alternative Lengthening of Telomeres (ALT), is limited by the repetitive and homologous nature of telomeric regions within individual chromosome arms. Single-molecule long-read nanopore sequencing of native telomeres allows direct measurement of telomere length and alignment to specific chromosome ends, as well as differential methylation profiling and variant calling in human cells. **Methods:** We developed a novel approach for direct telomere targeted sequencing, without amplification, using a 3' overhang telomere hybridization probe coupled to nanopore sequencing adapters. The resulting full-length telomeric reads were basecalled using a custom telomere basecalling model trained using *Bonito*, filtered and aligned to the telomere-to-telomere human reference CHM13v2.0. Next, we used the methylation-calling tool *Remora* to characterize raw-read CpG methylation and ran *Jellyfish* to identify variants of the canonical TTAGGG human telomeric repeat. **Results:** We specifically determined the telomere lengths of individual chromosomes in multiple human cancer cell lines, at single chromosome resolution. The approach yielded a 30-fold increase of mapped telomeric reads compared to whole genome sequencing. Cancer cells with TERT and ALT mechanisms, as well as tumor/normal matched cells, can be distinguished based on their telomere length profiles and distributions. Additionally, differential methylation was observed at CpG islands within the subtelomeric regions at specific chromosome arms. Lastly, telomeric repeat variant calling revealed an increase in non-canonical telomeric variants in cancer relative to normal cells. **Conclusion:** Here we demonstrate a robust, single molecule sequencing method for characterizing telomeres, including length measurements, CpG methylation, and variant information. Single-molecule, chromosome-specific telomere characterization gives a clearer overview of telomere biology and provides information for potential biomarker discovery and clinical interventions.

Omics Technologies Posters - Wednesday

PB3161. Targeted transcriptome sequencing enables exponential scaling of combinatorial barcoding.

Authors:

V. Tran¹, P. Matulich¹, E. Papalexi², R. Koehler¹, C. Roco¹, A. Rosenberg¹; ¹Parse BioSci.s, Seattle, WA, ²Parse BioSci.s Inc., Seattle, WA

Abstract Body:

In the thirteen years since its inception, single-cell RNA-sequencing (scRNA-seq) has rapidly spread across multiple fields of research, leading to many new discoveries. As technologies have matured, the number of cells that can be processed in a single experiment has seen exponential growth, with workflows now assaying up to one million cells in an individual experiment. While high throughput sequencing methods have facilitated the discovery and characterization of various cell types, sequencing costs can be prohibitively high for routine use. Many applications of scRNA-seq are focused on cell type identification, gene regulatory networks, or biomarker discovery. These applications often do not require surveying the entire transcriptome, but rather require the interrogation of specific sets of well-characterized genes. In these cases, sequencing the entire transcriptome is unnecessary and adds substantial sequencing costs to the project. To increase throughput and minimize sequencing costs, the development of targeted gene enrichment methods is required. Building on Evercode™ Whole Transcriptome's technology of split-pool combinatorial barcoding, we applied a method to enrich a subset of genes in the final single cell sequencing libraries. To illustrate the power of our technology, we enriched a whole transcriptome library of 1 million peripheral blood mononuclear cells (PBMCs) using gene panels that targeted immune cell markers and pathways. The percent of reads on target increased from as low as 1.6% in the whole transcriptome libraries to 70% in the targeted libraries. The unenriched whole transcriptome library of these 1 million cells were sequenced on a NovaSeq using an S4 flow cell. In contrast, a gene panel library enriching approximately 1,000 genes was sequenced with NextSeq 550 using a High Output flow cell (generating ~500 mean reads/cell). Despite a nearly ten-fold reduction in sequencing reads, the resulting clustering yielded very high concordance of cell type identities when unenriched and enriched libraries were compared. This modular enrichment strategy allows researchers to use either custom or preexisting gene enrichment panels with any gene subset of interest. This approach enables researchers to simultaneously reduce sequencing costs while drastically scaling up the number of cells and samples across experiments.

Omics Technologies Posters - Thursday

PB3162. Tempo: an unsupervised Bayesian algorithm for circadian phase inference in single-cell transcriptomics

Authors:

B. Auerbach¹, G. A. FitzGerald¹, M. Li²; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

The circadian clock is a 24-hour cellular timekeeping mechanism that temporally regulates human physiology. Answering several fundamental questions in circadian biology will require joint measures of single-cell circadian phases and transcriptomes. However, no widespread experimental approaches exist for this purpose. While computational approaches exist to infer cell phase directly from single-cell RNA-sequencing (scRNA-seq) data, existing methods yield poor circadian phase estimates, and do not quantify estimation uncertainty, which is essential for interpretation of results from highly sparse scRNA-seq data. To address these unmet needs, we developed Tempo, a Bayesian variational inference approach that incorporates domain knowledge of the clock and quantifies phase estimation uncertainty. Through simulations and analyses of real data, we demonstrate that Tempo yields more accurate estimates of circadian phase than existing methods and provides well-calibrated uncertainty quantifications. We further demonstrate that these properties generalize to the cell cycle. Tempo will facilitate large-scale studies of single-cell circadian transcription.

Omics Technologies Posters - Wednesday

PB3163. Temporal motif discovery in biological interaction networks

Authors:

A. Jazayeri; Washington Univ. in St. Louis, St. Louis, MO

Abstract Body:

Consideration of the time aspect is essential in monitoring cellular signaling and processes, identifying functional regions within signaling genes, and investigating transcription regulation and phenotypic evolution. With the revolution in high-throughput sequencing technologies in recent years, collecting and analyzing the temporal biological structure and expression data have attracted increasing attention. On the other hand, one of the fundamental analysis and mining tools used for biological network analysis and interactomics is motif discovery. This problem, also known as frequent subgraph mining, aims to identify interaction patterns observed more frequently in the original network than in the randomized version of the same network. However, consideration of the temporality in frequent pattern mining and motif discovery increases the computational complexities associated with graph and subgraph isomorphism problems. Therefore, the typical approach adopted by the algorithms proposed for motif discovery in temporal networks is to transform the temporal networks into a series of networks where each network represents an aggregated version of interactions recorded over the inter-network intervals. In this study, we show that even though this type of representation reduces the computational complexities, it negatively impacts the expressiveness of the network data and, consequently, the derived insight from detected frequent patterns. Therefore, we develop a novel lossless model for temporal network representation. By developing novel definitions for graph and subgraph isomorphism for temporal networks, we create a series of algorithms for identifying motifs in a data set of heterogeneous temporal networks. We applied the proposed algorithms to different simulated data sets of interactomes representing gene-gene interactions from mice models' and human tissues' biological expression data. The results show that the patterns detected by the proposed algorithms are significantly different from patterns detected by the common approaches proposed in the literature for motif discovery, whether the networks are considered static, or a sequential representation of the networks is adopted.

Omics Technologies Posters - Thursday

PB3164. The impact of ageing on the transcriptome profile of human germline and somatic cells

Authors:

M. Pham, L. Chappell, Y. Hooks, M. Stratton, R. Rahbari; Wellcome Sanger Inst., Cambridge, United Kingdom

Abstract Body:

Male germline cells have been reported to acquire considerably fewer mutations compared to the somatic cells. A few possible explanations for this phenomenon have been proposed, including a lower rate of stem cell turnover and an increase in transcription-coupled repair. However, evidence purported to support these hypotheses remains inconclusive. Thus, our study focuses on other potential processes influencing mutations, particularly transcriptional alterations in the human testes, to interrogate how the germline cells might be protected from mutations during ageing.

We applied a laser capture microdissection (LCM) approach to collect the germline cells from seminiferous tubules of 16 normal testes, and the somatic cells from colonic crypts of eight normal colons (as an example of somatic counterparts) from the same donors (age range 21-71). A total of 192 small pools of whole transcriptome data were generated from those LCM patches. This data enables us to investigate the gene expression changes, especially those related to the DNA repair genes, as well as the effect of ageing on transcriptome across the two tissues. Additionally, we generated single-cell whole transcriptome data from ~2000 germline cells from the above donors using a novel single-cell sorting method, which allows isolation of whole cells from frozen tissues. At least four cell populations during spermatogenesis, including spermatogonial stem cells, primary spermatocytes, secondary spermatocytes and spermatids, could be isolated with this method. We then used this single-cell RNA data to demultiplex LCM transcriptome data, reconstruct developmental trajectories in testicular tissues and infer the effect of ageing on repair pathways in male germline cells. Ultimately, our detailed molecular study of age-related transcriptional changes provides mechanistic insights into how the germline cells are protected during ageing.

Omics Technologies Posters - Wednesday

PB3165. The impact of insert length on variant calling quality in whole genome sequencing.

Authors:

B. Lajoie, A. Altomare, K. Blease, R. Kelley, E. Miller, J. Moreno, C. Thompson, J. Zhao, S. Kruglyak; Element BioSci.s, San Diego, CA

Abstract Body:

Accurate variant calling is critical for whole genome sequencing applications, including rare disease and oncology. The availability of high-quality truth sets provided by NIST has enabled various benchmarking efforts across both sequencing platforms and NGS algorithms. Based on these benchmarking results, we have an understanding of the most accurate variant calling methods, the impact of greater read length, and the properties of the remaining difficult regions. However, we do not yet have a careful examination of the impact that varying insert length distributions have on accuracy. This is in part because amplifying long inserts is challenging for certain sequencing chemistries.

Our hypothesis is that longer inserts, like longer reads, could improve alignment in certain genomic regions, thus improving overall variant calling accuracy. The intuition is that a short insert may match several repetitive locations in the genome, whereas a longer insert may have at least one end outside of the repetitive region, thus providing an anchor for the alignment.

We began with a simulation study to determine the appropriate conditions to try experimentally. The open source NEAT simulation framework (<https://github.com/ncsa/NEAT>) was first used to generate inserts with typical mean insert length of ~350 bp. Benchmarking of the simulated reads produced SNP and Indel metrics similar to what we observed from benchmarking public sequencing data from the Precision FDA Truth Challenges. We then synthetically varied the insert length distributions and repeated the benchmarking. As we increased the insert length, the total number of variant calling errors decreased, particularly in the false negative category for positions that span the “difficult regions.” This supported our initial hypothesis.

Next, we generated sequencing libraries with mean insert sizes ranging from 350 bp to 1000 bp. We then benchmarked alignment and variant calling quality as a function of insert length. Consistent with the simulation, we saw a higher fraction of reads aligned at high mapping quality and lower number of total variant calling errors as a function of insert length. We conclude that some of the benefits of longer reads can be attained through the use of longer inserts.

Omics Technologies Posters - Thursday

PB3166. Tissue specific transcript and protein isoforms in the human neural retina

Authors:

T. Riepe¹, M. Stemerding¹, R. Salz¹, E. de Vrieze¹, J. Gloerich¹, B. Ferrari², S. Ferrari², S. Roosing¹, F. P. M. Cremers¹, P. A. C. Hoen¹; ¹Radboudumc, Nijmegen, Netherlands, ²Fondazione Banca degli Occhi del Veneto, Venice, Italy

Abstract Body:

The retina is the light sensitive tissue at the back of the eye. It converts light into electric stimuli that can be interpreted by the brain. Like the brain, the retina is known to be enriched for alternative splicing. However, our knowledge about the transcript and protein isoforms expressed in the retina is not complete. Genetic defects in genes that code for proteins involved in normal retina function can cause inherited retinal diseases (IRDs). IRDs affect about 1 in 2,000 individuals worldwide and cause vision loss. It is estimated that approximately 1/3 of causative variants in IRD-associated genes disrupt splicing. Several previous studies revealed retina-specific splicing factors and isoforms. A more complete atlas of the retina transcriptome and proteome will contribute to the identification of IRD-associated variants, disease mechanisms and treatment options. In this study, we analyzed three human post-mortem neural retina samples with PacBio long read RNA-sequencing and mass-spectrometry-based proteomics to create an atlas of retina transcript and protein isoforms. We identified more than 125,000 unique isoforms, which resulted in more than 55,000 Open Reading Frames (ORFs). More than half of the isoforms were classified as novel, and more than two-third of the identified transcript isoforms from IRD-associated genes were novel. Both novel exons and exon skipping events were found in IRD-associated genes. Using a PacBio GENCODE hybrid database, more than 15,000 peptides from more than 2,500 proteins were identified. 20 of these peptides were novel peptides that confirm a novel transcript isoform. One interesting novel peptide mapped to an intron retention event in *ATP1A3*, which is a candidate gene for autosomal dominant cone-rod dystrophy. There are several reasons why we detect many more retina-specific transcripts than peptides. First, proteomics coverage is still limited and chances to detect a peptide that maps uniquely to a novel transcript are low. Second, a lot of transcript variation affects noncoding regions, without any effect on the ORF, but with possible consequences for translational efficiency and protein expression levels. Third, not all novel ORFs are translated into stable proteins. To conclude, the retina cell atlas provides novel insights into the retina transcriptome and proteome and can be used as reference for future IRD research. Moreover, it demonstrates the need to study tissue specific transcriptomes and proteomes.

Omics Technologies Posters - Wednesday

PB3167. Transcriptome analyses of congenital heart disease tissue from participants with Trisomy 21.

Authors:

S. Morton^{1,2}, D. M. McKean³, P. Narayan³, D. Reichart³, D. M. DeLaughter³, H. Wakimoto³, J. M. Gorham³, D. Quiat¹, Q. Zhang³, V. Strohmenger³, K. Brown³, B. McDonough³, J. A. Willcox³, C. A. Gifford⁴, Y. de Soysa⁵, M. Jang³, B. G. Benoit⁵, S. R. DePalma³, T. Ward³, D. Bernstein⁶, M. Brueckner⁷, B. D. Gelb⁸, E. Goldmuntz⁹, R. P. Lifton¹⁰, J. W. Newburger¹, A. E. Roberts¹¹, H. Yost¹², M. Tristani-Firouz¹², A. G. Robling¹³, R. N. Mitchell³, W. K. Chung¹⁴, R. Kim¹⁵, J. Cleveland¹⁶, J. G. Seidman³, C. E. Seidman³; ¹Boston Children's Hosp., Boston, MA, ²Harvard Med. Sch., Boston, MA, ³Harvard Med. Sch., Boston, MA, ⁴Gladstone Inst.s, San Francisco, CA, ⁵Univ. of California San Francisco, San Francisco, CA, ⁶Stanford, Stanford, CA, ⁷Yale, New Haven, CT, ⁸Icahn Sch. of Med. at Mount Sinai, New York, NY, ⁹Children's Hosp. of Philadelphia, Philadelphia, PA, ¹⁰Rockefeller Inst., New York, NY, ¹¹Boston Children's Hosp., Boston, MA, ¹²Univ. of Utah, Salt Lake City, UT, ¹³Indiana Univ., Indianapolis, IN, ¹⁴Columbia Univ., New York, NY, ¹⁵Cedar Sinai Hosp., Los Angeles, CA, ¹⁶Univ. of Southern California, Los Angeles, CA

Abstract Body:

Introduction: Trisomy 21 (T21), a prevalent human genetic syndrome occurring in 1:600 live births, is characterized by congenital heart defects (CHD), craniofacial dysmorphologies, neurocognitive, pulmonary, endocrine abnormalities, and increased risks for comorbidities. Despite a definitive genetic cause, the molecular mechanisms by which T21 perturbs development and homeostasis of human tissues remains poorly defined.

Hypothesis: We hypothesized altered cardiac gene expression in T21 might provide clues to the pathogenesis of its clinical manifestations.

Methods: We performed RNA sequencing (RNAseq) of CHD cardiac tissue from 49 T21 and 226 euploid (eCHD) patients aged 0-19.8 years and compared genome-wide transcript expression between groups. T21 individuals were equally likely to be male (47% vs 57%, P=0.19) but younger (mean 1.4 vs 3.1 years, P=0.02) than eCHD individuals. Septal defects were the most common types of CHD (65 eCHD and 38 T21), and atrioventricular canals were present in more T21 patients (24 T21 vs. 10 eCHD, OR 20.7, p<2.2E-16). The most common tissue samples was right atrium (83 eCHD and 37 T21). Using single nuclear RNA sequencing (snRNAseq) on an independent set of ten T21, 13 eCHD and 12 healthy adult hearts we assessed expression within cell lineages.

Results: In addition to chr21 genes, T21 increased *SOST* (chr17) expression 11-fold above eCHD tissues (P=1.2E-8). *SOST* and *ZNF467* (chr7), a *SOST* transcriptional activator, were co-expressed (r=0.32; P=1.7E-7). In endothelial cells, *SOST* levels were 12-fold higher in T21 than euploid cells (P=2.7E-27) and were correlated with ch21 *DSCAM* (Down syndrome cell adhesion molecule; r=0.77, P=1.9E-05).

Conclusion: Wnt signaling is critical for atrioventricular canal, craniofacial and long bone development, maintenance of bone mass, and endothelial regulation of pulmonary vascular homeostasis. Consequently, chronically increased tissue *SOST* expression that inhibits Wnt activities might promote prototypic T21 manifestations: atrioventricular septation defects, diminutive craniofacial bones, shortened stature, low bone mineral density and increased risks for osteoporosis and pulmonary hypertension. These findings imply therapeutic potential of anti-sclerostin antibodies for multiple T21 phenotypes.

Omics Technologies Posters - Thursday

PB3168. Transcriptome sequencing of RNA isolated from small volumes of blood stabilized in Tempus solution: a technical assessment of different extraction methods and DNase treatment

Authors:

S. Tomei, H. Manjunath, F. Vempalli, L. Mathew, L. Liu, L. Wang, G. Wang, K. Wang, O. Soloviov, S. Lorenz, M. Kalikiri; Sidra Med. and Res. Ctr., Doha, Qatar

Abstract Body:

Transcriptome profiling of human whole blood is used to discover biomarkers of diseases and to assess phenotypic traits. RNA sequencing technologies offer many advantages for transcriptomic profiling over other technologies, including the ability to detect novel transcripts, a wide dynamic range of transcript detection, high specificity and sensitivity and the ability to detect low-abundance transcripts. Recently, finger-stick blood collection systems have allowed a less invasive and quicker collection of peripheral blood that does not necessarily require medical infrastructures. This sampling of small volumes of blood offers practical advantages, allowing large-scale projects. The quality of gene expression data is strictly dependent on the steps used for the sample collection, extraction, preparation and sequencing. Here we have: i. compared the manual and automated RNA extraction of small volumes of blood using the Tempus Spin RNA isolation kit (ThermoFisher Scientific, USA) and the MagMAX™ for Stabilized Blood RNA Isolation Kit (ThermoFisher Scientific, USA), respectively; and ii. assessed the effect of TURBO DNA Free treatment on the transcriptomic data of RNA isolated from small volumes of blood. RNA Libraries were prepped using the QuantSeq 3' FWD mRNA-Seq Library Prep Kit (Lexogen). Library QC was performed on the LabChip GXII and quantified using KAPA Library quantification kit by qPCR. Libraries were pooled and sequenced on the Illumina HiSeq4000 system. The data QC was performed as recommended by Illumina. Reads were mapped to the human genome GRCh38.p13 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.28, Dec 2013) using STAR_2.6.1d aligner, and featureCounts v2.0.0 was used to generate the raw counts. We used DESeq2 (v1.32.0) to normalize read counts with standard settings. Normalized data was transformed using variance-stabilizing transform (VST) and removed batch effect using limma::removeBatchEffect from Lima package (v3.48.2). Heatmaps, correlation matrices and PCA plots were generated as relevant. Transcriptomic profiles were overall consistent, however the samples isolated manually displayed a higher variability in the transcriptomic data as compared to the other samples. The TURBO DNA Free treatment affected the RNA samples negatively, decreasing the RNA yield and reducing the quality and reproducibility of the transcriptomic data. We conclude that automated extraction systems should be preferred over manual extraction systems for data consistency, and that the TURBO DNA Free treatment should be avoided when working on RNA samples isolated manually from small volumes of blood.

Omics Technologies Posters - Wednesday

PB3169. Transcriptome signature of sodium intake in and links to cardiovascular traits

Authors:

A. Gaye; NHGRI, Bethesda, MD

Abstract Body:

Background: In average, Americans consume more than 3,400 milligrams of sodium per day, an amount much higher than recommendations in the US 2015-2020 Dietary Guidelines. Studies have shown that dietary intake high in sodium, fat and sucrose are strongly related to an increased risk for cardiovascular disease (CVD). However, the molecular mechanisms by which these nutrient impacts CVD are not fully understood. The aim of this analysis is to conduct an unbiased transcriptome analysis in African Americans (AA), a population with disproportionate burden of CVD. **Methods:** The analysis is a cross-sectional study of whole blood transcriptome mRNA sequencing (mRNA-seq) of 472 AA from the MHGRID and GENE-FORECAST studies. Random Forests (RF) was used to define a transcriptome profile empirically related to salt intake, and evaluate the relationships between that profile and other diet variable (fat, sucrose and calories) and cardiovascular traits including hypertension status (HTN), systolic and diastolic blood pressure (SBP and DBP) and kidney function (eGFR). Weighted Gene Co-expression Network Analysis (WGCNA) was subsequently conducted to identify clusters of co-expressed genes within the transcriptome profile defined by RF. **Results:** From the 13,359 protein coding transcripts that passed QC filters, RF defined a transcriptome profile of 816 mRNA that classified subjects with high vs low sodium intake with an area under the curve (AUC) of 0.82. That transcriptome profile classified fat, sucrose, calorie intake and the cardiovascular traits with AUC=76 (fat), AUC=67 (sucrose), AUC=0.77 (calorie), AUC=0.72 (HTN), AUC=0.71 (SBP), AUC=0.67 (DBP), AUC=0.71 (eGFR). WGCNA analysis identified 3 network modules that include genes such as ALMS1, RAP1, CXCR4, TRAF3, and IL16 involved in pathways that regulate vascular tone, flow, remodeling and sustained high blood pressure. **Conclusions:** This study employed machine learning predictive models to examine the blood transcriptome and define molecular signatures of mRNA transcripts associated with sodium intake and cardiovascular traits, in one of the largest mRNA-seq sample set of AA to date. The findings provide new insights into biomarkers and molecular pathways that may mediate the effects of sodium, fat and sucrose on cardiovascular health.

Omics Technologies Posters - Thursday

PB3170. Transcriptomic analysis of circulating endothelial colony-forming cells in patients with sickle cell anemia and ischemic stroke.

Authors:

J. Nicoliello Pereira de Castro, S. Matilde da Silva Costa, A. Lima Camargo, M. Tomie Ito, B. Batista de Souza, V. de Haidar e Bertozzo, T. Adalton Rosa Rodrigues, R. Casagrande Saez, M. Castro Ozelo, F. Ferreira Costa, M. Barbosa de Melo; UNICAMP, Campinas, Brazil

Abstract Body:

Background: Cerebrovascular alterations such as ischemic stroke (IS) are common and severe complications of Sickle Cell Anemia (SCA), being fatal in 15% of cases. The complex pathophysiology encompasses an interaction of hemolysis, thromboinflammation, endothelial activation and vasoconstriction that culminate in vaso-occlusion prior to ischemic accidents. In this context, the vascular endothelium has a pivotal role in this cascade of events. Objectives: To evaluate the differential gene expression profile in Endothelial Colony Forming Cells (ECFCs) from SCA patients with IS compared to SCA patients without IS and investigate the relevant biological pathways in ischemic stroke. Methods: RNA samples from ECFCs of SCA patients with IS (N=4) and SCA patients without IS (N=4) were sequenced by RNA-Seq. Normalization and differential gene expression analysis were performed by the *edgeR* package in R. Gene Ontology (GO) analysis for Biological Process category of the Differentially Expressed Genes (DEGs) was performed using *DAVID* tool. Next, we applied *stringApp* plugin in Cytoscape tool to construct protein-protein interaction network (PPI) of genes present in enriched GO terms, and *cytoHubba* plugin to identify the most hub-genes of the network. Results: We found 2,469 DEGs: 1,836 superexpressed and 633 DEGs underexpressed. There were 23 significantly enriched GO terms, such as defense response, inflammatory response, immune response and taxis. We identified 1,519 protein-coding genes in the main PPI network, with 1,556 interactions. The top 5 hub-genes were: *AKT1* (AKT serine/threonine kinase 1, Log₂FC = 2.57), *HRAS* (HRas proto-oncogene, GTPase, Log₂FC= 3.71), *PIK3R1* (Phosphoinositide-3-kinase regulatory subunit 1, Log₂FC = -2.13), *CDC20* (Cell division cycle 20, Log₂FC = 2.62), *MAPK11* (Mitogen-activated protein kinase 11, Log₂FC = 2.71). These protein-coding genes are predominantly present in signaling cascades of cell proliferation, inflammation and angiogenesis. *AKT1*, the most relevant protein in our network, is a serine/threonine kinase of the Akt pathway, activated in several physiological and pathological mechanisms. In this context, this protein has already been identified as a potential interaction target for pharmacological compounds of ischemic stroke treatment. Conclusions: Next-generation sequencing technologies are strategic tools for uncovering the complexity of several diseases and their complications. Thus, RNA-Seq analysis is a relevant approach that may enable the discovery of new molecular pathways of biological importance for IS in SCA.

Omics Technologies Posters - Wednesday

PB3171. Transcriptomic analysis reveals long-term effect of SARS-CoV-2 infection on host genetic regulation

Authors:

W. Zhu¹, H-H. Chen², H. Polikowsky¹, D. Shaw¹, L. Petty¹, D. Kim³, X. Zhang⁴, M. Yaser⁵, P. Sharma⁶, K. Young⁷, J. McCormick⁸, S. Fisher-Hoch⁸, K. North⁷, C. Huff⁹, J. Below¹⁰; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³UNC-Chapel Hill, Chapel Hill, NC, ⁴Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁵Univ. of North Carolina Chapel Hill, Chapel Hill, NC, ⁶Univ. of North Carolina, Chapel Hill, NC, ⁷Univ North Carolina, Chapel Hill, NC, ⁸The Univ. of Texas, Houston, TX, ⁹The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, ¹⁰Vanderbilt Univ Med Ctr., Nashville, TN

Abstract Body:

Background: COVID-19, caused by severe acute respiratory syndrome virus (SARS-CoV-2), is highly infectious and has led to a worldwide pandemic since 2019. Over 500 million cases and 6 million deaths have been reported globally as of today. Although 80% of cases only experience mild to moderate symptoms, approximately 5% patients developed critical symptoms. Moreover, beyond the typical recovery timespan, ~2.3% patients suffered from lingering symptoms of COVID-19 for weeks to years. This Post Acute Coronavirus Syndrome (or so-called “Long COVID”) is a multisystem disorder including a wide range of symptoms, commonly affecting the respiratory, cardiovascular, and hematopoietic systems. **Methods:** To investigate acute and long-term impacts of COVID-19 on transcriptomic regulation, we sequenced whole blood RNA samples isolated both before and after SARS-CoV-2 infection for 215 COVID-19 cases confirmed with presence of antibody against the nucleocapsid (to differentiate antibodies present due to vaccination from natural infection), and 188 sex-, age- and interval-matched healthy controls (antigen-negative) in Cameron County Hispanic Cohort. Another 220 single timepoint RNA samples acquired from matched COVID-19 cases and controls were also included in our analysis. We applied linear mixed-effects models adjusted for sex, age, and PEER factors to compare changes of gene expression over time between individuals with incident COVID-19 and controls. **Results:** Our analysis identified 9 genes whose change in expression is significantly associated with incident COVID-19 after multiple testing correction ($p < 2.86 \times 10^{-6}$): *ABCA4*, *ANGPTL6*, *ARHGEF10*, *KLF15*, *LINC00431*, *LOC101928514*, *MSLN*, *OLFM4* and *RPL3L*. Further analysis restricted to symptomatic infections revealed another 14 significant genes: *RNF5P1*, *OVCH1.AS1*, *LOC100506159*, *LILRB5*, *KRT79*, *IQSEC3*, *CPA5*, *CTXN2*, *CRLF2*, *SIGLEC12*, *CNTNAP3P2*, *CCL23*, *HBB* and *ZNF890P*. Prior evidence links several genes to immune processes or COVID-19 symptoms. Based on functional annotation by DAVID, *CCL23*, *LILRB5* and are involved in immunoregulatory and inflammatory processes. Our *RPL3L* finding validates a recent report from re-analysis of a cross-sectional differential expression study of COVID-19 (Vastrad 2020). *HBB*, a gene known to cause sickle cell disease, is significantly associated with pneumonia susceptibility (Chen 2021). **Conclusion:** In summary, our study profiled transcriptomic changes associated with SARS-CoV-2 infection and symptomology, identifying long-term effects of COVID-19 on the expression profile of 23 genes. These results provide compelling insight into COVID-19 pathophysiology.

Omics Technologies Posters - Wednesday

PB3172. Ultrasensitive Duplex Sequencing for Diverse Applications on the Singular Genomics G4 Platform.

Authors:

E. Schmidt, **D. Fitzgerald**, P. Van, M. J. Hipp, C. C. Valentine, J. Salk, J. Higgins; TwinStrand BioSci.s, Seattle, WA

Abstract Body:

Duplex Sequencing (DS) is an extremely accurate DNA sequencing technology which has broad applications across life science and medicine where detection of low frequency mutations is critical. The method relies on double-stranded consensus-making to achieve an error rate below one-in-ten-million. Nearly all DS studies to date have been carried out on Illumina sequencers. The Singular Genomics G4 platform is a newly available instrument with novel chemistry designed to deliver ~2x faster sequencing results. To assess relative platform performance, we benchmarked several TwinStrand DuplexSeq Assays on the Singular G4 platform vs. the Illumina NovaSeq 6000. To evaluate data yield over a range of inputs, libraries were seeded with 50-1000ng of high-quality human DNA and enriched with a small (2.4kb) control panel. Resulting peak DS depths ranged from 2,300-47,000 (G4) and 2,800-57,000 (NovaSeq), with the slight differences likely due to variation in experimental conditions. To evaluate the sensitivity and specificity of mutant detection across a range of VAFs, we prepared synthetic low frequency mixtures of DNA from cell lines carrying seven different cancer-associated *TP53* mutations (SNV & indels) spiked into DNA derived from a healthy donor (expected VAFs from 6E-3 to 9E-5). Spike-in mixes, as well as pure healthy donor DNA, were sequenced to an average DS depth of >50,000x across all exons of *TP53*. The measured mutant frequencies down to the lowest level spike-in matched that which was expected within variation predicted from Poisson sampling statistics on both platforms. Zero mutant counts at spike-in sites were identified in negative controls; thus the analytical performance of the assay was 100% sensitivity and 100% specific for detecting clones to below one-in-ten-thousand. To evaluate performance for measuring ultra-low frequency mutations, 200 ng of DNA from blood of a healthy donor was used to prepare libraries using TwinStrand's DuplexSeq Human Mutagenesis Assay (48 kb panel). A total of 70 unique mutant nucleotides out of 452,802,985 total duplex nucleotides (mutant frequency 1.03E-7 +/- 2.5E-8) and 51 mutant nucleotides out of 485,440,864 (mutant frequency 1.55E-7 +/- 3.2E-8) were respectively identified in the Illumina and Singular libraries. No statistically significant differences in overall mutant frequency or spectrum were observed between the platforms. In conclusion, three independent forms of analytical testing of TwinStrand Duplex Sequencing across extremes of DNA input, rare cancer clone detection, and ultra-sensitive mutagenesis applications yielded nearly identical performance on the Singular and Illumina platforms.

Omics Technologies Posters - Thursday

PB3173. Ultra-sensitive TCR/BCR clonotyping and immunophenotyping.

Authors:

A. Chenchik¹, M. Makhanov¹, T. Liu¹, D. Hu¹, K. Ghias¹, P. Diehl², L. Kobzik¹; ¹Cellecta, Inc., Mountain View, CA, ²Cellecta, Mountain View, CA

Abstract Body:

TCR and BCR repertoire profiling holds great potential for understanding disease mechanisms and for the development of new therapeutics in infectious disease, autoimmunity, and immuno-oncology. This potential could be greatly improved by combining information about receptor clonotypes with immunophenotypes of T and B cells. To facilitate these studies, we developed a novel technology for combined profiling of all human TCR and BCR variable regions and phenotypic characterization of immune cells. The developed TCR/BCR immunophenotyping method involves multiplex RT-PCR amplification and sequencing of CDR3 regions of TCR and BCR genes and a set of the most informative T- and B-cell phenotyping genes. Bioinformatic analysis of NGS data allows profiling of TCR/BCR clonotypes, and identification of major immune cell subtypes and their activation status. Preliminary studies indicate the assay has unparalleled throughput, sensitivity, and improved cost-effectiveness for high-throughput immunity biomarker discovery applications.

Omics Technologies Posters - Wednesday

PB3174. Undergraduate research assistant

Authors:

M. Williams¹, Y. Zhang¹, E. Perillo², E. Parsaee¹, G. Tremel², J. Jung², P. Chen¹, J. Flint¹, J. R. Ecker³, J. Beechem², C. Luo¹; ¹UCLA, Los Angeles, CA, ²NanoString Technologies, Seattle, WA, ³Salk Inst./HHMI, La Jolla, CA

Abstract Body:

Single-cell technologies have revolutionized the characterization of heterogeneous human tissues allowing unbiased census of constituent cell types and their transcriptomic and epigenomics signatures. However, the mapping of molecular signatures onto three-dimensional tissue structure remains highly challenging since most single-cell methods can only analyze disassociated cells or nuclei. We have developed photonic-indexing sequencing (pi-seq) strategies for in situ spatial barcoding with single-cell and subcellular resolution. The pi-seq strategy writes high complexity barcodes into the tissue using sequential ligation of DNA indices with the ligation reaction controlled by high-resolution patterned illumination. The rapid but lower-resolution mode of pi-seq enables spatial genomics with a ~200 μm resolution using a regular epi-fluorescence microscope and no other specialized equipment or reagent. The high-resolution mode of pi-seq uses 2-photon scanning microscopy to barcode single cells with exceptional spatial specificity. Compared to other spatial genomic platforms, pi-seq can be readily scaled to large tissue areas and allows the selection of cell-of-interest using immunofluorescent labeling as guides. The sequencing library generated by pi-seq is fully compatible with commercial short-read sequencing platforms.

Chromatin accessibility and cytosine modifications are well-established epigenomics marks playing critical roles in transcription regulation in normal and disease tissues. We have integrated pi-seq with well-established single-cell open chromatin profiling technique ATAC-seq, and single-cell methylome technique snmC-seq2 to develop methods for the spatial profiling of chromatin accessibility (pi-ATAC-seq) and methylcytosine (pi-mC-seq). These spatial epigenomic approaches have been applied to mouse brain slices to investigate the spatial diversity and organization of brain cell types, using the extensive scATAC-seq and snmC-seq2 profiles generated by the BRAIN initiative as the reference.

Omics Technologies Posters - Thursday

PB3175. Untargeted metabolomics profiling in patients with and without epilepsy.

Authors:

K. Oja^{1,2}, M. Ilisson², K. Reinson^{2,1}, K. Muru^{2,1}, T. Reimand^{2,1}, T. Esko¹, T. Haller¹, J. Kronberg¹, M. Wojcik^{3,4}, G. Michelotti⁵, A. O'Donnell-Luria⁶, S. Pajusalu^{2,1}, K. Ounap^{1,2}; ¹Univ. of Tartu, Tartu, Estonia, ²Tartu Univ. Hosp., Tartu, Estonia, ³Broad Inst. of MIT and Harvard, Boston, MA, ⁴Boston Children's Hosp., Boston, MA, ⁵Metabolon, Morrisville, NC, ⁶Boston Childrens Hosp. / Broad Inst. / HMS, Somerville, MA

Abstract Body:

Epilepsy is a central nervous system disorder with abnormal brain activity. Our understanding of its pathophysiology is still growing. In recent years, research has become more focused on finding biomarkers that would help with differential diagnosis and provide opportunities for the development of new anti-epileptic drugs. Metabolomics is a promising approach to understanding the pathophysiology of epilepsy. We devised a hypothesis free study comparing the metabolic profiles of persons with and without epilepsy. Untargeted metabolomics analysis was performed in two cohorts using ultra performance liquid chromatography (UPLC) instruments paired with mass spectrometry (UPLC/MS). First cohort consisted of 31 pediatric patients with suspicion of a genetic disorder from Tartu University Hospital, who remained unsolved after trio exome sequencing. The second cohort consisted of 824 adult participants from the Estonian Biobank, for whom untargeted metabolomics analysis data was available. We performed Welch's t-tests separately in the two cohorts to compare groups with and without epilepsy. P-values were adjusted to multiple testing using the Benjamini-Hochberg procedure. Epilepsy was present in eleven pediatric patients and in 37 Biobank participants according to their medical health records. The two cohorts had very different general characteristics (the median age was respectively 9 and 71 years). Most pediatric patients also had developmental delay (25/31), intellectual disability (21/31) and brain structural anomalies (17/31). The most relevant comorbidities in the Biobank cohort were disorders of lipoprotein metabolism and other lipidemias (288/824). Still, changes in lipid pathways were seen in both cohorts. In the pediatric cohort 18 significantly differing metabolites were detected (corrected p-value <0.05). These included sixteen different glycerophosphatidylcholines (GPC), dimethylglycine and eicosanedioate (C20-DC). In the Biobank cohort nine significantly altered metabolites were found, these were mainly triacylglycerides, which are also precursors in the GPC synthesis pathway, and one phosphatidylethanolamine. Although the two cohorts revealed different significantly altered metabolites, they both had changes mainly in the phospholipid pathways. Further research is required in order to understand, whether these changes could be used as biomarkers. Funding: Estonian Research Council grants PRG471, PSG774.

Omics Technologies Posters - Wednesday

PB3176. Using an NGS readout for high-throughput proteome-wide analysis in large population health studies.

Authors:

C. Lawley¹, L. Wik², N. Nordberg², J. Broberg², J. Björkesten², E. Assarsson², S. Henriksson², I. Grundberg², C. Westerberg², E. Liljeroth², A. Falck², M. Lundberg², L. Folksersen³, A. Malarstig⁴; ¹Olink, San Francisco, CA, ²Olink, Uppsala, Sweden, ³Nucleus Genomics, Ltd, New York, NY, ⁴Karolinska Inst., Stockholm, Sweden

Abstract Body:

Understanding the dynamics of the human proteome is crucial for identifying biomarkers to be used as measurable indicators for disease severity and progression, patient stratification, and drug development. The Proximity Extension Assay (PEA) is a technology that translates protein information into actionable insights across large samples sizes in both healthy and disease samples. The high-throughput nature of the assay is enabled by linking protein-specific antibodies to DNA-encoded tags that can be read out on a next generation sequencer. We have combined the PEA technology described above with automated sample preparation and a high-throughput sequencing readout for parallel measurement of ~3,000 proteins for up to 384 samples at a time, generating over 1 million data points per run. Characterizing the proteome alongside genetic and clinical data enables a pQTL framework to not only validate known clinical targets and identify new clinical targets but to also suggest repurposing opportunities of clinical candidates for new indications. Here we will summarize results where proteomics is impacting large population health studies (e.g., UK Biobank, SCALLOP) to advance precision and personalized medicine.

Omics Technologies Posters - Thursday

PB3177. Using combinatorial barcoding to simultaneously profile the transcriptome and immune repertoire of 1 million T cells.

Authors:

E. Papalexi, P. Matulich, V. Tran, G. S. Kim, D. Diaz, C. Roco, A. Rosenberg; Parse BioSci.s Inc., Seattle, WA

Abstract Body:

T cells are able to recognize and eliminate a wide variety of immunologic threats while maintaining self-tolerance. Pathogen recognition and clearance is ensured by a process called V(D)J recombination, during which a T cell obtains a unique set of V, D and J gene segments for all the chains (α and β or γ and δ) that make up its T cell receptor (TCR). Each recombined TCR detects a specific disease-associated antigen peptide, which triggers the appropriate adaptive immune response.

Understanding the relationship between TCR sequences and T cell activation during disease pathogenesis and progression can assist in the development of next-generation therapeutics with more favorable and sustainable outcomes. Recent advances in single-cell sequencing allowed for simultaneous profiling of TCRs and full transcriptomes leading to the characterization of key T cell populations with pathogen recognition and disease clearance capabilities. Despite their success, these methods rely on microfluidics devices or plate-based protocols with limited sensitivity and throughput (1,000s-10,000s of cells) making the study of disease-relevant T cells time consuming and costly.

To overcome these limitations, we have extended Parse's combinatorial barcoding technology (originally based on Split Pool Ligation-based Transcriptome sequencing or SPLiT-seq) to simultaneously characterize the TCRs alongside the full transcriptomes of up to 1 million T cells. Using this approach, we were able to recover paired TCR sequences in more than 75% of individual T cells together with their corresponding transcriptomes. We find that clonotype assignments are highly accurate with low rates of misassigned chains. We further obtained an accurate estimate of TCR repertoire diversity in 5 healthy donors and identified signatures that correspond to viral infection.

In summary, we report on the extension of our highly flexible and scalable combinatorial barcoding technology to allow researchers to profile up to 1 million of T cells in a single experiment and investigate their functional responses during infection, cancer, autoimmunity, or therapeutic interventions.

Omics Technologies Posters - Wednesday

PB3178. Using single cell CRISPR/dCas9-based regulatory element screening to dissect complex genetic loci.

Authors:

L. Bounds¹, A. Barrera¹, P. Giusti-Rodriguez², T. E. Reddy¹, Y. Li³, M. I. Love⁴, P. F. Sullivan³, A. A. Allen⁵, G. E. Crawford¹, C. A. Gersbach¹; ¹Duke Univ., Durham, NC, ²Univ. of Florida Coll. of Med., Gainesville, FL, ³Univ North Carolina, Chapel Hill, NC, ⁴UNC-Chapel Hill, Chapel Hill, NC, ⁵Duke Univ, Durham, NC

Abstract Body:

The Major Histocompatibility (MHC) Locus is the most SNP-dense region in the human genome and has been linked with >100 polygenic disorders. However, it remains unknown which variants affect gene expression. To dissect cis-regulation within the locus, we performed noncoding CRISPR/dCas9-based regulatory element screening with single-cell RNA-sequencing readout and identified putative regulatory element (pRE)-gene pairs in three diverse cell types: human induced pluripotent stem cells (hiPSCs), hiPSC-derived neural progenitor cells (hNPCs), and the K562 cell line.

We designed a sgRNA library targeting 581 elements with 150 TSS-targeting positive controls and 1,273 nontargeting negative controls (12,723 sgRNAs). We engineered hiPSCs, hNPCs, and K562 cells to express the repressor dCas9-KRAB, and hiPSCs and hNPCs to express the activator dCas9-p300. We transduced each cell line with the sgRNA library at a multiplicity of infection greater than one and profiled 1.2 million single cell transcriptomes seven to nine days post-transduction.

We performed differential expression testing between all cells with a given sgRNA versus all cells in which a different sgRNA was observed and recovered >170 significant pRE-gene links in each cell type. 16.3% of elements demonstrated regulatory function in all cell types when perturbed with dCas9-KRAB. Of these regions, 45 pREs regulate the same gene, and the change in gene expression was correlated across cell types. pRE-gene links present in all cell types versus cell-type specific links span shorter chromosomal distances and lead to greater changes in gene expression upon perturbation.

In all three cell types, we identified pREs that responded to perturbation but did not overlap a DHS in the corresponding cell type. In dCas9-KRAB NPCs, significant pREs that overlapped a DHS had greater accessibility than the NS pREs that overlapped a DHS. In the activation screen, there was no significant difference in the signal between the significant and NS pREs that overlap DHSs. We also observe pREs that do not overlap DHSs. There is no significant difference in accessibility of these regions and the “closed” pRE-gene links have a greater median pair distance versus those in accessible regions. We find similar trends in the other cell types. Taken together, our results highlight the importance of performing both repression and activation experiments as local chromatin context may influence the perturbation modality.

Finally, we demonstrate that hybridization-based targeted enrichment of the gene expression libraries increases on-target reads >160-fold and achieves >90% sequencing saturation with ~5,000 reads per cell.

Omics Technologies Posters - Thursday

PB3179. Using single nuclei multiomics and spatial transcriptomics to identify drivers of early pulmonary fibrosis

Authors:

A. Vannan¹, C. Taylor², M-I. Chung¹, L. Peter¹, M. Salisbury², T. Blackwell², J. Kropski², N. Banovich¹; ¹Translational Genomics Res. Inst., Phoenix, AZ, ²Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Pulmonary fibrosis (PF) is a heterogeneous clinical syndrome that represents the end-stage of chronic interstitial lung diseases (ILD). The most common (20%) and severe form of PF is idiopathic pulmonary fibrosis (IPF), which typically leads to respiratory failure within 5 years of diagnosis. While current therapies for IPF can slow disease progression, they have not been shown to improve survival or quality of life. The limited progress developing effective disease-modifying therapeutics has been due in part to an incomplete understanding of the sequence and coordination of molecular programs that drive disease pathogenesis. The advent of single cell genomics has revolutionized our understanding of the molecular and cellular mechanisms underlying PF. We and others have recently identified multiple previously unrecognized cell types/states that appear almost exclusively in individuals with PF and demonstrated cell-type specific dysregulation of gene expression in PF-lungs. While impactful, these studies have been limited to individuals with end stage disease, as samples were obtained from explant lungs removed during a transplant, and a comprehensive understanding of pathogenic molecular mechanisms across the natural history of PF is lacking. To investigate this, we have leveraged a unique cohort comprised of unaffected family members of familial pulmonary fibrosis patients. In this high-risk cohort, individuals underwent chest CT and bronchoscopy with transbronchial lung biopsy at the time of study enrollment, and have been followed for 5-10 years with serial imaging. We observe four trajectories within this cohort 1) no lung abnormalities, 2) newly arising lung abnormalities 3) stable lung abnormalities, 4) lung abnormalities progressing to symptomatic PF. To assess the molecular and cellular changes associated with early lung abnormalities and assess differences between these trajectories we performed single nuclei multiomics (RNA and ATAC) as well as spatial transcriptomics using the Vizgen MERSCOPE on tissues from three individuals within each category. We have identified patterns of dysregulation in both gene regulation and cellular composition that are distinct from the observed changes in late stage disease. Using the spatial transcriptomics data we have assessed these patterns in-situ and identified patterns of expression correlated with structural changes within the lung. Together, these data provide unique insights into the molecular mechanisms that drive the early pathogenesis of pulmonary fibrosis.

Omics Technologies Posters - Wednesday

PB3180. Utilization of Agena massARRAY to improve quality assessment of saliva samples for clinical genome sequencing

Authors:

C. Ho¹, L. Liao¹, P. Tong¹, N. Hammond¹, W. Qiao²; ¹Stanford Hlth.Care, Palo Alto, CA, ²Stanford Univ., Palo Alto, CA

Abstract Body:

Whole-genome sequencing (WGS) is becoming an increasingly powerful diagnostic tool for identifying genomic aberrations in patients. Given the inherent high cost of genome sequencing, a stringent quality assessment of each sample is typically performed in clinical laboratories to reduce clinical testing failures due to mislabeling and/or unintended specimen swaps. Moreover, the increasing use of saliva specimens for germline genetic testing necessitates additional specimen quality monitoring, given its variable bacterial content. Although DNA extracted from saliva is generally considered to be of comparable quality to that derived from blood, recent studies have shown that non-human contaminating DNA in human saliva can interfere with whole genome sequencing results. As such, it is necessary and important to evaluate the quality of human DNA extracted from each saliva sample and adjust the library input amount accordingly for downstream clinical genome sequencing. Therefore, our laboratory developed and validated a DNA fingerprinting assay using the commercially available MassARRAY genotyping kit, which interrogates 44 single nucleotide polymorphisms, three gender markers, and five sample integrity quality control (QC) markers. The clinical sensitivity, specificity, and accuracy of the fingerprinting assay were all >99.9%, with 95% confidence interval for clinical accuracy being 94.2-100%, supporting its use in clinical testing. In addition, potential bacterial contamination levels among patient saliva samples were estimated based on the ratio of amplifiable copies for human DNA to total DNA. 42 clinical and reference material samples were included in the DNA fingerprinting assay validation and an additional 24 saliva samples were used specifically to determine the percentage of human DNA in each sample, which subsequently were compared to the corresponding clinical genome sequencing ($\geq 40X$) mapped reads percentage to define the appropriate QC metric. Importantly, a linear relationship was observed between amplifiable copy values and mapped read percentage, which was leveraged to inform saliva WGS library loading. An improved weighting strategy based on the required mapped reads percentage that would pass genome sequencing QC metrics was subsequently established and implemented for clinical WGS library loading of saliva samples, and >95% sequencing QC passing rate to obtain 40X genome coverage was achieved. These results support the use of the massARRAY fingerprinting assay, coupled with our novel DNA quality assessment/management strategy, to effectively reduce clinical genome sequencing testing failures for saliva samples.

Omics Technologies Posters - Thursday

PB3181. Utilization of multi-omic, multiplexed cell atlases and benchmark analysis demonstrates purity of product in the manufacture of induced pluripotent stem cell derived natural killer cells

Authors:

M. Denholtz, A. Fernandez-Perez, K. Palomares, S. Kothapally Hanok, C. Chen, J. Yao, J. Hernandez, Z. Fazal, Y. Sui, B. Bober, B. Rezner, R. Bjordahl, T. Lee, D. Robbins, B. Valamehr; Fate Therapeutics Inc, San Diego, CA

Abstract Body:

Despite the promise of autologous cellular immunotherapies, the complexities of manufacturing, including the required single-patient manufacturing runs, extended patient treatment time, heterogeneity related to cellular engineering, and the overall high-cost of manufacturing, present substantial challenges to their broad clinical adaptation. To circumvent these challenges, our novel platform utilizes genetically engineered, clonal master human induced pluripotent stem cells (iPSCs) lines to derive allogeneic T (iT) and natural killer (iNK) cell based cellular immunotherapies. These master iPSC lines undergo directed differentiation to generate a nearly unlimited supply of a specialized, homogeneously engineered product, creating a true “off-the-shelf” cellular immunotherapy. The manufacture of iPSCs-derived cellular immunotherapy products includes a proprietary and complex multistep differentiation process to generate iNK cells. This process is designed to mimic *in vivo* hematopoietic development and yields exceptionally pure, homogenous cell products. To fully characterize this unique manufacture process which includes the generation of various cell types during the intermediate stages of the differentiation process, we developed a combined experimental and analytical pipeline to characterize the intermediate cell types in our iNK cell differentiation process, and to nominate validated flow cytometry markers for analytical development. Utilizing public multi-omic single cell atlas data sets we used benchmark analysis utilizing knowledge transfer algorithms to identify, characterize, and quantify distinct cell types present at the various stages of our differentiation process. Key intermediate cell types were further characterized via generation of internal, multiplexed, multi-modal cell atlases, thus allowing for the direct comparison of the transcriptional and cell surface protein repertoires of the intermediate cell types and final product. Utilizing this in-depth characterization pipeline, we confirmed the high level of homogeneity of the assessed iNK cell products (greater than 95 percent), defined key intermediate hematopoietic cell types in our differentiation process, and identified shared and engineering-specific differentiation pathways.

Omics Technologies Posters - Wednesday

PB3182. Utilizing paraformaldehyde fixation to expand opportunities for single-cell RNA-sequencing.

Authors:

B. Meta¹, D. Corney¹, Y. Han¹, Y. Qiu¹, J. Lozach², P. Vishwanath¹, M. Stephens¹, H. Latif¹, G. Zhou¹; ¹Azenta, South Plainfield, NJ, ²Rosalind Bio, San Diego, CA

Abstract Body:

Traditional approaches to performing transcriptional profiling have relied heavily on bulk RNA-seq which, while straight-forward and cost-effective, eliminates all evidence of heterogeneity that may have been present within a complex and diverse tissue. For this reason, high-throughput droplet- and microfluidic-based approaches for performing single-cell RNA-seq (scRNA-seq) have become powerful techniques to answer important questions in a wide range of fields. However, a significant bottleneck to widespread and routine usage has been the specialized and involved sample preparation requirements prior to performing scRNA-seq. Typically, tissues must be dissociated and processed immediately, which makes it difficult to batch and transport samples off-site, or cryopreserved, which is associated with challenges related to poor cellular viability upon sample thawing.

To further broaden the ability for researchers to deploy scRNA-seq, we describe our experience deploying a pre-commercial workflow developed by 10x Genomics to resolve these sample management and logistics challenges. On-site paraformaldehyde (PFA) fixation at the point of specimen collection allows samples to be collected, fixed, and transported to off-site laboratories for downstream processing. To demonstrate the workflow, various sample types were collected and either immediately fixed or cryopreserved for paired analysis by either probe-based transcriptome-wide gene expression profiling or using current single-cell 3' end counting. In all cases, fixed and cryopreserved cells demonstrated comparable clustering, gene and UMI detection, and gene expression metrics. Furthermore, inference of PBMC cell type identity determined by gene expression signatures showed comparable frequencies across both cryopreserved and fixed samples. Finally, downstream data analysis and visualization was performed in the Rosalind analytics platform, thereby enabling straightforward and collaborative interpretation with the ability to incorporate data obtained from other modalities for multiomic analysis. Taken together, this new approach expands the ability to perform single-cell transcriptome-wide analysis of gene expression at high sensitivity on challenging sample types and sample collection locations.

Omics Technologies Posters - Thursday

PB3183. Utilizing REVEAL SingleCell for single cell spatial transcriptomics data storage, analysis, and visualization

Authors:

C. Bragdon, S. Sarangi, U. Mudgal, Z. Pitluk, K. Sharma; Paradigm4, Waltham, MA

Abstract Body:

Single cell 'omics methods such as RNA-seq, CITE-seq, and ATAC-seq are valuable techniques for generating molecular profiles of cell type distributions and understanding regulatory networks in tissue samples. However, due to the required dissociation of cells in these techniques, there is loss of spatial information that might otherwise inform important questions about tissue organization such as disease microenvironment in Alzheimer's disease, inflammation centers, and tumor microenvironment in cancers. Spatial transcriptomics (ST) has become the key method to measure gene activity in a tissue sample and preserve the spatial information. While ST currently needs to improve read depth, transcriptome coverage, and there is a tradeoff between throughput and cell level resolution, the technique is a powerful tool that improves mechanistic and functional understanding of disease processes to build and test novel hypotheses. The growth in ST datasets and technologies, and integration with other single cell 'omics is creating the need for efficient storage and querying of multi-omics single cell and imaging data easily across multiple samples. In this poster, we present features of the REVEAL SingleCell analytical platform that support working with ST data and allow development and testing of novel use cases. We demonstrate loading of Visium ST data through the R-API using 10x, Seurat, or h5 objects signifying interoperability and ease of use. Rapid data retrieval of expression values as well as image metadata from ST datasets is also achieved through the R-API allowing efficient integration with downstream analysis workflows. Next, we present timings for scalable analytics such as computation of z-scores and identification of top expressed genes across Visium ST 'spots' in a sample. We also demonstrate creating and analyzing integrated ST datasets (iSets) from multiple samples, simulating creating bespoke ST atlases. Finally, we present a potential use case with REVEAL SingleCell that allows a user to select images associated with ST samples followed by selection of regions and querying gene expression.

Omics Technologies Posters - Wednesday

PB3184. Visualize RNA biomarkers to spatially interrogate complex tissues using the RNAscope™ HiPlex v2 in situ hybridization assay

Authors:

A. Dikshit, S. Basak, C-W. Chang, K. Collins; Advanced Cell Diagnostics, Newark, CA

Abstract Body:

Precise characterization of cell types in complex tissues is essential to determine their functional significance. Targeted gene expression analysis enables profiling of diverse cell types found in heterogeneous tissues such as the tumor and brain. The tumor microenvironment (TME) is comprised of tumor cells, immune cells, stromal cells, and extracellular matrix. Interrogating spatial interactions and activation states of immune cells in the TME is crucial for implementing successful therapies. Similarly, establishing gene expression profiles is critical for better understanding the pathology underlying many CNS disorders. This study demonstrates the use of highly specific RNAscope HiPlex v2 in situ hybridization (ISH) assay for detecting target gene expression in fresh frozen brain and FFPE tumor samples. The RNAscope HiPlex v2 assay has the capability of iteratively multiplexing up to 48 targets in fixed and fresh frozen samples and up to 12 targets in formalin fixed paraffin embedded (FFPE) tissues. The novel FFPE reagent effectively reduces background autofluorescence, improving the signal to noise ratio. We have leveraged this technology to investigate spatial expression of 12 oncology and immunology target gene as well as 12 targets implicated in the pain modulation in the brain. In the tumor tissues, we visualized T cell infiltration and identified T cell subsets within tumors using *CD3*, *CD8*, *PD1* and *IFNG* markers. Cytotoxic T cell phenotype and activated T cell phenotype were quantified using HALO image analysis. We further identified subsets of pro- and anti-inflammatory macrophages by *CD68* and *CD163* expression and detected chemokines and cytokines such as *CXCL10* and *CCL22*. Quantitative analysis indicated higher degree of macrophage and tumor associated macrophage (TAM) infiltration in the lung tumor compared to the cervical tumor. In the mouse brain, differential expression of opioid receptors, *Oprm1*, *Oprd1* and *Oprk1* was observed in different regions of the brain, implying region-specific function. Neuropeptide Y known to be involved in pain perception and transmission was observed to localize in the astrocytes and microglia of the nucleus accumbens. Using a highly sensitive multiplexed RNAscope HiPlex v2 ISH assay, we have demonstrated a technique to spatially resolve 12 targets in FFPE human tumor and frozen mouse brain tissues. The FFPE reagent efficiently quenched background autofluorescence in the FFPE tissues. This technology is highly valuable for investigating spatial interactions and can provide insights into the biological crosstalk among various cell types in complex and heterogeneous tissues.

Omics Technologies Posters - Thursday

PB3185. WES, with its limitations, continues to provide an attractive mode for genomic data collection

Authors:

S. Rockowitz, C. French, W. Shao, A. Sharma, P. Sliz; Boston Children's Hosp., Boston, MA

Abstract Body:

The transition from whole exome sequencing (WES) to whole genome sequencing (WGS) has opened opportunities for new types of analysis and may increase diagnostic yields, yet, the deployment of WGS over WES presents several challenges including: regime of sample collection, cost and data management. At BCH, we developed a hybrid WES/WGS sample collection system where we have collected WES and/or WGS samples from 6,188 individuals from 2,645 families (1,990 with WES only, 203 with WGS only, 452 with WES and WGS). As we collected both sets of data, we systematically evaluated the operationalization from sample collection to diagnosis and re-analyses. Lack of a streamlined process for the collection of remote CLIA WGS blood samples led cohorts to prioritize the collection of buccal swabs for WES. Preliminary analysis of these data indicates that when WGS was collected, diagnosis rates were increased over WES across numerous cohorts. Further, automated re-analyses workflows held additional diagnostic promise for initially undiagnosed cases.

Omics Technologies Posters - Wednesday

PB3186. Whole genome sequencing of low input tagmentation-based libraries results in high quality somatic variant calling comparable to ligation-based PCR-free libraries.

Authors:

E. Rice¹, L. Jiang¹, X. Zhang¹, C. Alba¹, G. Sukumar¹, M. D. Wilkerson², C. L. Dalgard²; ¹Henry M. Jackson Fndn., Bethesda, MD, ²Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD

Abstract Body:

Whole genome sequencing (WGS) of tumor tissue may indicate novel treatment paradigms through molecular classification of human cancers. While surgically obtained tumor samples may yield DNA below the microgram input required by established ligation-based PCR-free library preparation workflows, the lower, 100 to 300 nanogram inputs of novel PCR-free tagmentation-based methodologies may broaden the WGS applications of such samples. Comparative analysis of tagmentation- to ligation-based workflows along with benchmarking to a high-confidence reference is prudent to ensure maintenance of high-quality calls for future translational and clinical applications. To address this validation, replicates of tagmentation- and ligation-based sequencing libraries were generated for the HCC1395 breast cancer-derived tumor cell line and its matched germline normal lymphoblastoid cell line previously characterized by the SEQC2 Somatic Mutation Working Group. WGS was performed on the Illumina NovaSeq 6000 to a targeted sequencing depth of 30x for germline and 90x for tumor, and variant calling was performed. Sequencing quality control metrics, germline variant calls, and somatic mutation calls were compared to assess technical performance, concordance, precision, and sensitivity across methods. Tagmentation- and ligation-based libraries both achieved above threshold Q30 percent, above target mean coverage, greater than 90% aligned reads, and highly similar average insert size. Germline variant calls analyzed pairwise displayed 99.98% concordance between workflows. Germline variants showed greater than 98% precision and sensitivity using ligation-based variants as a reference. For somatic mutation calling using method-specific pairs, 88% precision and 81% sensitivity in exonic regions was observed against a ligation-based baseline. Against the SEQC2 reference call set, ligation- and tagmentation-based method SNP calls resulted in 93% and 89% precision and 73% and 74% sensitivity, respectively. A high fraction of false negative calls from the SEQC2 high/medium confidence space was associated with low tumor variant allele frequency (TVAF) of <0.1. Applying a cutoff of >0.1 TVAF for SEQC2 calls only increased ligation- and tagmentation-based method sensitivity to 82% and 81%, respectively, underscoring a highly similar outcome for both workflows. In summary, high concordance with ligation-based methodology, and similar precision and sensitivity compared to reference data provides confidence for tagmentation as a suitable method for attaining high quality somatic mutation calls from minimal starting material.

Omics Technologies Posters - Thursday

PB3187. Why long-read sequencing is poised to become key to clinical genetics: an episodic ataxia study.

Authors:

S. Audet^{1,2}, V. Triassi^{1,3}, C. Michaud^{1,2}, E. Bareke¹, V. Ferraro^{4,2}, L. Touma^{4,2}, A. Duquette^{4,2}, M. Tétreault^{1,2}; ¹CRCHUM, Montreal, QC, Canada, ²Dept. of NeuroSci.s, Univ. of Montreal, Montreal, QC, Canada, ³Dept. of Bioinformatics, Univ. of Montreal, Montreal, QC, Canada, ⁴Dept. of Neurology, CHUM, Montreal, QC, Canada

Abstract Body:

Despite a rapid evolution of technologies available in clinical settings, such as next generation sequencing (NGS) which has had a major impact on the identification of pathogenic variants, many patients remain without a molecular diagnosis. For rare neurological disorders, standard diagnostic approaches generally achieve a success rate of 20-50%. One of the major obstacles faced with NGS is the complexity of interpreting variants of unknown significance (VUS). Hence, the many applications of long-read sequencing (LRS) are highly compelling to efficiently bridge the identification of variants and the interpretation of their impact. Indeed, allowing the capture of full-length gene transcripts in a quantitative manner generates a great amount of functional information which could facilitate VUS interpretation in several manners.

Our lab conducted a pilot study for which the aim was divided in two parts: 1) evaluating the performance of a multi-omics approach on the diagnostic yield of complex ataxia cases and 2) defining how powerful LRS is for the validation of candidate VUS. The hypothesis was that diagnosis would be reached for more than half of the cohort, and that LRS would be a major contributor to the validation steps of our investigation.

A cohort of eight patients with complex episodic ataxias were recruited following a thorough yet unsuccessful clinical investigation to define the cause of their disease. DNA, RNA and proteins were extracted from peripheral blood mononuclear cells, enabling the combination of whole-genome and RNA-sequencing, as well as validation experiments. LRS was performed using Nanopore technologies. The multi-omics approach has allowed the identification of excellent candidates in each patient, four of which have been functionally validated through LRS (4/8; 50%). The interesting part of the result is the diverging contexts in which long-reads were used to evaluate the impact of the variants: to demonstrate the trans configuration of two *SPG7* variants, to prove the alternative splicing and transcriptome shift of *ELOVL4* and *PMPCB* in two separate patients, and to precisely quantify the number of *ATXN2* repeats in another case. The tool's versatility was unsurprisingly a major factor in its capacity to contribute to the functional validation.

These results highlight the potential of multi-omics and LRS in the context of clinical genetics.

Additionally, they should have a direct impact on patients through the psychological relief linked to an official molecular diagnosis, a better understanding of the pathology, and better care management.

Finally, these findings may hopefully lead to novel therapeutic targets in the future.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3188. A comprehensive analysis of the reusability of public omics data across 2.8 million research publications

Authors:

M. Vahed¹, N. Darci-Maher², K. Peng¹, J. Brito¹, A. Rajesh¹, A. Smith¹, R. Thompson³, A. Nellore⁴, J. Jacobs⁵, D. Duong⁶, E. Eskin⁷, S. Mangul⁸; ¹Univ. of Southern California, Los Angeles, CA, ²Univ. of California, Los Angeles (UCLA), Oakland, CA, ³Oregon Hlth.& Sci. Univ., Portland, OR, ⁴Univ. of Pennsylvania, Philadelphia, PA, ⁵QIAGEN Digital Insights, Redwood City, CA, ⁶Univ. of California Los Angeles, Culver City, CA, ⁷Univ California los Angeles, Los Angeles, CA, ⁸Univ. of Southern California, Sch. of Pharmacy, Los Angeles, CA

Abstract Body:

There is growing evidence that data sharing enables important discoveries across various biomedical disciplines. When data is shared on centralized repositories in easy-to-use formats, other researchers can examine and re-analyze the data, challenge existing interpretations, and test new theories. Additionally, secondary analysis is economically sustainable and can be used in countries with limited resources. Advancements in high-throughput omics technologies have reshaped modern biomedical research. However, once a research team publishes critical findings derived from an omics dataset, secondary analysis can play a crucial role in enabling and verifying the reproducibility and generalizability of published results. On the other hand, there is a lack of knowledge of how widely and what types of genomic data are reused and how it contributes to novel biological discoveries. We have performed a data-driven examination of reuse patterns of the reusability of public omics data across 2,882,007 open-access biomedical publications (published between 2001 and 2020; across 13,753 journals). Our search included two omics data repositories, NCBI Sequence Read Archive (SRA) and NCBI Gene Expression Omnibus (GEO). We searched for SRA or GEO accession IDs inside the publications using regular expressions. We tested the accuracy of this assumption using a subset of datasets for which investigators manually linked their dataset records with PubMed identifiers and found it to be accurate. Using the proposed protocol, we were able to identify 88,429 publications with SRA or GEO identifiers and compare the number of studies performing primary analysis versus the number of studies performing secondary analysis. We found out publications mentioning GEO have a higher proportion of secondary analysis than those saying SRA, and the balance of publications mentioning GEO with performing secondary analysis has increased over time, from roughly 20% in 2004 to approximately 60% in 2019. Of the publications citing SRA, this rate has remained around 30% from 2009 to 2019. Considering the data in units of datasets and the number of times they are reused, we found that except for a few initiatives, omics data is poorly reused, and over 59% of the data in GEO, and over 70% of the data in SRA, is not reused even once. Datasets in GEO have been reused 4.27 times, while datasets in SRA have been reused 1.89 times. Our study establishes the current state and trends of secondary analysis of omics data and suggests that an easy-to-use format is needed to enable omics data reusability.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3189. A comprehensive database of Long-chain fatty acid oxidation disorder gene variants.

Authors:

H. Richbourg¹, V. Rangel Miller¹, S. Daugherty¹, O. K. Japalaghi¹, D. Marsden¹, H. McLaughlin², A. Willcock³, N. Miller⁴; ¹Ultragenyx Pharmaceutical, Novato, CA, ²Invitae, South Lyon, MI, ³Invitae Corp., San Francisco, CA, ⁴Ultragenyx, Brisbane, CA

Abstract Body:

Long-chain fatty acid oxidation disorders (LC-FAOD) are rare autosomal recessive conditions caused by defects in genes encoding mitochondrial enzymes that convert long-chain fatty acids into energy. Patients with LC-FAOD may have characteristic elevations in plasma acylcarnitine profiles, identified through newborn screening follow-up and by testing patients with ongoing clinical signs and symptoms. We developed a comprehensive database of variants in the 6 genes associated with LC-FAOD (*ACADVL* [VCLAD], *CPT1A* [CPT1], *CPT2* [CPT2], *HADHA* [LCHAD or TFP], *HADHB* [TFP], and *SLC25A20* [CACT]) by integrating data from three sources: a systematic review of all published medical literature, a sponsored LC-FAOD gene panel testing program, and Ultragenyx clinical programs. Variants were annotated with molecular diagnoses, newborn screening test results, detailed clinical and biochemical phenotypes, and ACMG variant interpretations. As of 5 May 2022, the LC-FAOD gene variant database reports 3830 variants from 2970 patients (809 female, 930 male, 1231 unknown) with one or more LC-FAOD gene variants. These represent 957 unique variants: 682 (71%) pathogenic (P) or likely pathogenic (LP), 246 (26%) variants of uncertain significance (VUS), 1 (0.1%) benign, 7 (1%) likely benign, and 17 (2%) with conflicting information. Single nucleotide variants (658, 69%) are the most common type followed by small deletions (132, 14%), small insertions (40, 4%), deletions (>100bp) (9, 1%), indels (12, 1%), and insertions (1, 0.1%); the remaining 105 variants (11%) document a protein-level change with no cDNA information available. 2185 patients had a positive LC-FAOD molecular diagnosis (at least 2 P/LP variants) distributed as follows: *ACADVL*, 901/42%; *CPT2*, 501/23%; *HADHA*, 461/21%; *HADHB*, 202/9%; *SLC25A20*, 65/3%; *CPT1A*, 55/3%. 18 additional patients are reported as LC-FAOD double heterozygotes (P/LP or VUS) with 18 unique gene variant combinations, most commonly including *ACADVL* (15) and/or *HADHA* (8). 119 patients with a *CPT2* diagnosis also had ≥ 1 of 3 known thermolabile variants in this gene: c.1055T>G, c.1102G>A, c.1939A>G. This comprehensive database of LC-FAOD gene variants will be available on a searchable, interactive website and will continue to compile data on variants identified in ongoing clinical studies, genetic testing programs, and literature searches. Centralized resources like the LC-FAOD genes database are critically important in rare diseases where clinical information on variants is scarce and VUS are frequent and insufficient to support a molecular confirmation of diagnosis.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3190. A comprehensive Japanese genetic variation database TogoVar.

Authors:

T. Katayama¹, N. Mitsuhashi¹, M. Kawashima², L. Toyo-oka³, Y. Moriya¹, S. Kawashima⁴, T. Takagi³; ¹Database Ctr. for Life Sci., Kashiwa, Chiba, Japan, ²Japan Sci. and Technology Agency, Chiyoda-ku, Japan, ³Toyama Univ. of Intl. Studies, Toyama, Japan, ⁴Database Ctr. for Life Sci., Chiba, Japan

Abstract Body:

TogoVar (<https://togovar.biosciencedbc.jp>) is a database that offers integrated Japanese allele frequency information derived from the Japanese Genotype-phenotype Archive, the GEM Japan Whole Genome Aggregation Panel, the ToMMo 8.3KJPN Allele Frequency Panel, and the Human Genetic Variation Database. As the sample size of the Japanese population in the international variant database is still limited and Japanese genome information is scattered in distributed databases with controlled access, it is essential for Japanese human genetics research to provide a public dataset that accumulates Japanese allele frequencies from these data sources. This year we have constantly enhanced the amount of data in collaboration with the cohort and sequencing projects in Japan that resulted in over 100 million variants. To aid the interpretation of these variants, TogoVar also integrates additional information: (1) allele frequencies from the gnomAD database so that researchers can make a comparison of allele frequencies with a specific ethnicity, (2) annotations of variants including molecular consequence and variant effects calculated by the Ensembl Variant Effect Predictor, (3) pathogenicity information provided by the ClinVar database, and (4) relevant literature information from LitVar and PubTator tools. For fully utilizing the integrated information, we have implemented the advanced query interface that enables researchers to combine conditions of variant types, consequences, clinical significances, and allele frequencies in each dataset with the logical AND/OR operators. The query results will be navigated to a variant report page as well as newly introduced gene and disease report pages that provide a summary of variants relevant to a specific gene and a disease. Lastly, from the perspective of the system architecture, TogoVar provides two unique features: (1) tables and charts in TogoVar are provided as a set of reusable modular elements that can be embedded in any web application as a WebComponents module, and (2) the integrated data is internally standardized as a knowledge graph so that developers can make advanced graph queries for analysis. Here we report new advancements and the future directions of the TogoVar database.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3191. A globally representative panel of human genomes: Resource and tutorials to support analyses of diverse populations

Authors:

M. Yohannes^{1,2}, Z. Koenig^{1,2}, J. Goodrich^{1,2}, M. Wilson^{1,2}, G. Tiao¹, A. Kim^{1,2}, L. Nkambule^{1,2}, X. Zhao^{2,1,3}, S. Hao^{2,1}, N. Sahakian¹, M. Talkowski^{2,1,3}, H. Brand^{2,1}, K. Karczewski^{1,2}, M. Daly^{2,1,4}, E. Atkinson^{5,1,2}, A. Martin^{2,1}; ¹The Broad Inst. of MIT and Harvard, Cambridge, MA, ²Massachusetts Gen. Hosp., Boston, MA, ³Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, ⁴Inst. for Molecular Med. Finland, Helsinki, Finland, ⁵Baylor Coll. of Med., Houston, TX

Abstract Body:

Due to a lack of resources supporting their analysis, diverse populations are often excluded from genomic studies. The 1000 Genomes Project (1kGP) and Human Genome Diversity Project (HGDP) are among the most valuable genomic resources because of the breadth of global diversity they capture as well as their open sharing policies that allow release of unrestricted individual-level data. Recently, these two resources were sequenced to high coverage, but have yet to be well harmonized. To alleviate the disparity in resources, we are releasing a joint-called reference panel of more than 155 million high-quality variants obtained from harmonizing a deeply sequenced set of over 4,000 whole genomes from 1kGP and HGDP. Additionally, we are providing detailed tutorials for conducting many of the most common quality control steps and analyses with these data in a scalable compute setting.

These tutorials outline steps for conducting quality control, ancestry analyses, and joint calling with additional datasets. They were developed from best practice analysis scripts so researchers of any level can run through each step to develop an understanding of the data and how to work with it using state-of-the-art processing pipelines. They are currently available on Github

(https://github.com/atgu/hgdp_tgp/tree/master/tutorials) and utilize the Hail Python library, an open source scalable tool for genomics that intuitively interfaces cloud computing and genetic analyses.

In sum, the release of these tutorials along with the reference panel aims to aid researchers in learning cloud based tools in addition to introducing methods for working with diverse ancestry datasets. Public release of this resource will foster representative genomic studies to fully capture the breadth of global genetic variation.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3192*. A novel reference architecture for multi-party federation: enabling joint analysis of large-scale clinical-genomic data across distributed Trusted Research Environments.

Authors:

C. Davison¹, I. Knarston¹, I. Peral¹, B. Goncalves¹, S. Keshari Sahu¹, R. Zochling¹, E. Kyriakou¹, T. Seeger¹, N. Raine¹, R. Fennessy², C. Bacchelli¹, P. Moss³, M. Chatzou Dunford¹, P. Prieto Barja¹, M. Avery², S. Nik-Zainal²; ¹Lifebit Biotech Ltd, London, United Kingdom, ²Univ. of Cambridge, Cambridge, United Kingdom, ³Genomics England, London, United Kingdom

Abstract Body:

Enabling researchers access to the vast amounts of linked clinical and genomic data available is crucial to answering the world's most pertinent research questions. Current real-world genomic and clinical data is predominantly siloed to protect data privacy due to strict regulatory frameworks. Siloed data sets are notoriously difficult for researchers to access and analyse. Establishing a secure computing environment that holds data, often called Trusted Research Environments (TREs), enables researchers to access that data for analysis while securing sensitive, identifiable patient data. However, joint analysis of multiple TREs is impossible within this current model, and data cannot be moved to a centralised repository without compromising data security. Multi-party federation solves this problem, allowing researchers to access and analyse data from multiple TREs enabling more data to be analysed simultaneously, leading to an increased potential for novel discoveries with some estimates that 10x the data leads to 100x the findings. The DARE UK-backed multi-party trusted research environment federation consortium, which includes the University of Cambridge, NIHR Cambridge Biomedical Research Centre, Genomics England, Eastern AHSN, Cambridge University Health Partners and Lifebit, is developing a novel reference architecture to define how federated analysis can be performed on large-scale clinical-genomic data across distributed TREs. This novel demonstration of a multi-party federation will securely bridge the TREs of a UK leading research institute, NIHR Cambridge BRC and a public-sector clinical research endeavour, Genomics England. Lifebit has deployed its TRE platform, CloudOS, across these large scale data resources and is developing novel Application Programming Interface technology to enable TRE communication, in addition to a scalable airlock system for secure export into and out of the TRE. All novel technology for this project will be fully open-source and developed in alignment with Global Alliance for Genomics and Health (GA4GH) standards. Lifebit's cloud-based platform will allow in situ analysis to be run on each TRE individually and then aggregate the results from both TREs to be generated in a 'safe-haven'. This groundbreaking study and use case will contribute novel insights for future multi-party federated analysis - demonstrating best practice reference architecture, the necessary open-source technology and data governance recommendations standard for how TREs can communicate to allow joint analysis.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3193. A Private, universal registry for patient engagement in biomedical research - Xia-Gibbs Syndrome as a model.

Authors:

A. Hansen¹, E. Little¹, D. Belliston¹, M. Murugan², J. Hu², C. Hansen¹, R. Gibbs²; ¹Geneial, Houston, TX, ²Baylor Coll. Med., Houston, TX

Abstract Body:

Many isolated collections of data from affected individuals have been established as disease-agnostic ‘biobanks’ or indication-specific ‘registries’. However, for orphan genetic disorders, despite best intentions to aggregate data, individual biobanks and registries often fail to achieve a critical mass of data necessary to attract commercial interest, leaving patient communities with a lack of resources to accelerate research and development. A critical mass could be achieved by pooling data from disparate biobanks, or by augmenting registries with additional samples found in biobanks, but competitive interests across data managers and the sensitivity of biomedical data have combined to create an environment where **health-privacy and data ownership concerns universally limit access to—and value of—these data resources.**

To address these concerns, we are developing a platform for data interoperability, pooling, and exchange that safeguards patient privacy and data ownership. **Using the Xia-Gibbs Syndrome Registry as a model, we have demonstrated proof-of-concept for a privacy-preserving disease registry search layer** that operates directly on encrypted data without the need for decryption at the point of analysis, enabling data aggregation and analysis to be securely outsourced to a third party unable to view raw data or analysis results. We tested these features against over 200 registry fields for over 150 participants and reported the baseline, unoptimized performance of each operation, measuring performance across a range of registry sizes to understand computational scalability, demonstrating a worst-case scenario of a linear relationship between registry size and computational cost.

In parallel, we have developed a HIPAA-compliant variant annotation service and ‘universal’ registry platform with a modular, hierarchical framework for structured phenotypic data management that supports both registry-specific field customization and cross-registry analysis. With further development these tools will enable privacy- and ownership-preserving targeted sample matchmaking and cohort creation across a network of independent, standardized, interoperable data repositories. **Ultimately, this will increase privacy protections for research participants and data owners while simultaneously increasing data availability and effective sample size for researchers through a decentralized, interoperable platform.**

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3194. A reference panel composed of Alzheimer's Disease enriched single-nucleotide and structure variants for structure variants discovery on SNP array data.

Authors:

P-L. Cheng, H. Wang, W-P. Lee; Pathology and Lab. Med., Philadelphia, PA

Abstract Body:

Compared to whole-genome sequence (WGS), SNP array is cost-effective and indeed has more data available for Alzheimer's Disease (AD) research. However, due to technology limitations, SNP array is limited to capture complete genomic information and is unlikely to support structure variant (SV) discovery which has been proven as a contributor to the genetic basis of human disease. Imputation, which is the process of inferring unobserved genotypes in a sample, may mitigate the challenge, but current reference panels contain either none or a limited number of SVs.

As the quality of a reference panel may be affected by sample size, target variant density, and genetic relationship of samples, a pure SV only panel is infeasible. Thus, we took SNPs and SVs detected on 16,905 WGS data from the Alzheimer's Disease Sequencing Project (ADSP), and composed a reference panel by employing SHAPEIT4 and MINIMAC3, which could use to impute other datasets by MINIMAC4.

To explore the feasibility, we used two regions of known SNPs and SVs in the same linkage disequilibrium (LD) block and examine imputation results. The first region is in MAPT gene where rs8070723 and a 238bp deletion are in the same LD block, while the second one is located at NCK2 where rs143080277 and a 5510bp deletion are in the same LD block. We created a panel for each region for the deletion and SNVs in +/- 100kbp. Then, we assess the results by imputing an independent SNP array dataset with ~1600 samples. The result showed concordant genotypes of SNPs and the deletions. It suggests the feasibility of including SVs and SVs for a reference panel building. We expanded the experiment to build a reference panel genome widely. The project could further clarify the relationship between SVs and SNPs, enable SV discovery in numerous array data of AD, and facilitate finding novel genetic associations of AD.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3195. A systematic assessment of the completeness of TCR databases across *Mus musculus* strains.

Authors:

Y. He, Y-N. Huang, Y. Patel, C. Ronkowski, S. Mangul; Univ. of Southern California, Sch. of Pharmacy, Titus Family Dept. of Clinical Pharmacy, Los Angeles, CA

Abstract Body:

Immunogenetics databases facilitate the study of genetic polymorphisms in the immune system and the analysis, diagnosis, and treatment of diseases. It allows researchers to tap into the evolution of TCR diversity to study numerous diseases, such as cancers, autoimmune diseases, inflammation, and infectious diseases. In laboratory settings, lab mice (*Mus musculus*) account for the majority of non-human vertebrate data and are vital for in-vivo and in-silico studies. However, many studies neglect certain strains used to build mouse models. Disparities exist on how immunogenetics databases represent diverse mouse strains. In this study, we examined the completeness of the IMGT database, a comprehensive database for mouse TCR genes, representing diverse lab mice strains. To conduct the assessment, we analyzed the TCR-sequences of 100 mice samples from the Sequence Read Archive (SRA) using the BCR and TCR analysis toolkit, MiXCR. MiXCR aligns the sample TCR-Seq reads to the IMGT database across diverse mice strains. From the output of MiXCR, we will be able to analyze the mismatches in the VDJ genes and unmapped TCR reads of the diverse mice strains. The proposed methods allow us to evaluate the completeness of the IMGT database representing diverse mice strains by counting the number of mismatches and unmapped reads. Mismatches include substitutions, deletions, and insertions in the VDJ genes of the mice samples compared to the reference. Our evaluation of the completeness of immunogenetics databases representing diverse mice strain types would reveal the strains of lab mice which are underrepresented in the IMGT database. The results can highlight the disparities in the completeness of TCR databases across mice strains and guide further efforts aimed to complete such databases.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3196. ALS/FTD Compute: an open, centralized repository of genomic data for ALS/FTD research

Authors:

B. Traynor¹, C. Dalgard², P. Keagle³, J. Glass⁴, J. E. Landers³; ¹Natl. Inst Aging, Bethesda, MD, ²USUHS - TAGC, Bethesda, MD, ³Univ. of Massachusetts Med. Sch., Worcester, MA, ⁴Emory Univ., Atlanta, GA

Abstract Body:

BackgroundThe discovery of genes responsible for amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) has revolutionized our understanding of these related diseases and changed the clinical approach to patients diagnosed with this spectrum of fatal neurodegenerative conditions. It also has provided targets for drug development and targeted gene therapy efforts to slow progression. *ALS/FTD Compute* was established to maintain the momentum of gene discovery in these disorders by unifying existing whole-genome sequencing efforts of ALS and FTD patients into a central repository. **Methods**Using the AnVIL cloud environment, we established the *ALS/FTD Compute* as an open central repository. This platform consolidates data from multiple patient and control subject sequencing efforts on a single cloud provider. This effort will facilitate data harmonization improving the quality of the data available for analysis. The ultimate goal is to lower the costs associated with genomic analysis in the ALS/FTD space while democratizing access to this harmonized cohort. **Results**Major sequencing efforts are currently participating in *ALS/FTD Compute*. These include Answer ALS, Genomic Translation for ALS Care GTAC, New York Genome Center ALS Consortium, National Institutes of Health, and Project Mine USA, illustrating the collaborative spirit among the various groups and their recognition of the value of this efficient model of collaboration. To date, whole-genome sequence data for 10,000 samples have been uploaded to *ALS/FTD Compute*. An additional ~35,000 control samples will be included in our harmonization efforts. A cloud environment eliminates the prohibitive costs of large-scale data movement and provides the infrastructure to perform complex genomic analyses. Additionally, the AnVIL platform is built upon well-established components that facilitate various workflows, including Terra, Gen3, Jupyter, Galaxy, Bioconductor, and the Dockstore. AnVIL actively collaborates with other genomic data resources by adopting the FAIR (Findable, Accessible, Interoperable, Reusable) principles. **Conclusions***ALS/FTD Compute* sets the stage for the next iteration of gene discovery in ALS and FTD by adopting an efficient, open and inclusive platform. Our platform demonstrates the utility and efficiencies of the central repository approach in rare neurological diseases.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3197. An improved platform for sharing genomic information of underrepresented populations: BIPMed 2.0

Authors:

I. Lopes-Cendes^{1,2}, T. C. de Oliveira^{1,2}, W. de Souza^{1,2}, C. S. Rocha^{1,2}, B. S. Carvalho^{1,2}; ¹Univ. of Campinas (UNICAMP), Campinas, Brazil, ²Brazilian Inst. of NeuroSci. and Neurotechnology (BRAINN), Campinas, Brazil

Abstract Body:

Introduction: The Brazilian Initiative on Precision Medicine (BIPMed), established in 2015, provided the first platform for sharing genomic and health-related information of admixed populations in Latin America (www.bipmed.org). Over the past two years, we performed a major overhaul in the BIPMed genomic platform so that it can continue serving the medical and research community worldwide. **Materials and Methods:** We have performed several modifications to the BIPMed genomic platform, including implementing CRAM and VCF file formats. Also, we are making the BIPMed genomic data publically available in two reference genomes, hg19 and hg38. Furthermore, we have developed a new tool to import variants from VCF files to the BraVE application programming interface (API) based on the GA4GH Genomics API. **Results:** Currently, the platform has seven databases, including genomic and genetic information from about 900 individuals. Two of these databases are from the 'reference' Brazilian population, containing datasets of admixed individuals ascertained based on place of birth and not on disease phenotypes. Five are disease-specific databases, including diverse phenotypes ranging from cancer to neurological disorders. The two reference databases are composed of unique SNPs identified using whole-exome sequence and SNP arrays. Genetic variation within the databases can now be searched using three online tools (BEACON - GA4GH -, LovD - Leiden University - and the BraVe application, which has been developed in-house). Using this new implementation, we found that data processing in the web interface is considerably faster and fixes some critical issues regarding missing information in the variants file. **Conclusion:** Keeping public genomic databases running demands time and effort, requiring computational solutions for hosting, displaying, searching, and interacting with the datasets. Furthermore, the web interface should provide a good experience to suit users with different scientific backgrounds. By releasing BIPMed 2.0, we are keeping with our mission of supporting the implementation of precision medicine in Brazil. Funding: FAPESP, SP, Brazil.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3198*. Assessing the completeness of immunogenetics databases across diverse populations

Authors:

Y. Huang¹, Y. Meng¹, N. Patel², J. Mehta², B. Hua², M. Fayzullina², H. Alachkar¹, S. Mangul¹; ¹Dept. of Clinical Pharmacy, Sch. of Pharmacy, Univ. of Southern California, Los Angeles, CA, ²Dept. of Pharmaceutical Sci., Sch. of Pharmacy, Univ. of Southern California, Los Angeles, CA

Abstract Body:

Although the advanced development of bioinformatics tools has driven the popularity of AIRR sequencing (AIRR-Seq) studies and has promoted the scientific community to profile human AIRR efficiently, it has been previously disclosed that most of the study participants in the AIRR-Seq studies were of European ancestry, which lacks samples representing the underrepresented populations. Expanding the knowledge of the underrepresented populations in the field of AIRR studies will enhance the understanding of the phenotypic differences in immune cell receptor repertoires and elucidate the different responses to immune-related diseases across diverse populations. In the study, we examined the completeness of the international ImMunoGeneTics information system® database (IMGT database) for representing diverse populations. By leveraging the bioinformatics software, MiXCR, we will be able to comprehensively examine the mismatches in different ancestry group samples' read in the VDJ gene and evaluate the completeness of the IMGT database across diverse ancestry groups. MiXCR aligns and compares the TCR-Seq reads to the IMGT database. In our preliminary results, we analyzed ten European samples and ten Asian samples from Sequence Read Archive (SRA) and counted the number of substitutions, insertions, and deletions in the V and J genes of the T cell receptor sequences (TCR-Seq) from the output of MiXCR and counted the number of the unmapped reads of the samples. We compared the number of mismatches of the samples across diverse ancestry groups and discovered that the samples of European ancestry had fewer mismatches (Mean number of mismatches is 0.77) in the V genes than samples of Asian ancestry (Mean number of mismatches is 4.59) and a similar trend of mismatches are observed in the J genes among the European and Asian group. The result indicates that the IMGT database is more completely representing European ancestry compared to Asian ancestry. We are currently expanding the analysis to the publicly available datasets on the SRA and aim to run MiXCR across all TCR-Seq samples with available ancestry labels from SRA. Unveiling the completeness of the IMGT database representing diverse populations could highlight the need to improve ancestry diversity in those underrepresented populations and guide future immunogenomics studies to improve ancestry availability and distribution.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3200. Blockchain and Artificial Intelligence-Enabled Stratified Trial System (BESTS) - A patient driven data sharing platform that leverages genomic and health data to accelerate clinical trial recruitment for precision therapies

Authors:

L. Brady¹, G. Coyne², K. Power¹, Z. McCrea¹, O. Teltsh¹, D. Murphy², W. Northey², M. White¹, E. Gilbert¹, M. Grealley¹, R. Ileras³, C. Doherty^{4,5}, N. Delanty^{6,1}, M. Meagher², K. Marshall⁷, O. Rigby², G. L. Cavalleri¹; ¹Royal Coll. of Surgeons in Ireland, Dublin, Ireland, ²Akkure, Dublin, Ireland, ³Microsoft, New York, NY, ⁴Trinity Coll., Dublin, Ireland, ⁵St. James's Hosp., Dublin, Ireland, ⁶Beaumont Hosp., Dublin, Ireland, ⁷Microsoft, Dublin, Ireland

Abstract Body:

Background: Precision medicine requires the integration of clinical and genetic data to better understand disease, tailor treatment strategies and deliver improved patient outcomes. To facilitate the development of increasing targeted therapies, clinical trials will also need to become more personalized and stratified. However, recruitment continues to be a significant barrier for trial sponsors, clinical research teams and patients. This challenge can be met by connecting and stratifying patients to trials via their clinical and genomic data. Here we present BESTS- a cloud based platform developed for collaborative use by patients, healthcare providers (HCPs) and clinical research organisations. It allows patients to be intelligently matched to, and make their health data available for use in genomically-stratified clinical trials, while retaining complete control and ownership of that data. *Methods:* Twenty five participants representing potential users of BESTS engaged in a series of qualitative research methods including delphi interviews and ethnographic fieldwork to inform platform design and build. Applying a data privacy impact assessment approach, a dynamic consent module with an appropriate governance and rule-based permissions structure enabled by blockchain technology, was designed. A genomics module was developed to allow patients upload pre-existing genetic data or generate new data by accessing sequencing through the platform. A genomics working pipeline that facilitates automatic reanalysis was built based on GATK4 algorithms and deployed on Cromwell with Jupyter Notebook. An R script was added to create a shortlist of variants for ACMG classification. A proof of concept of the full pipeline was performed using whole genome and whole exome data. *Results:* Value propositions for patients included greater control around use of their data, personalised trial matching, contributing to research that benefits wider population and access to genetic sequencing. Value propositions for HCPs included identification of patients for trials, potential to demonstrate suitability as trial site, integration with electronic health records and access to characterised patient population. An early prototype of platform was developed and genomics pipeline deployed demonstrating successful delivery of appropriate output files. *Conclusion:* Embedding user centred design in the development of BESTS enables its realworld effectiveness and application. Access to genomic and phenome data through BESTS provides for high resolution recruitment to clinical trials, facilitating faster introduction of treatments into patient care pathways.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3201. Breast cancer related genes: a genome-first approach of the Geisinger and UK Biobank cohorts

Authors:

S. Kim^{1,2}, J. Kim¹, J. Haley³, M. Ramos¹, S. Rao³, B. Graubard¹, DiscoverEHR collaboration, H. Katki¹, D. Carey³, D. Stewart¹; ¹Natl. Cancer Inst., Rockville, MD, ²Johns Hopkins Med. Inst., Baltimore, MD, ³Geisinger, Danville, PA

Abstract Body:

Background: Breast cancer is the most common cancer in women and approximately 5-10% of patients harbor a germline pathogenic variant in a tumor-susceptibility gene. For low-to-moderate penetrance breast cancer genes, there is uncertainty about their prevalence, penetrance, and phenotype (other associated tumors). We undertook a genome-first approach to characterize these features in a set of 15 genes associated with breast cancer risk. **Methods:** Using the genome first approach from Geisinger (n=175,449) and UK Biobank (n=200,600), we investigated 15 genes (*ATM*, *BARD1*, *BRCA1*, *BRCA1*, *CHD1*, *CHEK2*, *MSH6*, *NF1*, *PALB2*, *PIK3CA*, *PTEN*, *RAD51C*, *RAD51D*, *RECQL* and *TP53*) previously reported to be associated with breast cancer risk. In both cohorts, pathogenic/likely pathogenic (P/LP) variants in 15 genes were annotated using ClinVar and InterVar. For each individual harboring a P/LP variant, history of breast cancer and other cancers was determined using the Electronic Health Records (EHR) and relevant registries; family history of cancer was also determined. Odds ratios (using matched non-carrier controls) were calculated to assess the risk of breast cancer of individuals with a P/LP variant. **Results:** In Geisinger, 7,229/175,449 individuals (4.1%) harbored at least one P/LP variant in one of the 15 genes; in the UK Biobank, 3,978/200,600 individuals (2.0%) harbored at least one P/LP variant in one of the 15 genes. Of these individuals, 504/7,229 (7.0%) in Geisinger and 389/3,978 (10.0%) in UK Biobank were diagnosed with breast cancer. Of these, 11/7,229 (0.2%) and 4/3,978 (0.1%) were male in Geisinger and UK Biobank, respectively. In UK Biobank, the risk of breast cancer was significantly increased in the individuals with P/LP variants in *BARD1*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2* and *RAD51D* with the p-value <0.0001. In Geisinger, the risk of breast cancer was increased in *ATM*, *BRCA1*, *BRCA2*, *NF1*, *PALB2*, and *TP53* (P<0.05). In the UK Biobank among the patients with breast cancer, 1,501/3,978 (37.7%) had a family history of breast cancer (in the mother or siblings). **Conclusions:** Our results of the genome-first approach showed that P/LP variants in *BARD1*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *RAD51D*, *NF1* and *TP53* were associated with increased risk of breast cancer. We anticipate that the information on the other cancers related to these genes and the family history of cancers will help to plan on cancer screening and genetic counseling for the family.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3202. Brotman Baty Institute Clinical Variant Database (BBI-CVD).

Authors:

A. Folta¹, L. Santos², A. Sedeno Cortes¹, M. Horike-Pyne¹, M. Dubard-Gault², F. Hisama¹, A. Stergachis¹; ¹Univ. of Washington, Seattle, WA, ²Fred Hutchinson Cancer Ctr., Seattle, WA

Abstract Body:

Precision genomics has the potential to revolutionize clinical medicine by improving diagnosis and treatment of patients with common and rare disorders. However, clinical and genetic data for patients with rare disorders exist in silos, which limits the utility of extant variant databases, and inhibits collaboration between clinicians seeing patients with rare disorders and researchers studying the causative genes. To help overcome this, we have developed the Brotman Baty Institute Clinical Variant Database (BBI-CVD), which is a REDCap database of clinical germline genetic variant information from patients with rare genetic disorders seen at a tertiary academic medical center. This database captures demographic data, genetic testing indications, and genetic testing results for patients with rare genetic disorders cared for at the UW and/or Fred Hutchinson Cancer Center clinics. This database is available to researchers hoping to obtain detailed clinical information on individuals with a particular genetic variant, and we have developed a tiered system of data accessibility. Specifically, overview information regarding all the variants in our database is displayed through a web-based data explorer. Deidentified genetic and non-genetic diagnostic testing reports on participants within the BBI-CVD is available to any researcher upon request. Finally, researchers with separate IRB-approved protocols may request access to additional database information to identify individuals suitable for clinical trials that target specific genetic variants. While this database is primarily for the use of researchers, it also tracks important information for clinicians like indication frequency per gene, which will help clinicians monitor the use of genetic testing through the institution with the potential to change genetic testing guidelines. We will present preliminary data on the several hundred patients entered to date to demonstrate the utility of our database and its potential to advance precision genomics research. Overall, we hope that the BBI-CVD will serve as a bridge between clinicians and researchers working to resolve the many variants of uncertain significance, and enable development of new functional assays, diagnostic, and therapeutic approaches to rare diseases.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday
PB3203. Burden of Mendelian disorders in a large middle eastern biobank.

Authors:

W. Aamer¹, A. Al-Maraghi¹, N. Syed², A. Al-kurbi¹, G. D. Gandhi¹, O. Al-Saei¹, K. Al-Malki¹, S. Abbasi¹, N. Agrebi³, F. Abbaszadeh⁴, A. Akil¹, R. Badii⁴, N. Krishnamoorthy¹, B. Lo^{3,5}, The Qatar Genome Program Research Consortium, Y. Mokrab^{1,6}, K. Fakhro^{1,6,5}; ¹Lab. of Genomic Med., Sidra Med., Doha, Qatar, ²Applied Bioinformatics Core, Sidra Med., Doha, Qatar, ³Lab. of Immunoregulation, Sidra Med., Doha, Qatar, ⁴Diagnostic Genomic Div., Hamad Med. Corp., Doha, Qatar, ⁵Coll. of Hlth.and Life Sci., Hamad Bin Khalifa Univ., Doha, Qatar, ⁶Weill Cornell Med.-Qatar, Doha, Qatar

Abstract Body:

Genome sequencing has enabled the discovery of hundreds of Mendelian genes that are implicated in various phenotypes and diseases. Recently with the advance of large biobank cohorts, the interrogation of the mutational burden in seemingly healthy individuals has provided useful insight complementing small-scale family studies. Here, we interrogated 6,045 genomes from Qatar Biobank with over 60 associated qualitative/quantitative traits and assess Mendelian disease burden in this representative Middle Eastern population. By examining known pathogenic variants from Clinvar in 2,442 Mendelian genes, we identified 17 carriers with phenotypes or clinical history consistent with the pathogenic burden documented in the literature. We also identified novel pathogenic variants in known Mendelian genes including 3 putative disease-causing alleles with extreme quantitative traits. We performed rare-variant burden testing in Mendelian genes identifying 18 genes with high mutational load, six of which are related to a diabetes panel. This increased load is consistent with the high incidence of diabetes in Qatar and the region and suggests potential actionable options specific for these mutations. Novel pathogenic variants discovered in this cohort will be valuable for future research with relevance to the wider population in the region and worldwide and serve as useful resource for public health strategies for tackling genetic disease.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3204. Clinical Variants for Brain Gene Curation: A Powerful and Under-Utilized Resource.

Authors:

M. Chopra¹, J. Savatt², T. Bingaman², M. E. Good², M. Wangler³, K. German⁴, M. Wasserstein⁵, S. Walkley⁵, S. Molholm⁵, A. Gropman⁶, C. Arneron⁷, L. Dewalt⁷, S. Shankar⁸, S. Mahida¹, V. Lanzotti⁹, J. A. Martinez-Agosto¹⁰, K. Izumi¹¹, S. Srivastava¹, K. Dies¹², J. Piven¹³, H. Heather¹³, M. Sahin¹, E. R. Riggs², J. Constantino¹⁴, Brain Gene Registry Consortium (Intellectual Disability and Research Centers); ¹Boston Children's Hosp., Boston, MA, ²Geisinger, Danville, PA, ³Baylor Coll. Med., Houston, TX, ⁴Univ. of Washington, Seattle, WA, ⁵Albert Einstein Coll. of Med., Bronx, NY, ⁶Children's Natl. Hosp., Washington, DC, ⁷Univ. of Wisconsin, Madison, WI, ⁸UC Davis, Sacramento, CA, ⁹Washington Univ., Washington, MO, ¹⁰UCLA, Los Angeles, CA, ¹¹Children's Hosp. of Philadelphia, Philadelphia, PA, ¹²Boston Children's Hosp, Boston, MA, ¹³Univ North Carolina, Chapel Hill, NC, ¹⁴Washington Univ., Saint Louis, MO

Abstract Body:

Establishing gene-disease validity and delineating new neurodevelopmental disorders (NDDs) rely on publicly accessible variant and phenotypic data. Variants identified during routine clinical care have the potential to contribute novel data to inform gene curation and phenotypic spectrum but are often siloed in medical records or internal databases. Although laboratories are encouraged to share variants with public databases, workflows are variable and accompanying phenotypic information is often incomplete. Case reports are prone to bias towards severity, and those beyond the initial report are less likely to be published. Such barriers result in the under-utilization of clinically ascertained genetic and health data. The National Brain Gene Registry (BGR) is an initiative harnessing clinically ascertained variants together with standardised neurobehavioral phenotyping and electronic health record (EHR) data to understand the role of genes implicated in NDDs. To evaluate our hypothesis that clinically ascertained variants are under-represented in publicly available resources, we selected 55 genes, focusing on those not previously curated for intellectual disability/autism by ClinGen (n=30), and those for which additional evidence could potentially change the existing classification (n=16). Eight BGR clinical sites searched for variants (VUS or above by ACMG criteria) in these genes using laboratory databases, EHRs and internal clinical databases. Variants were evaluated for presence in 7 disease databases - ClinVar, DenovoDB, Vericata, ADAMI DBD, LitVar, Decipher and HGMD. The search yielded 471 unique variants in 54 genes. Overall, 273 (58%) variants were absent in ClinVar, and 253 (54%) were absent from all databases, demonstrating the potential for this initiative to contribute novel data to the public knowledgebase. Furthermore, access to case-level data may impact variant classification. For the 198 variants present in ClinVar, 39 (20%) had aggregate ClinVar classifications of “conflicting” and 58 (29%) had aggregate classifications of “VUS.” In these instances, BGR participant phenotype and/or inheritance information could resolve discrepancies and shift to more definitive classifications. There were 40 variants (36%) with an aggregate ClinVar classification that conflicted with the BGR site-provided classification; this information could potentially be relayed to participants and result in a change in diagnostic course. The co-registration of genetic variant, phenotype and EHR data has the power to inform gene curation, the phenotypic spectrum of new and emerging Mendelian NDDs, and variant pathogenicity.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3205. Colorado Biobank Portal is a cloud platform for the interactive exploration of genome-wide association studies results

Authors:

N. Pozdeyev¹, N. Rafaels¹, S. Chavan¹, T. Phang¹, M. Lin¹, D. Mayer¹, I. M. Brooks¹, L. K. Wiley¹, R. Mathias², K. Barnes¹, C. R. Gignoux¹; ¹Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ²Johns Hopkins Univ, Baltimore, MD

Abstract Body:

The number of genome-wide association studies (GWAS) is growing rapidly, increasing the need for interactive and functional web tools to explore large-scale GWAS summary statistics data. We developed Colorado Biobank Portal (CBP, <http://hdc-sandbox-bioengine.uw.r.appspot.com/>), a web portal that provides access to GWAS results for 1258 phenotypes for ~34K participants in the Colorado Center for Personalized Medicine (CCPM) Biobank. Phenotypes were defined with phecodes, and GWAS analyses were performed using REGENIE whole genome regression modeling software. To increase computational efficiency and reduce cost, up to 10 phenotypes clustered by data missingness were analyzed in a single run. CBP stores statistical analyses results, Ensembl Variant Effect Predictor annotation, information on exon, transcript and gene structures, gene aliases, and rsID identifiers in a Google BigQuery database. Each variant is assigned a unique integer coordinate spanning across all chromosomes. The database was partitioned by integer coordinate to allow for efficient data query and retrieval. The application's data access layer was programmed in the Python Flask web framework. Visualization templates were written in HTML, CSS, and JavaScript and were borrowed from the Global Biobank Engine open source code (<https://github.com/rivas-lab/GlobalBiobankEngine>) or newly created. The CBP is deployed on the Google App Engine managed platform. Innovations of the CBP include: 1) Scalable design enabling rapid and inexpensive access to ~53 billion GWAS summary statistics. 2) Support for both "horizontal" (by variant, region, transcript, or gene) and "vertical" (by phenotype) data requests. 3) Real-time phenome-wide association analysis for any of the 49.4 million genetic variants. 4) Low cost and maintenance-free serverless deployment. 5) Convenient user access and management with the help of the Google Identity platform. The CBP is free to use by the researchers at the University of Colorado and by collaborators from outside academic institutions sponsored by CCPM faculty. This technology can be used by other Biobanks worldwide and is being further developed to support new types of analyses such as genetic ancestry and polygenic risk score estimates.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3206. CombVar: A pipeline to combine variant impact from prioritizers, integrators, and annotations

Authors:

A. Nato¹, V. Magalhães Borges¹, R. Mingroni-Netto², A. V. R. Horimoto³; ¹Dept. of BioMed. Sci., Joan C. Edwards Sch. of Med., Marshall Univ., Huntington, WV, ²Dept. of Genetics and Evolutionary Biology, Univ. of Sao Paulo, Sao Paulo, SP, Brazil, ³Dept. of Biostatistics, Univ. of Washington, Seattle, WA

Abstract Body:

Variant annotation and prioritization are strategies known to facilitate the identification of genetic variants that have higher contributory effect to a disease. However, numerous variant prioritizers, integrators, and annotations exist, such that it may be challenging to specifically select which one to use for a particular dataset. Here, we develop CombVar, a pipeline written in R that employs an algorithm to combine variant information from multiple annotation sources as well as variant prioritizers and integrators. For the preliminary version of this pipeline, we included: (1) variant prioritizers and integrators (i.e., VAAST Variant Prioritizer (VVP), Ensembl Variant Effect Predictor (VEP), UCSC Variant Annotation Integrator (VAI), Combined Annotation Dependent Depletion (CADD), and Machine Learning Var (MLVar)) and (2) variant annotation databases (i.e., ANNOVAR, SIFT, PolyPhen-2, and Mutation Taster). CombVar provides salient information about variants being analyzed, including: (1) general annotation (e.g., alternative sequence, type of mutation, location, gene, transcript, and frequency), (2) classification based on ACMG/AMP guidelines, and (3) factors overlapping with the variant (e.g., epigenetic elements). For each variant, CombVar estimates an overall score λ , which ranges from 0 to 10 (with 10 being the highest), based on the impact of the variant, making it useful in evaluating its degree of pathogenicity from different sources. Users will be able to choose the prioritizer(s) and/or annotation(s) and the corresponding weights specific for their analysis. Using CombVar, we analyzed 20 associated genetic variants with suggestive/significant level of significance in our in-house essential hypertension dataset. For this test, equal weight was used for all sources. The scores ranged from 0.17 to 4.22. The top 3 genetic variants with the greatest probable impact on the phenotype are: rs1805087 ($\lambda= 4.22$), rs79656879 ($\lambda= 2.34$), and rs12602540 ($\lambda= 2.08$).

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday
PB3207. Concordance in variant calls from short-read sequencing data of admixed individuals using different reference genomes

Authors:

T. de Oliveira, W. Souza, B. Carvalho, I. Lopes-Cendes; Univ. of Campinas, Campinas, Brazil

Abstract Body:

Introduction: The recent completion of the human genome (CHM13) by the T2T consortium has identified the missing 8% of the human genome contained in the references GRCh37 and CRCh38. The present work compares the impact of using the different human genome assemblies as a reference for variant identification in exome data generated from admixed Brazilian individuals. **Materials and Methods:** We analyzed 318 exomes from the Brazilian Initiative of Precision Medicine (bipmed.org). Raw sequences were aligned to GRCh37, CRCh38, and CHM13 using BWA-MEM. Local realignment and variant calling were performed with the GATK best practices workflow. We used the PLINK software to perform several filtering strategies, including thresholds in missing genotypes and samples, relatedness, and heterozygosity deviation. Variant lift-over and comparisons among the variant datasets were performed using the liftover tool. Concordant and discordant variants were compared using vcfeval. Identification of discordant reference patches (DISCREPs) and enrichment analysis for genomic features were performed as described in <https://github.com/hurleyLi/discreps>. **Results:** We identified similar numbers of filtered variants (~175,000) when using the three reference assemblies. However, approximately 7.6% of the variants were discordant when comparing calls using the different assemblies. Furthermore, around 60% of these discordant variants were uniquely identified in the CHM13 assembly. Also, we found DISCREPs in 0.6% of the tested genomic 10kb-windows, and these were enriched for variants in the comparisons between GRCh37 and CHM13-GRCh37 and when comparing GRCh38 and CHM13-GRCh38 (rate $q < 0.01$). Furthermore, we found that DISCREPs were also enriched in alternate haplotypes called with the GRCh38 assembly and in single tandem repeats found with the GRCh37. **Conclusion:** Our results indicate that discrepancies in variant calls are seen predominantly in regions known to contain repeat sequences. Also, these differences do not seem to be due to problems in the assemblies of the previous reference genomes. This is an ongoing project, and we plan to perform additional annotations using the CHM13 reference, including the verification of gene enrichments which will help us better quantify the impacts of the changing reference genomes on common and rare variants. We believe that our results are relevant for improving genetic diagnosis practices in admixed individuals.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3208. Distance-based panel generation optimizes gene selection for targeted gene panel design

Authors:

O. Isakov^{1,2,3}, S. Ben Shachar^{1,3,4}; ¹Clalit Res. Inst., Innovation Div., Clalit Hlth.Services, Tel Aviv, Israel, ²Raphael Recanati Genetic Inst., Rabin Med. Ctr.-Beilinson Hosp., Petach Tikva, Israel, ³The Ivan and Francesca Berkowitz Family Living Lab. Collaboration, at Harvard Med. Sch. and Clalit Res. Inst., Boston, MA, ⁴Sackler Faculty of Med., Tel Aviv, Israel

Abstract Body:

Background: Targeted gene panel sequencing is used to limit the search for causative genetic variants to genes associated with a certain phenotype. Collaborative tools, such as PanelApp, facilitate gene panels design based on expert review and include mostly genes with a strong association to the phenotype. This approach limits the sensitivity of the panel to identify causative variants in genes with weaker phenotypic association. Available commercial panels differ significantly from one another and some include genes without a clear association to the phenotype. Taken together, designing or selecting the most appropriate gene panel for a given phenotype is often challenging.

Aim: Review the gene content in panels for various phenotypes from different vendors world-wide and compare them with curated gene panels. Develop a method to improve gene panel design based on the collective wisdom represented by the publicly available gene panels.

Methods: Publicly available gene panel data was downloaded from the Genetic Testing Registry (GTR) and the PanelApp web server. Panels corresponding to 20 different phenotypes were selected for analysis. Naive-search panels were generated by unifying all the genes in all the panels identified by a naive text based phenotype search. Inter-panel gene content distance was calculated using the Jaccard index.

Distance based panels were generated by collecting all the genes found in panels similar to a prespecified PanelApp panel and scoring them according to the distance of their original panel from the PanelApp panel. For each gene in each panel, a literature search was run in order to elucidate the number of publications associating it with the phenotype. Performance metrics were then calculated for each panel based on the cumulative identified gene-phenotype associations.

Results: A total of 6053 panels from 225 different vendors were found in the GTR. For the selected phenotypes, a naive search identified, on average, 25.6 panels per phenotype (IQR [16.75,29.5]). High inter-panel discrepancy was noted for most phenotypes, with almost half of the genes (45.9%) IQR[35.6%,68.4%] found in less than 10% of the collected panels. Compared to the naive search panel, the distance based panel achieved a higher PPV while maintaining comparable sensitivity (difference = 0.05, 95% CI [0.01, 0.10]). Compared to the curated PanelApp panels, the distance based panels demonstrated a higher F1 score (difference = 0.08, 95% CI [0.04, 0.11]).

Conclusion: Panels generated using a distance based method maintain accurate gene content compared to curated panels while enabling a more comprehensive inclusion of phenotypically relevant genes.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3209. Dynamically querying thousands of genomes to identify genetically matched cohorts with Intel® Optane™ Persistent Memory.

Authors:

K. Schneider¹, M. Chowdhury¹, M. Tepper², J. Khan², S. Dongaonkar², C. Gignoux³, R. Layer⁴; ¹Univ. of Colorado Boulder, Boulder, CO, ²Intel Labs, Hillsboro, OR, ³Univ of Colorado Denver, Anschutz Med. Campus, Aurora, CO, ⁴Univ. of Colorado, Boulder, CO

Abstract Body:

Reproduction, migration, and random mutation events over hundreds of millennia have contributed to incredible genetic diversity across the human population. The nuance of this genetic diversity contributes to challenges for computation methods in precision medicine. For example, calculating polygenic risk scores (PRSs) is an effective way to identify an individual's risk of disease onset. This calculation requires a comparative analysis against the individual, a large cohort of individuals who present with some diseases, and a large cohort who do not. However, PRSs have little application across distinct genetic background (i.e. PRSs calculated across cohorts of European ancestry will not accurately describe risk for individuals of African descent). To mitigate this limitation, current methods use pre-defined cohorts of ancestry (e.g. European, African, Puerto Rican, etc.) to compute respective ancestry-based PRSs. These pre-defined cohorts still do not capture the complete diversity of the human population, and do not address the complexity of population-wide data being introduced by emerging datasets (e.g. H3Africa, PRS Diversity Consortium, IGNITE, eMERGE, etc.). To address these problems, it is necessary to identify cohorts of genetically similar individuals among a large population, disparate from pre-defined ancestry, with speed and accuracy. We propose a method to dynamically identify cohorts of genetically matched individuals by (1) performing highly accurate similarity searches for full-population, genome-wide analyses; and (2) enable precision queries for new samples. Our method adopts a similarity search method which leverages the Intel® Optane™ Persistent Memory system. We included custom encoding for phased genotypes on genetic variation data; and designed novel query algorithms to allow for quick whole-genome cohort identification for incoming samples. Our method improves upon current methods in both accuracy and speed of search, and introduces the novelty of dynamic searching for new patients. Adoption of this method will be a first step in precision medicine pipelines to more accurately describe how genetic variation between individuals contributes to the risk, onset, and progression of many known human diseases.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3210. Early development of a locus specific database for *GUSB*, the gene associated with Mucopolysaccharidosis type VII: hints of a higher predicted prevalence.

Authors:

S. Daugherty¹, V. Rangel Miller¹, D. Garica¹, R. Giugliani², O. K. Japalaghi¹, D. Marsden¹, H. McLaughlin³, A. Willcock⁴, M. Hegde⁵, N. Miller¹; ¹Ultragenyx Pharmaceutical, Novato, CA, ²HCPA/UFRGS, Porto Alegre, Brazil, ³Invitae, South Lyon, MI, ⁴Invitae Corp., San Francisco, CA, ⁵PerkinElmer, Lilburn, GA

Abstract Body:

Accurate interpretation of rare genetic variants is essential for diagnosis and, in ultra-rare disease, is confounded by variants of uncertain significance (VUS). Mucopolysaccharidosis type VII (MPS VII) is an ultra-rare, autosomal recessive, lysosomal storage disease caused by β -glucuronidase (*GUSB*) enzyme deficiency. MPS VII may present early with nonimmune hydrops fetalis (NIHF), which can lead to death in utero or early infancy, or may present later with skeletal dysplasia, dysmorphology and cardio-pulmonary signs. MPS VII is estimated to occur in 0.02-0.027/100,000 newborns with an overall prevalence of 0.07/million, but we hypothesize this number underestimates true frequency due to unrecognized atypical disease and perinatal lethal cases.

We integrated data from public databases and Ultragenyx clinical and sponsored genetic testing programs through March 2022 to develop a *GUSB* gene database. Prevalence estimates were calculated using allele frequency data from gnomAD and a published Bayesian estimation method for analyzing ultra-rare disease frequency. A total of 231 unique *GUSB* variants were collected and 189 were associated with an ACMG classification criteria: pathogenic (P) or likely pathogenic (LP) 61 (33%), VUS 118 (62%), benign 10 (5%). 90 (39%) of these variants were not present in ClinVar, including 34 published variants from HGMD. To predict prevalence of MPS VII based on these data, we first analyzed all P and LP variants. The Bayesian estimation gave a prevalence of 1.49/million (95% CI= 1-2/million) or \sim 1/670,000. Given the high rate of VUS, we next included all VUS variants with a Combined Annotation Dependent Depletion (CADD) phred score >20 , adding 27 VUS to the analysis, which increased the Bayesian estimation of prevalence to 9.11/million (95% CI = 7.5 -10.7/million) or \sim 1/110,000. This dataset revealed the high frequency of *GUSB* variants classified as VUS, underscoring the importance of parallel biochemical and molecular diagnosis of MPS VII. Our prevalence estimates suggest a higher potential prevalence of MPS VII and future studies are warranted to understand whether perinatal lethality, lack of diagnosis of atypical MPS VII, or both depress the predicted population frequency. We plan to expand the *GUSB* dataset with additional global evidence to further refine the prediction. Data-sharing efforts are critical to support accurate and consistent variant interpretation in ultra-rare disease, particularly for newborn screening (NBS) as MPS VII is screened for in multiple NBS pilots around the world.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3211. Efficient querying of genomic reference databases with *gget*

Authors:

L. Luebbert, L. Pachter; California Inst. of Technology, Pasadena, CA

Abstract Body:

A recurring challenge in interpreting genomic data is the assessment of results in the context of existing reference databases. Currently, there is no tool implementing automated, easy programmatic access to curated reference information stored in a diverse collection of large, public genomic databases. *gget* is a free and open-source command-line tool and Python package that enables efficient querying of genomic reference databases, such as Ensembl. *gget* consists of a collection of separate but interoperable modules, each designed to facilitate one type of database querying required for genomic data analysis in a single line of code. The manual and source code are available at <https://github.com/pachterlab/gget>.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3212. Efficient reinterpretation of rare disease cases using Exomiser

Authors:

L. Vestito¹, Genomics England Research Consortium, Monarch Initiative, J. Jacobsen¹, P. Robinson², D. Smedley¹; ¹Queen Mary Univ. of London, London, United Kingdom, ²The Jackson Lab. for Genomic Med., Farmington, CT

Abstract Body:

High-throughput and lower-cost sequencing have enabled the integration of whole-exome and whole-genome sequencing (WES/WGS) in clinical practice and the advent of large sequencing projects, like the UK's 100,000 Genomes Project. These extraordinary advances enabled the discovery of hundreds of novel gene-disease associations per year, published in peer reviewed journals. However, despite this impressive discovery rate, a substantial fraction of patients (75-80%) remain currently undiagnosed following WES/WGS. To address this, a periodic reinterpretation of the genetic data for undiagnosed individuals has been proven to increase the diagnostic yield by 10-15% and has the potential to bridge the gap between the dynamic scientific knowledge and clinical practice for patient benefit. A new diagnosis can lead to tailored therapeutic management, more accurate genetic counselling and better prognosis. Exomiser, a phenotype-driven tool to annotate, filter and prioritise likely causative variants, integrates numerous data sources and leverages information on variant frequency, predicted pathogenicity and semantic similarity between the Human Phenotype Ontology (HPO) terms describing patient's phenotype and the phenotypic annotations of human diseases (OMIM/Orphanet), orthologs in mice (MGI/IMPC) or zebrafish (ZFIN) model organisms, and phenotypes of protein-protein associated neighbours (STRING). We tested and compared the ability of seven releases of the Exomiser database from Feb 2019 until Feb 2022 to detect 37 real patient diagnoses from the 100,000 Genomes Project, in one of the 556 disease-gene associations that first appeared in OMIM since Feb 2019. We investigated in detail the distribution of Exomiser variant, phenotype, human phenotype and combined scores over time and settled on a threshold of variant score ≥ 0.8 and an increase in human phenotype score ≥ 0.2 between Exomiser releases as the optimal way to detect candidates based on novel emerging evidences published in clinically-relevant databases such OMIM, Orphanet or ClinVar. These thresholds highlighted 54 variants of interest in the 37 solved cases, with 31 being the correct diagnosis, representing impressive recall (84%) and precision (57%). Over time, Exomiser also successfully re-classifies these variants from variants of unknown significance (VUS) to pathogenic (P) or likely pathogenic (LP) in 34/37 of the cases, following an automated ACMG/AMP variant classification pipeline.

In conclusion, our data show the ability of Exomiser to highlight and re-classify novel candidates for diagnosis over time in light of novel emerging evidence of gene-disease associations.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3213. Genomic database query tracking to understand the prevalence of rare diseases: *ASAH1* in the VarSome database

Authors:

N. Confer¹, K. Crosby¹, A. Joyner², S. Solyom¹; ¹Aceragen, Durham, NC, ²Saphetor, Lausanne, Switzerland

Abstract Body:

Pathogenic variants in *ASAH1* cause acid ceramidase deficiency with a spectrum of phenotypes including Farber disease and spinal muscular atrophy with progressive myoclonic epilepsy (SMA-PME). *ASAH1*-related conditions are considered ultra-rare with fewer than 250 cases reported in the medical literature, but the prevalence is likely under-reported. VarSome is a community-driven search engine of human genome variant data that allows clinical and research professionals to search more than 130 harmonized datasets within a single interface. VarSome data searches are utilized in over 150 countries by hundreds of thousands of users to perform genome-wide variant evaluations by ACMG guidelines. In this study we used the VarSome database to aggregate *ASAH1* queries over the past two years to quantify the total number of variants, unique users and countries represented. These data were compared to historical literature to improve the prevalence estimation for Farber disease and SMA-PME, globally and by country. Since the first case of Farber disease in 1952, approximately 200 additional cases of Farber and 50 cases of SMA-PME have been published. These cases are from 32 countries with Egypt (14% of cases), the United States (13%), Iran (9%), France (9%) and Italy (8%) being the most represented. A recent review article lists 73 total published pathogenic/likely pathogenic (P/LP) variants in the *ASAH1* gene. For comparison, from May 2020 to May 2022, there were 787 distinct queries in the VarSome database for variants in the *ASAH1* gene. These queries were submitted by 717 unique users in 61 different countries, with Iran (16% of queries), Turkey (11%), Spain (8%), India (6%), and Russia (5%) being the most represented. There were 250 unique variants represented in these queries, with 69 P/LP (28%), 48 benign/likely benign (19%), and 133 variants of uncertain significance (53%). The number of countries and unique users represented in VarSome queries over the previous 2 years suggests the prevalence of *ASAH1*-related conditions is underestimated. This use of a large consolidated variant database to estimate prevalence for rare diseases may be superior to traditional literature-based approaches, while also essential to the conducting of natural history studies and successful enrollment of patients in clinical trials. These findings further support that published cases often under-represent the prevalence and geographic distribution of a rare disease.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3214. Global landscape of primary omics data generation and its secondary analysis across 193 countries and territories

Authors:

Q. Peng¹, N. Darci-Maher², Y. Patel¹, M. Vahed¹, A. Rathore³, G. Sharma³, S. Mangul¹; ¹Univ. of Southern California, Los Angeles, CA, ²Univ. of California, Los Angeles (UCLA), Los Angeles, CA, ³Inst. of Bioinformatics and Applied Biotechnology (IBAB), Bengaluru, India

Abstract Body:

Traditionally, novel scientific discoveries were a privilege of researchers who generated data, and no further analyses were conducted after the generation of original data. However, as we entered the era of omics, researchers have gradually attached more importance to re-analysis of existing open-access data. Instead of generating new data in the lab, some researchers tend to analyze existing data for new hypotheses in the form of a new research question or an alternative perspective on the original question. This process, known as secondary analysis, may enable novel discoveries across a variety of disciplines, including computational biology and medicine. Also, the secondary analysis excludes the cost of data-generating, therefore, helping in economic sustainability, making it a suitable research technique for countries with limited resources. Despite numerous advantages of secondary analysis of omics data, the scale of the secondary analysis across various countries of omics data is currently unexplored. Also, it remains unidentified whether secondary analysis was adopted more in developing countries or remains the privilege of developed countries. To address these emerging questions, we have performed a comprehensive analysis of omics datasets submitted via open-access, their sources, and the reusability of public omic datasets based on 4.2 million open-access publications published between 2001 and 2021. Among these publications, 148,239 papers containing SRA or GEO accession numbers were classified as studies generating data or performing secondary analysis. The selected papers were further classified as primary analysis or secondary analysis papers. The classification was done by comparing the publication date of a given dataset to the publication date of each paper including the dataset. The earliest-published publication was marked as primary analysis, and the rest of the publications (if they existed) were marked as secondary analysis. The countries where the data was generated and/or data analysis was performed were identified. The affiliated country of the last author was considered the country for the respective dataset. These data uncovered the scales and features of secondary analysis across 193 countries and territories. In this study, we also discussed factors limiting secondary analysis in countries with limited resources were discussed. Lastly, we identified challenges associated with secondary analysis and needs in bioinformatics training.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3215. Growing a research ecosystem in the cloud: Early insights from the All of Us Research Program.

Authors:

C. Lunt¹, K. Mayo², H. Master²; ¹NIH, Bethesda, MD, ²Vanderbilt, Nashville, TN

Abstract Body:

Background: The NIH *All of Us* Research Program makes phenotypic and genomic data available for more than 350,000 participants on a cloud-based research platform, the Researcher Workbench (RW). Approved researchers conduct projects in collaborative workspaces. To promote transparency, information about each workspace's research purpose and team are publicly available. De-identified, row-level phenotypic data has been available since May 2020, and the program released a new "controlled tier" of data in March 2022, containing the first nearly 100,000 whole genome sequences from *All of Us* participants. Here, we demonstrate how workspace usage and description data can be used to characterize the current *All of Us* research ecosystem. Specifically, we describe the current researcher community and examine emerging trends and research topics.

Methods: Researcher demographic information and workspace data were collected during researcher account and workspace creation and analyzed to inform RW utilization and collaboration. Additionally, Natural Language Processing was used to evaluate the emerging research topics from workspace descriptions.

Results: As of June 2022, the *All of Us* researcher community is composed of more than 2,000 researchers across different career stages and settings. Over 83% describe themselves as students, trainees, or early career investigators. These researchers represent 354 institutions from more than 45 states. More than 1,500 workspaces use "registered tier" data, and span a broad range of topics including education and methods development. Most research is centered on disease, and our analysis suggests that diabetes, cancer, and cardiovascular disease are the 3 most analyzed topics. Furthermore, researchers have created more than 400 controlled tier workspaces. Given the availability of diverse genomic data in this tier, a significant number of projects are focused on understanding health and disease in minority populations.

Conclusion: Cloud-based analysis platforms like the RW enable real-time characterization of growing research communities. Our findings suggest that this resource may help develop a biomedical research workforce hungry to advance our understanding of health in underserved populations.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3216. HLA-SPREAD: a natural language processing based resource for curating HLA association from PubMed abstracts

Authors:

D. Dholakia; CSIR-IGIB, Delhi, India

Abstract Body:

Extreme complexity in the Human Leukocyte Antigens (HLA) system and its nomenclature makes it difficult to interpret and integrate relevant information for HLA associations with diseases, Adverse Drug Reactions (ADR) and Transplantation. PubMed search displays ~ 146,000 studies on HLA reported from diverse locations. Currently, IPD-IMGT/HLA (Robinson et al., Nucleic Acids Research 48:D948-D955, 2019) database houses data on 28,320 HLA alleles. We developed an automated pipeline with a unified graphical user interface HLA-SPREAD that provides a structured information on SNPs, Populations, REsources, ADRs and Diseases information. Information on HLA was extracted from ~ 28 million PubMed abstracts extracted using Natural Language Processing (NLP). Python scripts were used to mine and curate information on diseases, filter false positives and categorize to 24 tree hierarchical groups and named Entity Recognition (NER) algorithms followed by semantic analysis to infer HLA association(s). This resource from 109 countries and 40 ethnic groups provides interesting insights on: markers associated with allelic/haplotypic association in autoimmune, cancer, viral and skin diseases, transplantation outcome and ADRs for hypersensitivity. Summary information on clinically relevant biomarkers related to HLA disease associations with mapped susceptible/risk alleles are readily retrievable from HLA-SPREAD. The resource is available at URL <http://hla-spread.igib.res.in/>. This resource is first of its kind that can help uncover novel patterns in HLA gene-disease associations.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday
PB3217*. How UK Biobank is democratising access to large-scale genomic and phenotypic data for discovery science

Authors:

B. Lacey¹, N. Allen², M. Effingham³, M. Pancholi³, L. Carson³, L. Burkitt-Gray³, R. Collins¹; ¹UK Biobank, Univ. of Oxford, Oxford, United Kingdom, ²UK Biobank, Univ. of Oxford, OXFORD, United Kingdom, ³UK Biobank, Stockport, United Kingdom

Abstract Body:

UK Biobank is scientifically unparalleled as a biomedical research study and is now arguably the world's most important health research resource. With its unique combination of scale, depth, maturity and accessibility, the resource is enabling researchers worldwide to make scientific discoveries that improve human health. UK Biobank has collected (and continues to collect) detailed data on genetics, biomarkers and lifestyle factors from half a million people across the UK, whose health has been followed up for about 15 years through linkage to electronic health records.

UK Biobank has integrated large-scale genomic data (in 500,000 participants) with deep phenotyping data (including lifestyle factors, physical measures, accelerometry and multi-modal imaging), coupled with long-term longitudinal health records. The ability to leverage significant funding from public-private partnerships to perform cohort-wide whole-exome and whole-genome sequencing (with genome-wide genotyping and imputation already available) has made UK Biobank one of the world's most important resource for identifying rare genetic variants that are particularly valuable for drug discovery. The addition of large-scale proteomic data is creating an even more powerful resource for the identification of protein quantitative trait loci for protein biomarkers of thousands of phenotypic traits; together with large-scale metabolomics, transcriptomics and whole-body imaging data, UK Biobank will enable a better understanding of disease biology and will support innovative drug development for more effective therapies.

To accommodate the rapid growth of the resource and to enable more researchers across the world to access these data without limitations of transferring, collating, storing, and accessing data at this scale, UK Biobank launched an innovative cloud-based Research Analysis Platform (RAP) in 2021, democratising access to large scale compute and novel technologies. The availability of financial research credits for early-career researchers and those from low and middle-income countries is further democratising access to this unique resource. The RAP also serves to foster greater collaboration between researchers around the world by allowing users to analyse multiple data types together and to benefit from the returned data from other researchers within the cloud-based platform.

Ready access for researchers around the world to the combination of in-depth genomic, other -omic, imaging, lifestyle and health information from half a million UK participants is enabling novel research that is advancing discovery science and improving human health.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3218. Identification and characterization of genomic variations from population scale whole genome sequencing in south India.

Authors:

B. Kahali; Ctr. for Brain Res., Indian Inst. of Sci., Bangalore, India

Abstract Body:

Representation of Indians in human genetic studies has been limited so far. Therefore, population-specific reference genome datasets as well as genome-wide association studies in Indian population are warranted. India's population of 1.3 billion comprising >4,500 well defined ethnic groups could have numerous distinct genetic variations resulting from the rich genetic and social diversity of the country, multiple waves of migration into the country over the past thousands of years, and unique socio-cultural practices followed by extant population groups. Here, we present a comprehensive analysis of SNPs and InDels in about 650 individuals from the TATA Longitudinal Study of Aging (TLSA) and Srinivaspura Aging Neurosenescence and Cognition (SANSCOG) population-based cohorts, including data from more than twenty population groups across South India. We identify and characterize about 30 million SNPs and 4.5 million InDels from deep (40X) coverage short-read whole genome sequencing data. The rare variants hugely outnumber the common variants, and about two-thirds of our identified SNPs and InDels are present in <1% individuals in our dataset. 3% of our identified variations are present in coding regions. Out of 12,000 frameshift causing variations, 4300 are rare in our study. We find that on an average each individual can carry 20 variants that are intolerant to protein coding loss of function changes. We elucidate the population structure and disease associations inferred from these genetic variants. We also report the performance of these genetic variants in imputation, to facilitate genetic studies in Indian populations across Asia and worldwide. This work will help pilot the detection and characterization of genetic variations in understudied population based and disease cohort studies and help address the underrepresentation of genetic variant discovery from non-European, especially South Asian populations. Finally, this dataset will be a resource for designing, and interpreting large scale association and functional genomic studies in Indian population that could pave the way for devising better diagnosis and treatment strategies for a substantial proportion of world population.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3219. Identifying Variants of Interest in Samples with Whole Genome Sequencing Data using Kids First Variant WorkBench

Authors:

Y. Guo¹, J. Costanza², A. Resnick¹, V. Ferretti², Kids First Data Resource Center; ¹The Children's Hosp. of Philadelphia, Philadelphia, PA, ²Ctr. Hosp.ier Univ.ire Sainte-Justine, Montréal, QC, Canada

Abstract Body:

Kids First Data Resource Center developed Kids First Data Resource Portal (KFDRP; <https://portal.kidsfirstdrc.org/>), which is a centralized data platform for both Kids First and collaborative cohorts. The recently released KFDRP component named Variant WorkBench (VWB) enables users to query, mangle, analyze and visualize germline genomic variants. We demonstrate the feasibility of analyzing germline Whole Genome Sequencing (WGS) data using VWB for customized families whose vcf files are provided by the user. The VWB uses Zeppelin notebooks for scripting and the current effort has four steps which are all written in notebook paragraphs. The first step is database initialization, loading variant annotations in the public domain, such as Cancer Hotspots, ClinVar, COSMIC, dbNSFP, gnomAD, TOPMed, as well as gene-phenotype links provided by OMIM, HPO, Orphanet, and the Deciphering Developmental Disorders Project. Then as the second step, we use Glow, an open-source toolkit for working with genomic data at biobank-scale and beyond, to read in the vcf file (either single- or multi-sample) then store the vcf file as a dataframe. In the third step, we join the newly generated dataframe with tables initialized in the first step to annotate the variants, and label variants with inheritance patterns according to the observed genotypes stored in the dataframe. Lastly, we flag a list of variants satisfying the user's predefined criteria such as minor allele frequency thresholds in general population, variant deleteriousness assessment scores, and gene-disease correlations, then export this list into an external file in the Cavatica system, enabling further review. As an example, we selected a family trio of a boy with kidney and urinary tract defects and healthy parents, input a three-sample vcf file into VWB, ran the notebook and discovered a missense variant in *TRAP1* which is homozygous in the proband and heterozygous in both parents. This variant is reported before as pathogenic in a cohort with Congenital Anomalies of the Kidney and Urinary Tract, indicating the ability of using VWB to identify variants linked to genetic disorders.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3220. Indiana Biobank: a resource of linked electronic health records, omics data, genetic data, and biospecimens for the research community.

Authors:

D. Lai¹, T-H. Schwantes-An¹, A. Roberts², M. Abreu¹, J. Lee¹, M. Zhang¹, B. Patz¹, T. Foroud¹; ¹Indiana Univ. Sch. of Med., Indianapolis, IN, ²Regenstrief Inst., Indianapolis, IN

Abstract Body:

Indiana Biobank (<https://indianabiobank.org/>) is a state-wide collaborative effort that combines state-of-the-art centralized biobanking and linked electronic health records (HER). Each participant's de-identified EHR are available via the Indiana Network for Patient Care (INPC) supported by the Regenstrief Institute at Indiana University School of Medicine. Currently, Indiana Biobank has DNA from over 50,000 broadly consented participants and actively adds over 100 new participants each week. More than 9,000 samples have been genome-wide genotyped and more than 6,000 samples have been whole exome sequenced; the rest of the samples will be genotyped and sequenced in the near future. Additionally, transcriptomics, metabolomics, and HLA haplotype data are available in subsets of participants. The EHR includes basic demographic information, diagnoses based on ICD9/10 codes, hospitalization information, laboratory test results, and prescription drug information obtained from patients at each visit, all de-identified by the Regenstrief Institute. Furthermore, information extracted from the physicians' note using natural language processing are also available, which is especially important for rare diseases not indexed by ICD codes. Recently, COVID-19 testing, vaccination, and hospitalization data from over 27,000 participants (including 4,700 with genotype data) were added. For samples that have been genotyped, 42.71% are male; 76.44% are white; and 21.87% are black. The average age at enrollment and last visit are 53.51 (SD=15.36; range: 1.77-96.68) and 58.43 (SD=15.33; range: 4.66-101.80), respectively. The majority of Indiana Biobank participants are unrelated. Based on ICD9/10 codes, >506 diseases have >100 cases and these data can be included in meta-analyses or for replication of previous genome-wide association studies (GWAS) findings. We have calculated >2,100 polygenic risk scores (PRS) from The Polygenic Score Catalog. Phenome-wide GWAS (PheGWAS) and genetic correlation studies are ongoing. With appropriate approvals, summary information (GWAS, PRS, and genetic correlations, etc.) will be made available to academic investigators via an online searchable platform. The Indiana Biobank is a growing resource for the research community that combines EHR, omics, and genetic data to improve patient care.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3221. JBrowse 2: a modular genome browser for visualizing synteny and structural variants

Authors:

I. Holmes¹, C. Diesh¹, R. Buels¹, G. Stevens¹, T. Martinez¹, S. Dider¹, C. Bridge², S. Cain², R. Haw², L. Stein³; ¹Univ. of California, Berkeley, CA, ²Ontario Inst. for Cancer Res., Toronto, ON, Canada, ³Ontario Institute for Cancer Res., Toronto, ON, Canada

Abstract Body:

Genome browsers are useful tools for research in human genetics, able to display and integrate biological information such as long and short read sequencing data, variant calls, and annotations. While traditional linear genome browsers have demonstrated broad utility for many applications, there is an increasing need for visualizing complex structural variants (SVs) beyond the linear view. To meet these challenges, we created JBrowse 2: an extensible open-source platform for visualizing and integrating genomic data. JBrowse 2 is a modular general purpose genome browser that can display linear, circular, synteny, and other types of views. To facilitate academic collaboration and innovation, JBrowse 2 has been designed with a pluggable interface that permits the development of features tailored to the specific needs of an individual or organization.

Unique to JBrowse 2 are its specialized structural variant views. JBrowse 2 can display discontinuous regions side-by-side or vertically stacked. Large or inter-chromosomal structural variants can be visualized with the read evidence (split long reads or paired end reads) shown with connected lines. Users can also visualize de novo assembled contigs or individual long reads aligned to a reference either with a synteny-style view or a dotplot view, each clearly showing inversions, deletions, and insertions. Whole-genome overviews of coverage (or other quantitative data) reveal copy number variation and large-scale duplications.

JBrowse 2 can be run on a web page, as a desktop application, inside Jupyter notebooks, from R sessions, or via the command line. A plugin framework enables developers to create new data adapters, track types, and view types. We have also integrated JBrowse 2 with R and Python ecosystems with JBrowseR, installable from CRAN, and dash_jbrowse, a widget that can embed JBrowse 2 in Jupyter notebooks.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3222. Lifebit delivers a federated Trusted Research Environment to allow researchers seamless access to 35,000 whole genomes in a novel COVID-19 cohort.

Authors:

N. Raine¹, M. Chatzou Dunford¹, P. Prieto Barja¹, C. Bacchelli¹, B. Goncalves¹, D. Silva¹, C. Chatzipantsiou¹, I. Peral Alvarez¹, D. Ardley¹, C. Davison¹, I. Knarston¹, T. Seeger¹, P. Moss²; ¹Lifebit Biotech Ltd, London, United Kingdom, ²Genomics England, London, United Kingdom

Abstract Body:

The COVID-19 pandemic has caused immeasurable global harm. However, the pandemic has also seen unprecedented collaboration and data sharing efforts worldwide. Genomics has played a central role in the COVID-19 response. Not least the fact that millions of whole genomes of SARS-CoV-2 have been uploaded to the international Global Initiative on Sharing Avian Influenza Data (GISAID) database to be shared globally. This near-real-time sequencing was the first time this technology was applied to rapidly inform public-health decisions and showed the world the power of genomics and global collaboration. However, COVID-19 remains a prominent public health issue with considerable impact beyond deaths and hospitalisations. One meta-analysis revealed that 56.9% of people infected with SARS-CoV-2 had at least one long-covid symptom at a median follow-up of 6-months. Genomics research could prove an invaluable resource to continue to track the evolution of SARS-CoV-2 and provide insights into patient response to infection, elucidating critical mechanisms of disease severity, potentially resulting in novel treatment options. To tackle these challenges, researchers require patients' linked genomic and health data. This data is being generated worldwide in multiple clinical trials; however, due to the sensitive nature of the data, it becomes siloed within hospitals and health services and proves difficult to access for research. To address this challenge and answer these key research questions, Genomics England (GEL), in partnership with the GenOMICC consortium, accomplished whole genome sequencing of 35,000 COVID-19 patients, 20,000 intensive care patients collected from 230+ hospitals across the UK and 15,000 people with mild symptoms. GEL selected Lifebit CloudOS as its new secure research platform or 'Trusted Research Environment' to enable rapid, secure access and analysis to this cohort. With a fully federated architecture, CloudOS has allowed GEL to securely retain this sensitive patient dataset within their own cloud environment, while CloudOS is deployed as a virtual abstraction layer. Federation allows approved researchers anywhere in the world to seamlessly access this large dataset of real-world data and perform analysis within the CloudOS platform. Analysis of this dataset has the potential to understand COVID-19 better, leading to discoveries of genetic factors relating to disease severity. Lifebit's federated CloudOS is a validated high-performance solution for population-level health data management. It addresses the problem of data sharing and collaboration with these datasets, which contains highly valuable but sensitive data.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3223. MANE Select and beyond: expanding the joint NCBI and EMBL-EBI transcript set

Authors:

J. Loveland¹, S. Pujar², A. Astashyn², R. Bennett¹, A. Berry¹, E. Cox², C. Davidson¹, O. Ermolaeva², C. Farrell², R. Fatima¹, T. Goldfarb², J. Gonzalez¹, D. Haddad², M. Hardy¹, T. Hunt¹, J. Jackson², V. Joarda², M. Kay¹, V. Kodali², K. McGarvey², J. Mudge¹, M. Murphy², S. Rangwala², F. Thibaud-Nissen², A. Vatsan², C. Wallin², D. Webb², A. Frankish¹, F. Cunningham¹, T. Murphy²; ¹EMBL-EBI, Hinxton, United Kingdom, ²NCBI/NLM/NIH, Bethesda, MD

Abstract Body:

Comprehensive gene annotation is essential for understanding the impact of clinically relevant variants. Historically there has not been a standard for clinical reporting which complicates the process of consistent interpretation and reporting. To address this, the Matched Annotation from NCBI and EMBL-EBI (MANE) collaboration between Ensembl/GENCODE and RefSeq has defined a high-value set of transcripts and corresponding proteins for use as a universal standard for variant reporting (Nature 2022;604(7905):310-315). Each MANE transcript represents an exact match between the exonic sequence of an Ensembl/GENCODE transcript and its counterpart in RefSeq, such that the identifiers can be used synonymously. The MANE Select set identifies a representative transcript for each human protein-coding gene with the MANE Plus Clinical set providing additional transcripts at loci where the Select alone is not sufficient to report all currently known clinical variants. MANE release 1.0 has MANE Select transcripts for 99.7% of human protein-coding genes, including all ACMG SF v3.03 genes. MANE transcripts are accessible from major genome browsers and key resources. Widespread adoption will increase consistency of reporting, facilitate exchange of data regardless of annotation source, and help streamline clinical interpretation. The MANE collaboration will continue, aiming to 1) reach 100% coverage of protein-coding genes on the human reference genome GRCh38, including those on patches and alternative loci; 2) capture novel pathogenic clinical variation in MANE Plus Clinical transcripts; 3) expand to include additional functionally important transcripts with exonic features absent from MANE Select and MANE Plus Clinical; and 4) incorporate a small set of transcripts from clinically relevant non-coding genes. MANE annotation is anticipated to be very stable, with proposed changes announced at <https://ftp.ncbi.nlm.nih.gov/refseq/MANE/> in advance of public release. We welcome input from the community to ensure that all aspects of the MANE collaboration provide maximum benefit to users. Please contact us at mane-help@ebi.ac.uk or MANE-help@ncbi.nlm.nih.gov. This work is supported by: Wellcome Trust-WT200990/Z/16/Z; EMBL-Core-Funds; NIH-U41HG007234. This work was supported in part by the National Center for Biotechnology Information of the National Library of Medicine (NLM), National Institutes of Health (NIH).

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday
PB3224. Metadata retrieval from sequence databases with *ffq*

Authors:

A. Galvez Merchan, A. Boeshaghi, L. Pachter; Caltech, Pasadena, CA

Abstract Body:

We present a command-line tool, called *ffq*, for querying user-generated data and metadata from sequence databases. Given an accession or a paper's DOI, *ffq* fetches metadata and links to raw data in JSON format at a rate of 10s per sample. *ffq*'s modularity and simplicity makes it extensible to any genomic database exposing its data for programmatic access. This enables the re-analysis of existing published data from numerous existing and future databases alongside novel experiments. *ffq* is open source, and the code can be found here: <https://github.com/pachterlab/ffq>

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3225. MetaMatchMaker: An artificial Intelligence tool for data discovery and meta-data harmonization

Authors:

G. Page¹, H. Pan², A. Blatecky³, K. Burdekin³, R. Chew³, I. Glaze³, C. Hamilton³, R. Henao⁴, C. Ives³, M. Mandal⁵, T. Reeve³, D. Williams⁶, X. Wu³, E. Earley³; ¹RTI Intl., Atlanta, GA, ²RTI Intl., RTP, NC, ³RTI Intl., Durham, NC, ⁴InfinML, Durham, NC, ⁵RTI Intl., Research Triangle Park, NC, ⁶RTI Intl., Research Triangle, NC

Abstract Body:

High quality data is a fundamental need in genomics, data sciences and the development of Artificial intelligence (AI) systems. The quality, nature, experimental design, and quantity of data determines the utility, performance, fairness, robustness, and scalability of the genetics analysis and AI models. However, data cleaning is often the least incentivized aspect of studies, viewed as ‘operational’ relative to the higher profile work of building novel models and algorithms. According to a 2018 Kaggle survey of data scientists, 50-65% of the time of a data study is devoted to data work including finding, gathering, harmonizing, and cleaning data. This is in part due to the highly manual nature of these steps, but this does not need to be the case. We have developed **MetaMatchMaker (M3)**, a suite of cutting-edge AI approaches and tools to reduce the time and costs of finding and integrating genomic, molecular, and clinical data for analysis. At its core, M3 is an advanced pretrained learning model (PLM) created using a natural language processing (NLP) neural network we call M3 neural network (M3:NN). With this approach we take advantage of the recent advancements in NLP entity matching to predict similarities between complex human language terms. **M3:NN** has been trained on data generated by data harmonizers with decades of experience abstracting and mapping terms between studies. Training data was drawn from ~30k unique, manually mapped study concepts between the PHENX toolkit and study terms within dbGAP (June, 2018). Subsequent model training included terms from BioLINCC and FITBIR. M3:NN supports two common use cases - data discovery and data linking. **M3:Find** is a data discovery tool (available at www.MetaMatchMaker.com) which lets users find publicly available data by matching user-supplied terms to study variables across numerous public databases and tools including dbGaP, BioLINCC, PhenX, and FITBIR. The M3:NN model predicts content similarity between terms based on the training data. **M3:Link** allows users to link meta data study terms between two or more studies, allowing for more rapid data integration across cohorts. This tool is also useful for building common data elements across multiple studies. M3:Link has been evaluated using dbGAP data submitted after July 1, 2018 to link 70,945 dbGAP variables from to the 9,148 PhenX variables at a score > 0.99 links to a correct variable 79% of the time and incorrectly 21% of the time. Use of MetaMatchMaker can reduce the time and budget for harmonizing complex datasets by simplifying the process of linking variables and meta data.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3226. MGeND: Integrated database of clinical and genomic information to encourage precision medicine in Japan

Authors:

M. Kamada¹, Y. Kawai², K. Tokunaga³, Y. Okuno¹; ¹Kyoto Univ., Kyoto, Japan, ²Natl. Ctr. for Global Hlth.and Med., Shinjuku-ku, Japan, ³Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan

Abstract Body:

In order to improve the accuracy of genomic medicine, it is necessary to accumulate genomic data with clinical interpretations that consider the diversity of genomes. To this end, we have been developing and operating the Medical Genomics Japan Database (MGeND), an integrated database of genomic and clinical information discovered in the East Asian population. We have been updating the database contents upon continuous submissions of data. The database previously provided variant information only for the genome assemblies specified by the registrant. We newly produced liftover of all variants between GRCh37 and GRCh38, and the mapped data were published on MGeND for download. In addition, we have improved the response time for downloads that took a long waiting time when the combination of items to be downloaded are specified. Currently, to resolve inconsistency of disease names, mapping between registered disease names in MGeND with the MONDO and NANDO disease ontologies is underway, where NANDO targets intractable disease Nanbyo in Japanese. Since MGeND deals with multiple public databases, it requires flexibility in the updating procedure. Therefore, we are refactoring the system to allow partial updates of corresponding information by introducing the Semantic Web framework. This enhancement enabled us to maintain data in the Resource Description Framework (RDF) so that we are able to represent complex data in a standardized data model. We have also introduced the TogoStanza framework that allows database developers to embed the information in our database as visualization modules in various applications, including other Japanese variant databases such as TogoVar. This presentation will introduce the system update and our efforts to develop a new system.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3227. National Center Biobank Network (NCBN) and genomic information for NCBN bioresources.

Authors:

K. Tokunaga¹, Y. Omae¹, Y. Kawai², E. Noiri¹, K. Kitajima¹, S. Hideyuki¹, Y-i. Goto¹; ¹Natl. Ctr. Biobank Network, Tokyo, Japan, ²Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan

Abstract Body:

The six national centers (6NCs) for advanced and specialized medicine in Japan conduct basic and clinical research on major diseases that have a significant impact on national health. Disease-specific bioresources and information collected by each NC are stored in a separate biobank. The National Center Biobank Network (NCBN) coordinates the biobanks and researchers of the 6NCs via an open access database (Catalogue Database: http://www2.ncbiobank.org/Index_en) for efficient provision of registered biological resources and data for use in research communities. The NCBN resources are characterized by their high quality and rich medical information and are available for life science research and for the development of novel testing methodologies (biomarkers), new treatments, and drugs for future healthcare in the scope of personalized medicine, through a deeper understanding of disease pathogenesis. In addition to the bioresource and medical information, whole genome sequencing analysis for 9,850 NCBN samples has been conducted since 2020. These can be utilized as control subjects in research into cancer and rare diseases in the Japanese population and their variant frequency information will be registered to NBDC Human Database and Medical Genomics Japan Variant Database (MGeND) after publication. The total number of obtained genome-wide data for NCBN bioresources is increasing for many diseases through whole genome sequencing analysis or SNP array analysis and it is now possible to search the Catalogue database for bioresources with genomic information. NCBN is creating a system that can provide both bioresources and genomic data to users.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3228. NCBI dbGaP FHIR API Provides Access to Thousands of Studies to Improve Data Integration and Interrogability for Biomedical Research

Authors:

L. Phan¹, E. Moyer¹, J. Bloom¹, L. Hao¹, M. Kimura¹, M. Feolo¹, P. Lynch², H. Xie², Y. Sedinkin², L. Amos², C. McDonald², L. Ziyabari¹, V. Lyoshin¹, K. Kaur¹, R. Russette¹, B. Kattman¹; ¹NIH, NLM/NCBI, Bethesda, MD, ²NIH, NLM/Lister Hill Natl. Ctr. for BioMed. Communications, Bethesda, MD

Abstract Body:

NCBI has developed an FHIR data-model-based API for fast, selective, secure, and reliable access to patient-level research study-related controlled data. The initial delivery of this API can pull data from dbGaP.

The Database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the results of studies that have investigated the interaction of genotype and phenotype in humans. It contains thousands of studies, >350K phenotypic variables, and ~2.5 billion observations from millions of subjects. Once a dbGaP study is processed and approved for release, it is made available through dbGaP FHIR API (GFA). The FHIR format can accelerate the use of clinical data for research. FHIR is a standardized way of transmitting health data from one health information system to another through an application programming interface (API). It is also compatible with analytic resources used in biomedical research, such as R and Python. The format is being widely promoted and adopted for use in clinical care. GFA provides two levels of access:

Open - provides study and variable metadata and summary (available to anyone with no restrictions). Users can search using multiple criteria including studies title, sponsor (IC), type (prospective, longitudinal, cohort, case-control), keyword, condition, and many others.

Controlled - provides individual-level genotype and phenotype data that have been de-identified (i.e., no personal identifiers, such as name, etc.) The new GFA can be used to download phenotype data from several large population cohort studies in dbGaP.

We will demonstrate GFA to assist researchers access diverse and large datasets. Users could then develop third-party applications to extract clinical data from the API to enable new discoveries and improved health.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3229*. NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS): 2022 Update.

Authors:

H. Issen¹, A. Kuzma¹, O. Valladares¹, E. Greenfest-Allen¹, P. Gangadharan¹, Z. Katanic¹, A. Wilk¹, Y. Zhao¹, L. Qu¹, M. Kirsch¹, M. Moon¹, A. Lerro Rose¹, J. Manuel¹, L. Armus¹, N. Saravanan¹, P. Keskinen¹, J. Cifello¹, O. Pathak¹, S-Y. Chou², W-P. Lee¹, Y. Leung¹, A. Naj¹, C. Stoeckert Jr.¹, G. Schellenberg¹, L-S. Wang¹; ¹Univ. of Pennsylvania, Philadelphia, PA, ²Lehigh Univ., Bethlehem, PA

Abstract Body:

Background: NIAGADS is a national genomics data repository that facilitates access of genotypic and sequencing data to qualified investigators for the study of the genetics of Alzheimer's disease (AD) and related neurological diseases. Collaborations with large consortia and centers such as the Alzheimer's Disease Genetics Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Alzheimer's Disease Sequencing Project (ADSP), and the Genome Center for Alzheimer's Disease (GCAD) allow NIAGADS to lead the effort in managing large AD datasets that can be easily accessed and fully utilized by the research community.

Methods: NIAGADS is supported by National Institute on Aging (NIA) under a cooperative agreement. All data derived from NIA funded AD genetics studies are expected to be deposited in NIAGADS or another NIA approved site. NIAGADS manages a Data Sharing Service (DSS) that facilitates the deposition and sharing of genomic data and association results with approved users in the neurodegenerative research community. In addition, researchers are able to freely use the NIAGADS Alzheimer's Genomics Database (www.niagads.org/genomics/) to search annotation resources that link published AD studies to AD-relevant sequence features and genome-wide annotations.

Results: As of June 2022, NIAGADS houses 88 datasets comprised of >144,000 samples including GWAS, sequencing, gene expression, annotations, deep phenotypes, and summary statistics. Qualified investigators can retrieve ADSP sequencing data with ease and flexibility through the NIAGADS DSS. To date, the ADSP and other contributing studies have completed whole exome sequencing (WES) of 20,503 samples and whole genome sequencing (WGS) of 16,905 samples. Raw WES and WGS files, quality controlled VCF files, and phenotype data files are available via qualified access. The next round of sequencing currently underway will generate around 18,000 additional genomes to be released in the second half of 2022.

Conclusion: NIAGADS is a rich resource for AD researchers, with the goal of facilitating advances in Alzheimer's genetics research. By housing datasets from many projects and institutions, NIAGADS enables AD researchers to meet their research goals more efficiently. Datasets, guidelines, and features are available on our website at <https://www.niagads.org>.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3230. North Carolina Macular Dystrophy (NCMD/MCDR1): Analysis of our entire database, a model disease of non-coding mutations.

Authors:

K. Small¹, M. Leys², P. S. Bernstein³, A. Scott⁴, E. M. Stone⁵, F. S. Shaya⁶; ¹Molecular Insight Res. Fndn., GLENDALE, CA, ²West Virginia Univ. Eye Inst., Morgantown, WV, ³Univ. of Utah Hlth. Moran Eye Ctr., Salt Lake City, UT, ⁴John Hopkins Wilmar Eye Inst., Belair, MD, ⁵Univ Iowa Hosps & Clinics, Iowa City, IA, ⁶Macula and Retina Inst., Glendale, CA

Abstract Body:

Purpose: We published the initial findings of 5 mutations in 12 NCMD families with 141 subjects. The purpose of this study is to clinically and molecularly study our entire NCMD database to determine if our initial findings continue to be substantiated. **Methods:** Ophthalmic examinations and whole genome sequencing (WGS) and/or targeted DNA Sanger sequencing was performed on our entire dataset of 53 families with 318 subjects. Junction PCR and Sanger sequencing was used to confirm point mutations and characterize duplications involving the MCDR1/MCDR3/*PRDM13* locus. **Results:** Of the total 318 subjects evaluated to date, 185 were found to be affected having DNA sequence changes consistent with MCDR1 on chromosome 6 or MCDR3 on chromosome 5. In addition to our 12 initial families, we report the findings of an additional 41 families with 75 subjects affected and 32 unaffected. Eight of these new families, 35 subjects, were found to have the original “V1” Chr6:99593030G>T mutation, in a non-coding region of the DNASE1 site upstream of *PRDM13*. Eleven families, 36 subjects, had the “V2” mutation Chr6:99593111G>C in the same DNASE1 site. One Asian family with 2 subjects had our previously reported Asian “V3” Chr6:99593164C>T mutation in the same DNASE1 site. Two new single nucleotide variants (SNVs) have recently been reported by us from our dataset, Chr6:99599064A>G in four members of one Czech family and Chr6:99959303G>C in four members of a Mexican family. A new tandem duplication Chr6:99560265-99616492 involving the same DNASE1 site, was recently reported by us in a Turkish family. **Conclusions:** North Carolina Macular Dystrophy (NCMD) is more prevalent than typically thought with a worldwide distribution making the name of this disease a gross misnomer. Continued identification of subjects and families and their mutations supports our initial discovery of mutations. A total of 8 of 13 worldwide NCMD mutations have been found by our group. All of the mutations (SNVs and duplications) appear to involve DNASE1 sites in non-coding regions. The mutations on chromosome 6 continue to appear to alter the expression of the retinal transcription factor *PRDM13* while the chromosome 5 mutations may affect the *IRX1* expression as we originally reported.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3231. Participant-provided case-level data from GenomeConnect can impact variant classification

Authors:

A. Morgan¹, E. Riggs², J. Savatt³, D. Azzariti⁴, C. Martin¹, H. Rehm⁵; ¹Geisinger, Danville, PA, ²Geisinger Hlth.System, Huntsville, AL, ³Geisinger, Lewisburg, PA, ⁴Broad Inst., Cambridge, MA, ⁵Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

Variant classifications may evolve over time with the emergence of new evidence. GenomeConnect, the ClinGen patient registry, enables sharing of de-identified variants, phenotype, and inheritance with ClinVar; other information, such as co-occurring variants, are available within the registry. Such data can impact variant classification. Here, we report on an assessment of variants shared by participants and their potential to inform other laboratories' variant classifications. Of the 2285 unique variants submitted to ClinVar by GenomeConnect, 213 (9.3%) had conflicting interpretations. Of these, 53 (24.9%) had a likely pathogenic/pathogenic (LP/P) classification on the participant's report. We focused on participant variants with case-level data that may inform other laboratory's classifications. These included 1) variants in autosomal dominant (AD) or X-linked genes that were found to be de novo in the GenomeConnect participant (potentially adding PM6 or PS2 to laboratories' classifications) or 2) variants in autosomal recessive (AR) genes co-occurring with a second LP/P variant in trans (potentially adding PM3). Variants unlikely to contribute to the classification (e.g., weak evidence types) were removed. We identified 12 unique variants where participant-contributed data could help be used to inform classification by another ClinVar submitting laboratory with VUS classification in the database. These included 6 de novo variants in AD genes and 6 variants in AR genes with another LP/P variant in trans. ClinVar submissions for these 12 variants had conflicting classifications (LP/P and VUS) from multiple submitters. For the 6 variants in AD genes, no laboratory submitter cited de novo inheritance as part of their classification. Two of the 6 variants in an AR gene did not note co-occurrence of a second variant in trans in ClinVar; such data are available in GenomeConnect. GenomeConnect participants also provided phenotypic details beyond what was included in the ClinVar laboratory submissions for all variants. These data highlight participant-provided individual-level data can provide additional evidence to help resolve discrepancies in variant classification. As a next step, the submitting laboratories for the 12 variants in conflict will be contacted by GenomeConnect to explore whether our participant-provided data could change their classification and resolve the discrepancy in ClinVar. This initial small-scale pilot and future work using GenomeConnect participant data could assist in future work on data sharing to inform discrepancy resolution, which can allow patients to receive updated and more accurate genetic test results.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3232. Partly Cloudy: analysis of blood lipids across All of Us and UK Biobank demonstrates the promise and limitations of cross-biobank analyses in the cloud

Authors:

M. Selvaraj¹, N. Deflaux², H. Condon³, K. Mayo³, M. Basford³, M. Effingham⁴, D. Glazer⁵, P. Natarajan¹, A. Bick³; ¹Massachusetts Gen. Hosp., Boston, MA, ²Verily Life Sci., Seattle, WA, ³Vanderbilt Univ., Nashville, TN, ⁴UK Biobank, Stockport, United Kingdom, ⁵Google, Mountain View, CA

Abstract Body:

Background: The rapid growth of genomic data precipitates a new research paradigm where data is stored centrally in Trusted Research Environments (TREs) such as the All of Us Researcher Workbench (AoU RWB) and the UK Biobank Researcher Analysis Platform (UKB RAP). Here we characterize the advantages and drawbacks of TREs in facilitating cross-analysis, as demonstrated by genome wide association study (GWAS) of blood lipids.

Methods: We conducted GWAS using lipid phenotypes (HDL-C; LDL-C, TC, TG) on UKB 200K whole exomes sequence (WES) and AoU 98K whole genome sequence (WGS) datasets (filtered to UKB exome capture region) in both siloed and pooled fashion on the UKB RAP and AoU RWB. In the siloed analysis, we carried out the GWAS on each cohort separately and meta-analyzed the results. In the pooled analysis, we combined datasets and carried out GWAS once for all samples. We compared the results from siloed and pooled analysis to each other and against a published multi-ancestry lipid study in the NHLBI TOPMed cohort.

Results: We curated lipid measurements for 37,754 AoU participants with WGS data and 190,982 UKB participants with WES data. Single variant GWAS was carried out with ~1.6M and ~2.1M variants from AoU and UKB cohorts respectively. After applying the AoU data dissemination policy which restricts dissemination of data on fewer than 20 participants (eg AC<40), we retained 30% and 23% of variants in AoU and UKB for further analysis. We meta-analyzed ~500K variants and identified 454 and 445 genome wide significant variants in siloed and pooled analysis, respectively. Overall there was strong correlation with known lipid loci ($R^2 \sim 92-97\%$) from both meta-analysis and pooled methods. Importantly, 84 variants were unique to siloed analysis and 75 variants were unique to pooled analysis. Examining these, we noted a significant increase in the proportion of variants predominantly from non-european ancestry individuals in the pooled analysis compared to siloed ($p=0.01$). We also show that pooled analysis required ~half as many computational steps as meta-analysis.

Conclusion: We demonstrate that results from siloed and pooled approaches to trans-biobank genetic analysis are similar but not identical. We found that pooled analysis is notably less complicated and has important implications for studying genetic variants in diverse individuals. These findings have important technical and policy implications for biobank TREs.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3233. Path to independence: Overview of challenges and opportunities of computational data-driven research in biology

Authors:

T. Ramesh¹, K. Chhugani¹, V. Jönsson², S. Mangul³; ¹Univ. of Southern California, Los Angeles, CA, ²Univ. of California, Santa Cruz, Santa Cruz, CA, ³Univ. of Southern California, Sch. of Pharmacy, Los Angeles, CA

Abstract Body:

Over the past decade, the rapid advancement of high-throughput technologies has reshaped modern biomedical research by vastly extending the diversity, richness, and availability of data and methods across various domains. Computational data driven, or dry lab, research focuses on developing and applying computational models and methods across various types of omics datasets. The researchers also perform secondary analysis- mine and reanalyze publicly available large scale open datasets to make novel biological discoveries. Here we have discussed the pitfalls, the opportunities and the significance of computational data driven research. With increasing importance to computational driven research in recent times, the paradigm shift in attitudes towards the computational research is also discussed -there is more receptiveness of the field by the biomedical community over the years, through increasing computational projects thriving in the leading consortium integration studies and publishing of the research in high caliber journals. Yet, there are fewer funding and powerful roles in the funding panels are assigned to computational researchers, suggesting the appreciation to the field is still not as significant. Computational research is pervasive and plays a multi-disciplinary role in almost every field and is not limited to biology, philosophy, sociology etc. In biology, the computational applications largely focus on genomics and translational bioinformatics which open doors to clinicians and wet lab experimentalists to come up with new diagnostics, therapeutics, or insights to the disease. Computational driven research is unique as it enables researchers to be creative and provide the freedom to discover novel diseases, invent new methods and tools or improve the existing ones to enhance data analysis. By establishing more meaningful collaborations between dry and wet lab research, the findings of computational data driven research becomes more interdisciplinary and crucial by validating the results from dry lab in wet lab. Transitioning into the future, many accomplished scientists believe that effective technical, research, and communication skills are key in thriving in the field of computational research. Computational research has a unique potential of being the bridge between the large amounts of data collected by wet labs and understanding this data. With effective collaborations with clinicians, and wet lab experimentalists, computational research is likely to flourish greatly and reach a remarkable level in the biomedical field in the future notably, in the fields of disease prevention, therapeutics and treatment decisions.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3234. PhenX presents Bone and Joint updates and expansion of Social Determinants of Health

Authors:

M. Krzyzanowski¹, M. Nelms¹, D. Williams¹, L. Cox¹, J. Schoden¹, C. Ives¹, A. Brandow², P. Carroll³, D. Felson⁴, K. Norris⁵, M. G. Cockburn⁶, W. Huggins¹, D. Maiese¹, L. Gridley¹, P. West¹, X. Wu¹, M. Mandal¹, H. Pan¹, N. Jones⁷, V. Marshall⁷, T. HENDERSHOT¹, C. Hamilton¹; ¹RTI Intl., Research Triangle Park, NC, ²Med. Coll. of Wisconsin, Milwaukee, WI, ³Johns Hopkins Univ., Baltimore, MD, ⁴Boston Univ. Sch. of Med., Boston, MA, ⁵Div. of Gen. Internal Med. and Hlth.Services Res., Univ. of California Los Angeles, Los Angeles, CA, ⁶Univ. of Southern California, Los Angeles, CA, ⁷Natl. Inst. of Minority Hlth.Disparities, Bethesda, MD

Abstract Body:

The PhenX (**Phenotypes and eXposures**) Toolkit (<https://www.phenxtoolkit.org>) is a web-based catalog of recommended measurement protocols and associated bioinformatics tools to assist with study design and facilitate cross-study data integration and analyses. PhenX protocols are recommended by domain experts using a consensus-based process that includes community input. The PhenX Toolkit currently includes measurement protocols from 29 research domains (e.g., Demographics and Diabetes) and six collections that add depth to the Toolkit (e.g., Social Determinants of Health) and has over 4,200 registered users. An established resource that continues to expand content for investigators, the PhenX Toolkit includes expansion of the Social Determinants of Health (SDoH) Research collection, updates to the Bone and Joint domain, and a new Sickle Cell Disease (SCD) Pain specialty collection. As part of the National Institute of Minority Health Disparities research framework and vision, the PhenX Toolkit received funding to expand the existing SDoH Research collection. New protocols were recommended by SDoH-X Working Group members, focused on topics including water access, housing insecurity, accommodations for mental or physical disabilities and incarceration history. These new protocols will enhance the existing individual and structural SDoH protocols already in the Toolkit. Supported by the National Heart, Lung, and Blood Institute (NHLBI) through co-funding to the “PhenX Toolkit-Expansion and Sustainability” (U41) project Y3-Y5, the Bone and Joint Working Group recommended new protocols to assess bone turnover and mineral content, fracture, pain, gout, arthritis, and fibromyalgia. The SCD Pain Working Group, identified a set of 24 protocols that encompass the burden of pain on an individual’s day-to-day life, the magnitude of pain experienced by an individual, the location of the pain experienced, the extent to which the pain interferes with the person’s daily life, the individual’s overall physical mobility, the specific subjective sensations associated with pain, and the coping strategies employed by the individual. We will present the final protocols recommended by the Working Groups with review and approval by the Sickle Cell Disease Research and Scientific Panel and the PhenX Steering Committee.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3236. PreSiBO: Database system for genome guided and data driven drug repurposing.

Authors:

D. Priyadarshi, N. Sahelijo, G. Jun; Boston Univ., Boston, MA

Abstract Body:

Precision medicine aims to develop patient driven treatment solutions using genetic profiles and corresponding drug options. Since clinical manifestation of Alzheimer's disease (AD) is complex in both neuropathology and underlying disease mechanisms, there is an urgent need to develop a tool enabling precision medicine in AD. We, therefore, developed an integrated database system after profiling predictor, signature, biomarker, and outcome (PreSiBO) features of genome-guided patient subgroups.

Methodology: To aid precision medicine in AD, we collected, collated and curated omics, protein interaction, and existing drugs data and connected these by leveraging their inherent hierarchical structure. The data included:

- 1) Predictors - genome-wide association study (GWAS) summary statistics, network based polygenic risk scores,
- 2) Signatures - transcriptome, methylome, lipidome, and proteome profiles,
- 3) Biomarkers - cerebrospinal fluid and MRI, PET imaging,
- 4) Outcomes - Clinical/neuropathological.

We created *PreSiBO subgroups* for precision medicine that profile subjects using genetic risk scores for biologically connected gene networks derived from the curated omics data. Gene-to-drug and protein-protein interaction data were incorporated into the database to identify existing FDA approved drugs within a gene-network for a targeted PreSiBO subgroup.

Implementation: PreSiBO was implemented using a MySQL based relational database system deployed on Amazon Web Services (AWS) for quick retrieval of these logically linked datasets. A custom web interface dashboard was also developed using R Shiny for easy search and navigation of the datasets in the database. We will, further, scale the system and implement machine learning algorithms to detect priority targets that have best matched approved drugs for drug repurposing, in each PreSiBO subgroup.

Conclusion: In summary, PreSiBO system has allowed us to better organize the diverse but logically connected precision medicine data to holistically understand genetic profiles for data driven drug repurposing in AD.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3237. Rare Disease Phenotyping: What is Next?

Authors:

D. Adams¹, C. Esteves², E. Burke³, M. Malicdan¹, W. Gahl⁴, A. McCray², Undiagnosed Diseases Network; ¹NIH, Bethesda, MD, ²Harvard Med. Sch., Boston, MA, ³NIH - NHGRI, Bethesda, MD, ⁴NHGRI (NIH), Bethesda, MD

Abstract Body:

Clinical phenotype has long been a critical component of both clinical care and research for patients with rare inherited diseases. The era of genomic sequencing, and particularly consequent reverse phenotyping, have increased the pace of evolution in disease definition. For newly described and ultra-rare disorders, phenotype can be a critical component for establishing coherent cohorts. The launch of the Human Phenotype Ontology (HPO) in 2007, and subsequently Matchmaker Exchange (ME) and GeneMatcher (GM) in 2013, heralded fundamental changes in how rare phenotypes are documented and shared. Each of these tools continues to be used broadly in the rare disease research community. Azzariti and Hamosh reported in 2020 that MME encompassed “150,000 cases from more than 11,000 contributors in 88 countries”. In the Undiagnosed Diseases Network, 1775 cases have been posted to the PhenomeCentral (PC) ME node, generating 746 matching inquiries. Importantly, most matches are triggered by gene name rather than phenotypic similarity. HPO continues to be updated with new terms, with the last posted release occurring in April of 2022. However, coding of some phenotype classes remains challenging, e.g., neurodevelopmental, neuroradiological and laboratory measurement. Computational tools that incorporate HPO matching typically use terms without metadata. Therefore, important information about severity, periodicity, age of onset and other factors are not utilized. Despite a significant published literature addressing the generation of HPO terms from text, robust open-source tools in active development are lacking. The challenges of obtaining funding to develop and maintain tools for documenting HPO terms has driven most projects in to an expensive commercial realm that comprises a barrier to use for smaller and less-resourced projects.

We present a summary of our experience with phenotype recording and subsequent utilization. We then propose a list of phenotype-related goals for the rare disease community, funders and users alike, to consider for the future development of this critical rare disease research infrastructure.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3238. Regeneron Genetics Center (RGC) Cohort Identification Application

Authors:

D. Sharma¹, M. Cantor¹, S. Malhotra¹, S. Gokhale¹, G. Mitra¹, J. Staples², M. Nafde¹; ¹Regeneron Pharmaceuticals, Tarrytown, NY, ²Regeneron Genetic Ctr., Tarrytown, NY

Abstract Body:

Integrating high-throughput exome sequencing and diverse phenotypic data is at the core of therapeutic target research and identification. The Regeneron Genetics Center has developed an interactive application that allows scientists to efficiently explore and visualize the phenotypic data for generating clinical cohorts for Genome-Wide Association (GWAS) analysis. The application represents massive clinical phenotypic datasets from various collaborations worldwide, representing over a million participants. The data is derived from multiple sources, such as electronic health records, registries, and patient surveys which makes this application unique since most platforms (e.g., i2b2) focus only on data from a single source or source type. The data is standardized and harmonized through mapping to standardized vocabulary concepts within RGC's internally developed Master Ontology, permitting the uniform representation of clinical concepts across data sources. The phenotype browser application makes it highly efficient to query the structured data via vocabulary concepts, thus, allowing for quick and flexible self-service selection of patients that share common clinical characteristics across multiple studies and expanding the scope of target discovery. The application has been deployed widely within RGC for clinical data exploration and supports drag and drop clinical cohort queries, cohort visualization and comparison, and interoperability with other enterprise applications through interaction with the backend relational databases through a browser-based user interface. For example, for the use case for generating a composite Breast cancer phenotype, a user can search for neoplasm of breast ICD-10 CM concept (C50 and its children concepts) within internal Master Ontology and find relevant data across all RGC's collaborations. Similarly, to locate Hemoglobin A1c (HbA1c) data, a user can search for HbA1c LOINC parts concept code (LP16413-4). Using the drag and drop functionality, all patients that are cases for breast cancer and have HbA1c > 7.5% can be added to the query for generating a composite phenotype. The application also allows for a graphical view of the query results by age, gender, race, and ethnicity. Users can in real-time edit the query, and get the updated results. The application's short-term plans include extending the cohort builder functionality to include longitudinal phenotype data and integrating deep learning models to identify more patients that meet the cohort inclusion criteria.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3239. Role of Python APIs, Big data and Cloud computing in understanding the Genetic basis of a Disease

Authors:

S. Murthy; Self Employed/ Freelancing, Hyderabad, India

Abstract Body:

Background: Datasets on signaling events appear central to investigators in understanding the genetic basis of a disease. Recently, there has been much focus on the use of emerging information technology (IT) tools in this regard. Say, for example, API (application programming interface) is one such area. APIs connect to multiple EHRs (electronic health records), labs, pharmacies, and patient portals and finally provide commercial firms access to general health data. Here, database development and storage of information using cloud technology is an important subarea to emphasize. However, there is poor literature with regard to the application of this IT-based platform particularly using python (a programming language) APIs or python based web APIs as useful tools in understanding the cross-talk between signals and genes. **Aim:** The aim of this project is to carry out a literature review on how python API and cloud computing are efficient in exploring and synthesizing the datasets/big data in understanding the genetic- disease specific signalling events. **Methods:** A web-based literature search was carried out. Electronic databases like PubMed were accessed between January 1, 2012, and May 31, 2022. The keywords used were ‘role of APIs in medicine’, ‘role of APIs in genetics’, ‘signals in host-pathogen interaction’, ‘APIs in health and genetics’, ‘the significance of APIs in healthcare industry’, ‘python in genetic disease management’, ‘big data in genetic disease management’, ‘big data in genetic disease management, and ‘cloud computing in genetic disease management, respectively, to recognize studies that apply the role of python in clinical settings in this area. **Results:** The database search result revealed only a limited number (below ten) of articles. **Conclusion:** The gaps in data may imply that an IT-driven understanding of signal machinery requires attention. Molecular signals may possess a strong link with the host’s unique genetic make-up and can be instrumental in influencing a particular disease cycle. So, datasets linking a disease, distinct signals, and the gene need to be created and stored in huge cloud servers. Such strategy when integrated with python APIs (API integration) may not only uncover undiscovered correlations, hidden patterns, and other insights through inspecting large-scale multiple sets of data but also facilitate systems to exchange the sources of data. This makes the data more accessible and understanding with further exploratory insights on database schemas for the healthcare community.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3240. STR Truth Set for Evaluating Genome-wide STR Calls

Authors:

B. Weisburd, N. Kurtas, V. Jalili, H. Brand, G. Tiao, H. L. Rehm; Broad Inst. of MIT and Harvard/Massachusetts Gen. Hosp., Cambridge, MA

Abstract Body:

Short tandem repeats (STRs) are scattered throughout the human genome. Recent advances in tools such as GangSTR and ExpansionHunter have made it possible to survey the full range of STR variation using whole genome sequencing data (WGS). However, the lack of a high-confidence genome-wide STR truth set makes it difficult to determine 1) which STR caller(s) to use for genome-wide analyses; 2) what set of STR loci to genotype; and 3) the accuracy of results. Current approaches to addressing these issues involve simulating STR expansions, Mendelian-violations analyses, and/or using standard long-read sequencing data as truth. However, simulated data doesn't sufficiently capture the complexities of WGS, Mendelian-violations analyses can inform about specificity but not sensitivity, and long read variant calls or diploid assemblies are anecdotally inaccurate for STR loci as described in [Dashnow 2021] and [Chiu 2021].

Here, we report a genome-wide STR truth set based on the Synthetic Diploid Benchmark [Li 2018]. It contains 178,909 STR variants found in the CHM1-CHM13 sample. To derive this truth set, we filtered the 4.1 million variants in the original synthetic diploid benchmark to insertions and deletions whose variant sequence + flanking reference sequence consist of at least 3 repeats of some shorter motif (excluding homopolymers). For example, if the original truth set contains a variant that denotes a deletion of CAGCAG from a reference context of 5 CAGs, we add it to the STR truth set as a contraction of 2 CAG repeats. We find this to be a good source of truth because it is based on haploid long-read assemblies which are more accurate than diploid assemblies. Additionally, since CHM13 is the basis of the new telomere-to-telomere (T2T) reference genome, we can further validate STR expansions by lifting them over to T2T, checking that at least one allele matches the T2T reference, and then lifting them back to GRCh38 to see if they map to their original GRCh38 locations. The fact that 98.3% of identified STR variants pass this QC process, and that we find nearly equal numbers of STR expansions (89,460) and contractions (89,449) highlights the quality of this truth set.

We then evaluate the accuracy of both short- and long-read STR callers on the 99.6% of truth set STR loci where the motif inside the variant is also found in the surrounding reference context. We additionally evaluate 178,246 other STR loci from the reference genome as true negatives. Among other findings, we see that ExpansionHunter and GangSTR have similar accuracy of ~70% when we define accuracy as $(TP+TN)/(\text{total \# of loci})$.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3241. Study design using the PhenX COVID-19 Collection and Variable Compare Tool

Authors:

D. Williams, M. Krzyzanowski, L. Cox, M. L. Engle, I. Glaze, W. Huggins, C. Ives, D. Maiese, T. Reeve, T. HENDERSHOT, C. Hamilton; RTI Intl., Research Triangle Park, NC

Abstract Body:

The PhenX (**Phenotypes and eXposures**) Toolkit (<https://www.phenxtoolkit.org>) is a web-based catalog of recommended measurement protocols and associated bioinformatics tools to assist with study design and facilitate cross-study data integration and analyses. PhenX protocols are recommended by domain experts using a consensus-based process which includes research community input. The PhenX Toolkit currently includes measurement protocols from 29 research domains (e.g., Demographics and Diabetes) and six collections that add depth to the Toolkit (e.g., Social Determinants of Health) and has measures included in 523 National Institutes of Health Funding Opportunity Announcements. The Toolkit released a COVID-19 Research Collection and is continuing to build tools for COVID-19 research. The COVID-19 Research Collection has six specialty collections to address research areas of concern to the community: Behaviors and Risks; Ethnicity, Race and Demographics; History, Treatment and Outcomes; Information Resources; Psychosocial and Mental Health; and Socioeconomic. Investigators are encouraged to use protocols from these collections when planning or designing a new study. Because health disparities were identified as an area of particular concern, the Social Determinants of Health Core collection is recommended to complement the COVID collections. Expansion of the existing COVID collection to include Long COVID protocols is in progress. The PhenX COVID-19 Variable Compare Tool (VCT) was developed to allow researchers to explore the protocols and variables in the COVID-19 Research Collection and the full instruments included in the PhenX COVID-19 Library. VCT features include a Keyword Search at the variable level, Side-by-Side comparisons of similar variables from two questionnaires, and visual representations of 3+ questionnaires via a heatmap. The Keyword Search allows researchers to address specific research topics by finding individual response items from questionnaires. For example, an investigator could search for “depression” to see which variables and response items relate to participants’ experience with depression as related to COVID-19. The Side-by-Side Comparison feature presents the degree of similarity between two questionnaires at the variable level. The heatmap visual representation gives a large-scale comparison between questionnaires and their number of similar variables. These tools can be employed to identify protocols that can be more easily harmonized with data from published COVID-19 research using the same questionnaires.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3242. The Alzheimer's Disease Sequencing Project - Follow Up Study ADSP-FUS: APOE genotype status and demographic characteristics across datasets

Authors:

P. Mena¹, A. Zaman¹, K. faber², L. Adams³, J. Inciute¹, P. Whitehead¹, T. Foroud⁴, D. Reyes-Dumeyer⁵, A. Kuzma⁶, H. Issen⁷, A. Naj⁷, E. Martin⁸, C. Dalgard⁹, G. Schellenberg¹⁰, L-S. Wang¹¹, R. Mayeux¹², B. Vardarajan¹², J. Vance¹, M. Cuccaro¹³, B. Kunkle¹, M. Pericak-Vance⁸; ¹Univ. of Miami, Miami, FL, ²Indiana Univ., Indianapolis, IN, ³Univ Miami, Miami, FL, ⁴Indiana Univ Sch Med, Indianapolis, IN, ⁵Columbia Univ., New York City, NY, ⁶Univ Pennsylvania, Philadelphia, PA, ⁷Univ. of Pennsylvania, Philadelphia, PA, ⁸Univ. of Miami Miller Sch. of Med., Miami, FL, ⁹USUHS - TAGC, Bethesda, MD, ¹⁰Univ Pennsylvania Sch Med, Philadelphia, PA, ¹¹Univ. of Pennsylvania, Pennsylvania, PA, ¹²Columbia Univ, New York, NY, ¹³John P. Hussman Inst. for Human Genomics, Miami, FL

Abstract Body:

Background: The ADSP-FUS is a National Institute on Aging (NIA) initiative focused on identifying genetic risk and protective variants for Alzheimer Disease (AD) by expanding the ADSP beyond non-Hispanic Whites of European Ancestry (NHW-EA) populations. Given the lack of diversity in the ADSP, the ADSP-FUS was designed to whole genome sequence (WGS) existing ethnically diverse and unique cohorts. The upcoming phase for ADSP-FUS, ADSP- FUS 2.0: The Diverse Population Initiative, focuses on inclusion of Hispanic/Latino (HL), non-Hispanic Black with African Ancestry (NHB-AA), and Asian populations. **Methods:** ADSP-FUS cohorts consist of studies of AD, dementia, and age-related conditions. Clinical classifications are assigned based on standard criteria and derived from clinical measures and history, as well as additional neuropathologic data. In addition to production of WGS, genome-wide array and *APOE* genotyping is acquired or performed for all ADSP-FUS samples. **Results:** The ADSP-FUS currently consists of 38 cohorts comprised of ~40,000 individuals, with plans to sequence > 100,000 individuals from diverse ancestries. Genotyping, sequencing, and clinical adjudication has been performed on 19,305 participants (cases N=6,679, median age=75; controls N=11,155, median age=72; ADRD N=1,471, median age=72). Median age varies greatly within individual cohorts: oldest for cases is 81.8+7.9, youngest cases is 56.1+3.3; oldest for controls is 85.3+8.0, youngest is 66.5+6.7. More than 60% of participants are female and are evenly distributed across cases (62%), controls (65%), and ADRD (62%). As expected the most prevalent *APOE* genotype is *APOE* 3/3 (% by cases/controls for 2/2=0.2,0.7; 2/3=4.2, 9.8; 2/4=2.7, 2.8; 3/3=37.3, 53.5; 3/4=43.5, 29.7; 4/4=12.1, 3.5). These proportions vary greatly within ethnicities, with the highest for *APOE* 4/4 observed in Amerindian participants (22%) and the lowest in NHW-EA participants (7.5%). Median Braak stage for AD cases is higher (4.9+1.2) than controls (2.1+1.3) and ADRD participants (2.8+1.0). **Discussion:** The results provide an overview of features of ADSP-FUS cohorts. The continued growth of the ADSP-FUS is central to the ADSP. As the ADSP-FUS expands in size and diversity, this genomic resource, available via NIAGADS, will be integrated with ADSP programs focused on phenotype harmonization, association analyses, functional genomics, and machine learning. In concert with these programs, the ADSP-FUS will accelerate the identification and understanding of potential genetic risk and protective variants for AD across all populations with the target of developing new treatments that are globally effective.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3243. The BaseJumper™ research platform for single cell multiomic analysis and interactive visualization

Authors:

V. Weigman, V. Amin, I. Salas Gonzalez, T. Tate, C. Culler; BioSkryb Genomics, Durham, NC

Abstract Body:

Translating the molecular alterations, especially those from a single cell, can be a powerful tool to understand underlying mechanisms of disease and somatic mosaicism. Along with managing the vast data volume from sequencing technologies, the complexity of these changes requires a multi-dimensional approach to first compute and identify the variance found within the cells of a study and then display these effects concomitantly. The platform was designed for researchers across computational skill levels, placing more energy on the ability ask direct questions of across the multiple forms within their dataset. We present a secure, cloud-based bioinformatics and visualization platform that can be run from any laboratory computer on a standard web browser. Samples can be pulled from across repositories: local and cloud, to create analysis projects that can be wholly contained within an institution or shared across collaborators for distributed analysis considerations. Available pipelines provided within the platform enable simple (SNV/insertion/deletion) and complex (structural/copy number/highly polymorphic) variant detection in addition to cell identification, cell state and pathway status from gene expression. Pipelines have been benchmarked across single cells of NIST standards like NA12878 achieving a ~94% accuracy and ~99% precision for variant identification. Context for variation across cells is provided through annotation to several known repositories for genetic variation (gnomAD, ExAC, 1000 genomics) and clinical relevance (COSMIC, ClinVar). These can be directly queried for prevalence or in addition to other filtering strategies, where schema can be saved and toggled within and across research groups. This filtering across the whole genome dataset (hundreds of thousands to tens of millions of variants) occurs in just a few seconds. Leveraging cellular or sample phenotypes, different methods can be directly applied to datasets to run association studies, and important genotypes presented, without the need for pulling extra compute. Native to the platform is the ability to toggle these associative markers across transcript and genome-level of paired datasets. Output from one analysis can be used to guide visualization or narrow focus of another data type. Visualization applications can take datasets across genomic and transcriptomic results and can be manipulated based on expert interpretation of the researcher to maximize biological leverage. Compiling all of this within a browser maximizes accessibility of data so you can perform analyses anywhere inspiration takes you.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3244*. The cellxgene suite is an online analytical platform and the largest repository of standardized single-cell data.

Authors:

B. Aebermann¹, P. Garcia-Nieto², S. Chambers¹, T. Harley¹, B. Raymor¹, M. Lombardo¹, K. Liang¹, M. Czerwinski¹, S. Badajoz¹, C. Megill¹, T. Smith¹, M. Dunitz¹, D. Hegeman¹, A. Mani¹, T. Huang¹, K. Katsuya¹, E. Bezzi¹, B. Martin¹, A. Tolopko¹, A. Infeld¹, M. Urisko¹, J. Cool¹, A. Carr¹; ¹CZI, Redwood City, CA, ²Chan Zuckerberg Initiative, Redwood City, CA

Abstract Body:

Cellxgene (cellxgene.cziscience.com) is a free-to-use online data portal hosting a growing corpus of more than 350 single-cell datasets with over 24 million unique cells from human and mouse. The portal hosts single-cell data from modalities that include gene expression, chromatin accessibility, DNA methylation, and spatial transcriptomics. All data are standardized to include raw and normalized counts, and annotated using an ontological shared vocabulary for cell and gene metadata.

Data are easily searchable and can be downloaded in multiple formats via web or by programmatic API calls. Additionally we deploy UI-based analytical tools for exploration of single datasets that do not require download. We will showcase the main tools hosted in the portal. First, the cellxgene explorer which displays an interactive 2-dimensional representation of cells in a dataset and allows users to color cells by gene activity or metadata (e.g. cell type, disease, technical features, etc.), subset and analyze subgroups of cells, perform differential gene expression and create scatter plots of gene expression. Second, scExpression which allows querying the expression of any gene across all human and mouse cell types available in the concatenated data from the portal.

Cellxgene is intended for community use and contributions. By supporting multiple modalities and data generated by labs around the world, the cellxgene suite of tools and data aims to maximize rapid use of data. To date, we support data from over dozens of labs and consortia such as the Tabula projects, LungMap, BICCN, Allen Institute for Brain Science, KPMP and the Human Cell Atlas. New contributions are welcome, the cellxgene team actively supports curation of data, and we work to ensure that self-curation is easy.

We are continuously improving the usability of cellxgene and adding new features tailored to the needs of cell and computational biologists. Groups interested in submitting their own data can contact the cellxgene team at cellxgene@chanzuckerberg.com to explore whether your data would be a good fit for the cellxgene resource.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3245. The ClinGen Allele Registry, Linked Data Hub, and Evidence Repository Support Curation Efforts Through Registration of Variants, Aggregation of Supporting Evidence, and Publication of Pathogenicity Classifications

Authors:

K. Riehle¹, A. R. Jackson¹, N. Shah¹, A. Zuniga¹, T. Farris¹, A. Kodochygov¹, K. Yu¹, M. Wright², C. Preston², K. Dalton², M. Mandell², G. Cheung², D. Ritter¹, L. Babb³, K. Ferriter³, E. McCormick⁴, X. Gai⁵, L. Shen⁵, M. Falk⁴, M. Cline⁶, A. Rubin⁷, C-H. Wei⁸, Z. Lu⁸, O. Griffith⁹, M. Griffith⁹, T. Klein², S. Plon^{1,10}, A. Milosavljevic¹, Clinical Genome Resource; ¹Baylor Coll. of Med., Houston, TX, ²Stanford Univ. Sch. of Med., Stanford, CA, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Children's Hosp. of Philadelphia, Philadelphia, PA, ⁵Children's Hosp. Los Angeles, Los Angeles, CA, ⁶UC Santa Cruz, Santa Cruz, CA, ⁷Walter & Eliza Hall Inst. of Med. Res., Parkville, Australia, ⁸Natl. Library of Med., Bethesda, MD, ⁹Washington Univ. Sch. of Med., St. Louis, MO, ¹⁰Texas Children's Hosp., Houston, TX

Abstract Body:

To support ClinGen curation efforts and also provide services for research and clinical genome interpretation, we developed a suite of web accessible services, including the ClinGen Allele Registry (CAR) for variant registration and naming, the Linked Data Hub (LDH) for aggregation of supporting evidence associated with variant pathogenicity assessments, and the Evidence Repository (ERepo) for publishing variants and evidence curated through the Variant Curation Interface (VCI). The VCI accepts a registered CAR canonical allele identifier (CA ID) as input in which aggregated supporting information (per variant) is queried and procured from the LDH. Upon completion of variant classification “publication” in the VCI, the structured data is transmitted through the ClinGen Data Exchange to the ERepo in SEPIO format (GA4GH) and is accessible by the public. The CAR, LDH, and ERepo are open access and accessible via both User Interfaces (UIs) and programmatically via the Application Programming Interfaces (APIs). Both the UIs and APIs are accessible at clinicalgenome.org/tools and are described in more detail below. The ClinGen Allele Registry contains over 2.57B canonical alleles and 300k CNVs registered from individual investigators, public external databases, and contributors from research and clinical communities. In order to support further growth of the CAR database, we have released a major update to support 64-bit identifiers, thus allowing UI-based registration of individual and small batches of variants and API-based bulk registration at a speed of between 100 to 100,000 variants per second (tinyurl.com/car-guide). The ClinGen Linked Data Hub supports Open, Linked Data principles by leveraging infrastructure to store structured documents linked to one another, to aggregate variant annotations excerpted from external databases and literature. Additionally, the LDH tracks provenance information, which is vital for supporting current and previous curations that may require reassessment. We are actively soliciting content contributions from additional external databases (e.g., BRCA Exchange, CIViC, hypothes.is, MaveDB, MSeqDR, tmVar/LitVar and others) in the implementation and testing of a “self-service” approach that will empower data owners towards contributing fresh variant data to inform curation and for any other use by the research and clinical communities. The ClinGen Evidence Repository, the designated distribution point for FDA-recognized ClinGen pathogenicity classifications, currently contains curations for over 3,600 curated variants from 19 Variant Curation Expert Panels categorized by 45 separate conditions.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday
PB3246. The importance of assessment of gene-disease association in the clinical practice.

Authors:

P. Bauer, E. Zonic, M. Ferreira, N. Ordonez, D. Saravanakumar, L. Almeida, I. Fernandes, N. Gulati, R. Higgins, R. Al-Ali, C. Pereira, O. Paknia, U. Hladnik, J. Pinto Basto, A. Bertoli-Avella; Centogene GmbH, Rostock, Germany

Abstract Body:

Background/Objectives:

One of the major challenges for diagnostic laboratories is the evaluation of potentially relevant variants within genes that are not yet conclusively associated to a genetic disease but have been identified in patients. With this study, we aimed to explore the gene-disease association evidence for 56 different genes and the significance of a large bio/databank for the classification process.

Methods:

The ClinGen Clinical Validity Framework for evaluation of gene-disease associations was applied. All but one of the 56 assessed genes had no publicly available gene-disease assessment in the ClinGen database, and 38 genes had no registered OMIM® phenotype. CENTOGENE's bio/databank with exome/genome data from previously tested individuals, in addition to data available in the literature or public databases were used for gene classification. According to the framework, Definitive, Strong, Moderate, Limited, or No Known Disease Relationship were assigned.

Results:

A Strong level of evidence has been reached for 21.1% of the genes, Moderate for 28.1%, Limited for 49.1%, and No Known Disease Relationship for 1.7% of the genes. A higher total number of points has been reached for 85.7% of the genes when using our bio/databank compared with using only externally obtainable data, allowing to increase the final level of classification in 19.6% of the genes.

Conclusion:

Our results demonstrate the importance of careful assessment of gene clinical validity data, along with the use of genetic data repositories. Implementation of ClinGen standardized scoring system for assessment of gene-disease association is relatively easy to apply and relevant in a clinical diagnostic setting. Diagnostic laboratories should implement such systems and contribute to closing the knowledge gap in genetic diagnostics.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3247. The International Mouse Phenotyping Consortium: A unique resource underpinning the study of human disease

Authors:

T. Groza¹, V. Muñoz-Fuentes¹, R. Wilson¹, H. Haseli Mashhadi¹, F. Lopez Gomez¹, O. Gunes¹, P. Cacheiro², A. Frost³, B. Vardal³, A. McCoy³, T. Cheng³, P. Keskivali-Bond³, L. Santos⁴, S. Wells³, D. Smedley², A-M. Mallon⁴, H. Parkinson¹; ¹EMBL - European Bioinformatics Inst., Hinxton, United Kingdom, ²Queen Mary Univ. of London, London, United Kingdom, ³Mary Lyon Ctr. at MRC Harwell, Didcot, United Kingdom, ⁴The Turing Inst., London, United Kingdom

Abstract Body:

The International Mouse Phenotyping Consortium (IMPC) is a world-wide initiative started in 2011 that links genes to function using single-gene knockout mice to inform mammalian gene function and human disease. Currently, more than 9,000 lines have been produced, many of which represent poorly understood genes. The IMPC data is produced via a systematic phenotyping pipeline, spanning across organismal physiological systems and comprising more than 250 phenotypic parameters collected on embryos, young adult and aged mice.

IMPC is a one-of-a-kind effort supporting the identification of new mouse models of rare and common human diseases, new gene functions and the development of novel methodological approaches that form the basis of new gene-disease associations. Moreover, it underpins successful experiments effectively uncovering pleiotropy, of particular importance when elucidating the genetic causes of syndromic disorders, as well as wide-ranging sexual dimorphism.

The IMPC data exists in a complex ecosystem of projects and ever growing databases which are evolving to deliver systematic analyses of cellular, organism level and population analyses. Data are curated, integrated and analysed and are made available to the public through data releases and a Web portal at mousephenotype.org. The latest release (April 2022) consists of over 80M data points and over 93,000 statistically significant phenotype hits mapped to human diseases. Interoperability with human disease and phenotype resources is achieved via widely-adopted community ontologies, such as the Human Phenotype Ontology. This enables, for example, the integration of the IMPC data directly into the interpretation of clinical genomes for National Health Service patients in the UK and in US based sequencing programs such as the Undiagnosed Disease Programme and Network.

From a clinical perspective, to date, IMPC has identified over 1700 disease models that can be used for mechanistic or therapeutic studies, where clinical phenotypes associated with known disease-gene associations are recapitulated in the mouse. In addition, thousands of novel, potential disease-gene associations have been detected that can inform on the candidacy of variants detected in patient sequencing studies. In addition to the significant impact made so far in large-scale functional studies, as complex disease analyses move from association studies to causative or 'effector' gene analyses, the functional data provided by IMPC is critical to understand the phenotypic consequences of variants, to understand which genes are essential and which contribute to diseases enabling downstream functional follow up.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3248. The NINDS Human Genetics Resource Center: 20 Years of Impact

Authors:

X. Jian¹, L. Scheinfeldt¹, S. Heil¹, A. Green¹, A. Amberson¹, S. Sander-Effron¹, R. Zhang²; ¹Coriell Inst Med. Res., Camden, NJ, ²Natl. Inst. for Neurological Disorders and Stroke-NIH, Bethesda, MD

Abstract Body:

The National Institute of Neurological Disorders and Stroke (NINDS) Human Genetics Resource Center was established in 2002 by the NINDS as a public biorepository of DNA and lymphoblastoid cell lines to facilitate the identification of genetic risk factors for neurological disorders. Each biospecimen is annotated with standardized de-identified clinical and demographic data. Since the establishment of the NINDS Repository, specimens from over 42,000 unique individuals have been successfully banked at the Coriell Institute for Medical Research and can be accessed through an online catalog (<https://www.coriell.org/1/NINDS>), and over 66,000 catalog samples have been distributed to researchers around the world. The collection includes samples from individuals diagnosed with cerebrovascular diseases (N>12,800), Parkinsonism (N>5,600), motor neuron diseases (N>2,500), Epilepsy (N>6,100), Tourette syndrome (N>4,200), Dystonia (N>3,700), and neurologically normal controls (N>7,500). To date, NINDS Repository samples have been used in over 600 peer-reviewed scientific publications, including genome-wide association studies, whole-exome sequencing studies, and structural variation studies, and these studies have been cited over 66,000 times. The genomic data collected by these studies have been deposited in the Database of Genotypes and Phenotypes (dbGaP), a NIH/NLM sponsored restricted-access data repository for studies investigating the contributions of genetic variation to phenotypic variation and disease (<https://www.ncbi.nlm.nih.gov/gap/>). The NINDS Repository catalog includes a resource page devoted to the annotation of genomic data collected from NINDS Repository samples that are available in dbGaP (<https://www.coriell.org/1/NINDS/NINDS-Samples-in-dbGaP>). Cumulatively, data from 11 dbGaP accessions are cross-referenced to over 7,000 unique NINDS Repository subject samples. Genomic data collections include genome-wide single nucleotide polymorphisms (n=6,229), genome-wide copy number variants (n=895), and whole exome sequencing (n=2,113). This public NINDS Repository collection of biospecimens, clinical and demographic data, and dbGaP annotation has made a significant research impact over the past 20 years and continues to offer extensive resources for use in new genetic and genomic studies of neurological disorders and human disease more generally.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday PB3249*. The One-Sided Matchmaking Platform (OSMP) - Accelerating genomic matchmaking and gene discovery through robust data sharing

Authors:

M. Osmond¹, M. Price¹, T. Hartley¹, C. Klamann², H. Le², J. Xu², K. Mo², H. Driver¹, M. Brudno², O. Buske³, K. Boycott¹; ¹Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, ²Univ. of Toronto, Toronto, ON, Canada, ³Phenotips, Toronto, ON, Canada

Abstract Body:

Gene discovery (associating a gene with a disease for the first time) is essential for continuing to solve undiagnosed rare diseases (RDs). A key piece of evidence for gene discovery is to identify at least one unrelated individual with overlapping phenotype and rare variant(s) in the same gene (called 'matchmaking'). Considering the scarcity of any given RD, finding similarly-affected individuals requires sharing data across borders. However, RD data is often siloed due to privacy concerns, presenting a major barrier to matchmaking efforts.

Recent initiatives such as the Matchmaker Exchange (MME), a federated network of RD databases, have enabled sharing of RD data between institutions for the purposes of matchmaking. Current matchmaking protocols such as the MME use a "two-sided" approach, where users flag candidate genes for individuals with undiagnosed RDs and potential matches are returned only when two parties flag the same gene. This excludes participation from large cohorts that do not have the resources to flag candidates themselves. Further, many cases in the MME do not provide information on phenotype or mode of inheritance up front, which results in lengthy email exchanges between users to determine if cases are a true match. Not only are two-sided approaches labour intensive, but in our experience, have false positive ratios as high as 85%.

To further expand and improve matchmaking approaches, we have developed the One-Sided Matchmaking Platform (OSMP). This web-based portal supports "one-sided" matchmaking, in which variants in novel candidate genes are identified directly from the genome-wide sequencing data within RD databases. An OSMP user may query a gene of interest across participating databases. Using application programming interfaces (APIs), the OSMP returns rare variants in the gene, along with participant-level phenotypic (HPO terms) and genotypic (zygosity, inheritance, variant quality) information. The portal's user-friendly design allows the queriers to filter and sort variant/participant information to rule matches in or out, reducing the need for time-consuming email correspondence. A pilot querying the last 6 months of new gene-disease associations in OMIM is being conducted using OSMP to explore its ability to improve the false positive ratio and identify matches of interest in the RD databases Genomics4RD (Canada) and Genomic Answers for Kids (United States). We anticipate that through robust data sharing practices, the OSMP will provide a more accessible and efficient solution to genomic matchmaking for RD researchers, and foster future connections with additional international repositories.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3250. The Osteogenesis Imperfecta Variant Database: current state

Authors:

S. Stolk¹, L. Zhytnik^{1,2}, T. van Welzenis³, M. Eekhoff¹, L. Wisse¹, A. Maugeri¹, R. Dagleish⁴, G. Pals¹, D. Micha¹; ¹Amsterdam UMC, Amsterdam, Netherlands, ²Dept. of Traumatology and Orthopedics, Inst. of Clinical Med., The Univ. of Tartu, Tartu, Estonia, ³Osteogenesis Imperfecta Federation Europe, Denemarken, Netherlands, ⁴Dept. of Genetics and Genome Biology, Univ. of Leicester,, Leicester, United Kingdom

Abstract Body:

ObjectiveThe Osteogenesis Imperfecta Variant Database (OIVD) contains comprehensive information on variants, their pathogenicity, and patients' phenotypes. Submission of newly found variants to OIVD will further enhance its significance for the OI community. The database was created around 1984 and is widely consulted for OI gene diagnoses and numerous research projects. In 2008 the database migrated to the Leiden Open Variation Database (LOVD), supported initially by the EC GEN2PHEN project. The current OIVD includes variants for 25 genes for OI and overlapping disorders. The database is recognized as a powerful tool for genetic diagnosis, prediction of disease progression, genotype-phenotype correlations, and translational and clinical research. Recently, the main curation of the database was transferred to Amsterdam UMC. We encourage variant submission to OIVD to ensure consolidation and harmonization in OI variant interpretation. **Methods**Variants, published in journals, should be submitted to OIVD. Anybody can become a submitter (https://databases.lovd.nl/shared/docs/LOVD3_submit.pdf). Variant reporting follows current standards (mandatory transcript identifier and use of the HGVS variant nomenclature guidelines), pathogenicity assessment uses the ACMG guidelines, and correct OI clinical types are reported. These measures apply to variants submitted both directly to LOVD-OI and to those from the literature added by the curators. All data are checked and verified by the curators of the database. All OI variant information can be found at <https://lovd.nl/OI-genes>. **Results**OIVD contains approximately 3250 unique OI variants in 6290 patients, with a complete coverage of variants till 2018. Completing and updating the database with published variants (2019-current) is being reinitiated by Amsterdam UMC. Currently, missing OI variants from 2019-2020, are being submitted, including 930 variants from the AGDx. **Conclusion**OIVD is the most comprehensive database of OI genetic variants and patient phenotypic data but few updates have been made since 2018. We strongly encourage our colleagues to submit their (un)published variants and the database is open for submissions. **Funding and acknowledgements**We would like to acknowledge following organizations: Department of Human Genetics, Amsterdam UMC; Leiden University Medical Centre; Osteogenesis Imperfecta Federation Europe (OIFE), The Ehlers-Danlos Society; OI Society Australia. The authors declare no conflicts of interest.

Ethical approvalNot applicable

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3251. TT-Mars: structural variants assessment based on haplotype-resolved assemblies

Authors:

J. Yang, M. Chaisson; Univ. of Southern California, Los Angeles, CA

Abstract Body:

Structural variant (SV) benchmarking is a critical for method development and evaluating the accuracy of genetic variation studies. Typically this has been performed by comparing against a curated SV callset. However this approach is not satisfactory when there are multiple valid representations of the same variant. This drawback is exacerbated in repetitive regions that not only are highly variable, but are also where new sensitive SV discovery methods using long reads find the most variation. We have developed an alternative approach for validating variation, TT-Mars, that scores variant calls based on the sequence implied by the SV rather than the call itself. TT-Mars compares implied genome content to high-quality haplotype-resolved genome assemblies generated from long-read sequencing. In our initial study, we demonstrated about 20.9% additional SVs from short read, and 50.9% from long read data could be validated by TT-Mars, compared to the state of the art, gold-standard based method, Truvari. Here, we present an update to TT-Mars to evaluate calls produced as a result of pangenome graph construction, and whole-genome alignments of de novo assemblies constructed from Oxford Nanopore (ONT) data. Both of these approaches require the flexibility to validate “fragmented” SV calls: multiple nearby calls that in aggregate represent the true call, but individually may not be valid. We find that for graph genomes, when phase data are available, about 4,600 aggregate calls may be validated, and for ONT data about 1,000 aggregate calls are validated.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3252. Using RGD to analyze genes associated with obesity.

Authors:

M. Vedi, J. R. Smith, G. Thomas Hayman, M. L. Kaldunski, S-J. Wang, S. J.F. Laulederkind, M. Hill, W. Demos, M. Tutaj, A. Gibson, L. Lamers, H. S. Nalabolu, K. Thorat, J. Thota, M. A. Tutaj, J. L. De Pons, M. R. Dwinell, A. E. Kwitek; Med. Coll. of Wisconsin, Milwaukee, WI

Abstract Body:

RGD (<http://rgd.mcw.edu>) provides a rich core of human genomic and phenotypic data with an infrastructure of standardized ontologies, disease portals, and many bioinformatics tools to allow users to explore and make disease-related connections among these datasets. These tools and data are integrated with nine other species used as models for human disease (rat, mouse, pig, chinchilla, dog, bonobo, 13-lined ground squirrel, vervet, and naked mole-rat) for translational research.

Obesity is a complex disease and a worldwide health concern due to its implication as a critical risk factor in multiple conditions, including stroke, dementia and COVID-19. As an example workflow, we used RGD's Multi-Ontology Enrichment Tool (MOET) to do an enrichment analysis of the top 200 differentially expressed genes in a gene expression profile dataset of peripheral blood mononuclear cells among normal weight and moderately obese subjects obtained from GEO (GSE69039). The top terms in biological process and pathway ontologies were cellular response to stress and regulatory pathway. The top term in the gene-chemical interactions (ChEBI) ontology was an antirheumatic drug which is interestingly associated with weight loss. To further analyze these genes, users can easily navigate from MOET to the other RGD tools such as Variant Visualizer to investigate clinical variants and GViewer for a genome-wide view of the genes.

RGD provides researchers with the Obesity & Metabolic Syndrome Disease Portal as one of fifteen Disease Portals that offer data for genes, QTLs and rat strains associated with the disease. As another example, we generated a list of human genes common between obesity and stroke using the Obesity & Metabolic Syndrome Disease Portal and Object List Generator and Analyzer (OLGA). This list was used as input for the Variant Visualizer, which found the *LDLR* gene to include many predicted damaging variants and was also linked to Hypercholesterolemia. Additional information about the *LDLR* gene can be found on the RGD gene report page, including disease annotations, clinical variants with associated conditions, and links to the gene report pages of other RGD species. The rat *Ldlr* gene report page shows the multiple *Ldlr* rat strain genetic models available for obesity and obesity-related conditions research. For example, the SD-*Ldlr*^{em1Sage} strain page shows obesity-related disease and mammalian phenotype annotations and PhenoMiner (quantitative phenotype) data. In conclusion, RGD aims to build functionality to support widespread use and assist researchers in finding and utilizing the data and models they need to explore to further gene-to-disease research.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3253*. Variant reclassification based on the allele frequency information from exome sequencing data of 20,455 patients enriched with under-represented populations.

Authors:

K. Kwon, H. Han, G. Seo, H. Lee; 3billion, Seoul, Korea, Republic of

Abstract Body:

Accurate variant classification is essential for making a molecular diagnosis, and therefore, multiple lines of evidence are carefully considered for determining variant pathogenicity. One of the most important lines of evidence used for diagnosing rare Mendelian disorders is the variant allele frequency (AF) information. Filtering out variants based on AF is the most effective way of narrowing down the number of variants that need interpretation, as it removes a substantial number of variants that are too common to be disease-causing. There are several publicly available databases (DB) such as gnomAD that provide AF information from large populations. Many of these large DBs tend to be dominated by few populations though and therefore it is less effective when filtering out variants for individuals from underrepresented populations. Accordingly, many variants remain of uncertain significance (VUS) and even incorrectly classified as pathogenic or likely pathogenic (P/LP) because of its absence in these public DBs. Here, we investigated if a subset of ClinVar P/LP variants and VUS can be downgraded to likely benign (LB) based on AF information from our internal exome sequencing data of 20,455 patients that are not part of any publicly available DBs. From a total of 1.1 billion variants identified in the exome of 20,455 patients, we extracted 399 ClinVar P/LP/VUS variants in clear autosomal dominant genes. These variants were not observed in gnomAD but observed in more than 2 individuals in our DB, while not being reported as disease-causing variants because of inconsistent phenotype. To account for incomplete penetrance and variable expressivity, AF of the most common ClinVar P variant of each gene was obtained from gnomAD and used as a threshold. There were 384 variants with an internal AF greater than this threshold and therefore could be considered LB. Seven of these variants were classified as P/LP in ClinVar, although none of them had ClinVar review status above 1 star. As expected, these 384 variants were mostly found in the “East Asian” (39%) and “others” (30%, non-major) population, suggesting that lack of AF information is the most likely reason for over-classifying the variants. It is not uncommon to find a variant not observed in gnomAD to be very common in a specific population within our dataset consisting of 58% Asians (40% East Asians, 15% South Asians, 3% other Asians). This study highlights the significance of using matched population data to filter variants and the importance of spearheading the effort of sequencing more underrepresented populations.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3254. VariantMatcher a tool to enable connections amongst individuals with interest in a specific variant

Authors:

R. Paulo Martin¹, E. Rodrigues¹, S. Griffith¹, C. Antonescu¹, L. Vail¹, J. Posey², Z. Coban-Akdemir², S. Jhangiani^{2,3}, K. Doheny¹, J. Lupski^{2,3,4,5}, D. Valle¹, A. Hamosh¹, N. Sobreira¹; ¹McKusick-Nathans Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ²Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, ³Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, ⁴Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, ⁵Texas Children's Hosp., Houston, TX

Abstract Body:

VariantMatcher (<https://variantmatcher.org>) was developed to connect individuals around the globe with an interest in a specific variant. It enables sharing of variant-level and phenotypic data from participants in research projects for discovery of disease-causing variants and genes. VariantMatcher contains rare (MAF < 1% in gnomAD) nonsynonymous SNVs identified in 6,235 VCF files (896,847 unique variants) of affected and unaffected individuals sequenced as part of multiple projects and their detailed phenotypic information. Users must register and be approved by site administrators. Users may upload up to 10 genomic coordinates per day and are notified of any match. The query format is “chr:coordinate refAllele > altAllele”. Queries using hg18 or hg38 are lifted over to hg19 prior matching. Phenotypic features can be added and shared in the email notifying the match, but the match is based on the genomic location. When a match occurs, both parties are notified by simultaneous emails. If a match is not made, the queried coordinates can be stored for future matching. VariantMatcher is also connected to the Beacon Network through the Beacon protocol. If a user querying the Beacon Network is informed that the variant is found in VariantMatcher, VariantMatcher registration is required to obtain additional information regarding the variant. As of June 1, 2022, VariantMatcher had 767 submitters from 47 countries; 5,630 variants were queried and 87 variants matched to 982 individuals. The 87 variants affected 82 genes. 64 of these genes are known disease-causing genes. 15 variants in these genes were classified as pathogenic or likely pathogenic in ClinVar, 28 were classified as variant of uncertain significance and 10 were classified as benign or likely benign. To date, no variant match was confirmed as a phenotype match but most of the matches helped to rule out the variant queried as the cause of the phenotype being investigated. Additional query capabilities are planned in upcoming releases, including: indels; variants by zygosity state; specific variant with phenotypic feature(s); and specific group of variants/gene (e.g., individuals with nonsense variants in gene X). In addition, VariantMatcher will implement the standards and protocols from Data Connect API (GA4GH) to connect VariantMatcher with other variant-level databases such as Franklin and MyGene2/Geno2MP. By connecting different variant-level databases and individuals, we expect to improve the variant classification process in both research and clinical settings and also to increase the discovery rate of novel disease-causing variants by increasing the specificity of matches.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday

PB3255. VuTR: Visualising the consequences of high-impact 5' UTR variants

Authors:

E. Dsouza¹, J. Martin-Talbot¹, D. MacArthur², N. Whiffin¹; ¹Univ. of Oxford, Oxford, United Kingdom, ²Ctr. for Population Genomics, Sydney, Australia

Abstract Body:

5'untranslated regions (5' UTRs) are directly upstream of the protein-coding sequence and have important roles in regulating RNA stability, localization, and rate of mRNA translation. Variants within 5' UTRs that disrupt these regulatory processes have been implicated in genetic disorders, however, annotating and interpreting any individual variant of interest is difficult. Approximately 30% of likely pathogenic variants in the ClinVar database potentially act through perturbing upstream open reading frames (uORFs), which are short translated elements in the 5' UTRs that negatively regulate downstream protein translation. Here we present, VuTR, to enable the exploration of annotated variants in 5'UTRs and their impact on uORFs. For any transcript defined by the Matched Annotation from NCBI and EMBL-EBI (MANE) project, the application displays its native 5' uORF architecture overlaid over its genomic sequence along with the likelihood that each uORF is translated given the surrounding Kozak consensus context and harnessing experimental data on translational efficiency. VuTR displays the effect of all 5' UTR variants from the ClinVar and gnomAD databases using annotations from the Ensembl Variant Effect Predictor (VEP) plugin, UTRannotator. Additionally, users are able to query and visualise the impact of any 5' UTR SNV or indel (up to 3bps) of interest. VuTR facilitates annotation and interpretation of variants in 5'UTR regions by both clinical scientists and researchers. In the future, the application will be expanded to include additional regulatory elements. VuTR is available through an easily accessible and open web application build on python3, Flask and nginx (<http://49.12.238.72:8080/>). Its source code and pipeline are released on a GPLv2 License and can be accessed through our GitHub repository (<https://github.com/Computational-Rare-Disease-Genomics-WHG/VuTR>).

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Thursday

PB3256. Web Tools and Resources for Generating and Accessing ACMG Criteria Specifications for Variant Classification by ClinGen Variant Curation Expert Panels

Authors:

N. Shah¹, A. Jackson¹, T. Farris¹, D. Azzariti², M. DiStefano^{2,3}, S. Harrison^{4,2}, M. Mandell⁵, C. Preston⁵, J. Goldstein⁶, M. Wright⁵, T. Klein⁷, S. Plon^{8,9,10}, A. Milosavljevic¹; ¹Baylor Coll. of Med., Houston, TX, ²Broad Inst. of MIT and Harvard, Cambridge, MA, ³Precision Hlth.Program, Geisinger, Danville, PA, ⁴Ambry Genetics, Salt Lake City, UT, ⁵Dept. of BioMed. Data Sci., Stanford Univ. Sch. of Med., Stanford, CA, ⁶Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, ⁷Dept.s of BioMed. Data Sci. & Med., Stanford Univ. Sch. of Med., Stanford, CA, ⁸Molecular and Human Genetics Dept., Baylor Coll. Med., Houston, TX, ⁹Pediatrics-Oncology, Baylor Coll. of Med., Houston, TX, ¹⁰Cancer Genomics Clinical and Res. Programs, Texas Children's Hosp., Houston, TX

Abstract Body:

The ClinGen Sequence Variant Interpretation (SVI) Working Group provides general recommendations for using ACMG/AMP criteria codes to enhance consistency in usage and transparency of variant classification rationales. In addition to following these general recommendations, ClinGen Variant Curation Expert Panels (VCEPs) define and apply their own gene-disease specific recommendations for each of these ACMG evidence codes. We here present multiple web-accessible tools that facilitate generation and dissemination of these specifications in both human- and machine-readable formats. ClinGen Criteria Specification (CSpec) Registry database contains 22 criteria specification documents, defined by 20 VCEPs for more than 50 genes. The CSpec Registry UI (<https://cspec.genome.network>), is a user-friendly web interface for browsing, filtering, and searching criteria specifications by gene and disease. The CSpec data messaging service publishes SVI approved specifications to the ClinGen Data Exchange thereby notifying other ClinGen Curation Tools, like the Variant Curation Interface (VCI), of the approval and release of these specifications. This enables CSpec integration into the VCI curation and ClinGen workflows, as well as provides direct access to the up-to-date specifications by the curators as they classify variants. In addition, the CSpec Editor allows VCEP coordinators and other members to create, update, and release approved specifications for public sharing. The CSpec Registry REST-API service (<https://cspec.genome.network/srvc>) provides programmatic access to the structured content in JSON and JSON-LD formats so that other resources, including those external to ClinGen, can integrate ClinGen approved evidence code specifications into their variant classification and genome interpretation processes. By allowing both programmatic and web user interfaces for developing and releasing the criteria specifications, these tools elevate the transparency and consistency of variant curation processes across ClinGen while empowering research and clinical communities globally. In addition, these resources will be instrumental in supporting our ongoing efforts to automate batch variant curation. Programmatic access further empowers automation of variant classification beyond ClinGen.

Genetic, Genomic, and Epigenomic Annotations, Databases and Resources Posters - Wednesday
PB3257*. Web-CSEA: Web-based Cell-type-Specific Enrichment Analysis of Genes

Authors:

Y. Dai¹, R. Hu¹, A. Liu², K. Cho¹, A. Manuel¹, X. Li¹, X. Dong³, P. Jia², Z. Zhao⁴; ¹UTHlth.at Houston, Houston, TX, ²UTHlth., Houston, TX, ³Brigham & Women's Hosp, Cambridge, MA, ⁴Univ Texas HSC Houston, Houston, TX

Abstract Body:

Background: Human complex traits and common diseases show tissue- and cell-type- specificity. Recently, single-cell RNA sequencing (scRNA-seq) technology has successfully depicted cellular heterogeneity in human tissue, providing an unprecedented opportunity to understand the context-specific expression of complex trait-associated genes in human tissue-cell types (TCs). There is a pressing need for advanced computational methods and web services that could characterize the signature genes of human cell types and, reversely, assess the cell types that a particular gene set may be enriched. **Methods:** We curated a total of 111 scRNA-seq panels of human tissues and 1,355 TCs from 61 different general tissues across 11 human organ systems and developmental stages. We adapted our previous decoding tissue-specificity (*deTS*) algorithm to train and measure the enrichment for each tissue-cell type (TC). To overcome the potential bias from the number of signature genes between different TCs, we further calculated a permutation-test p-value with ~20,000 previously calculated p-values derived from genetic trait-associated gene sets of diverse human diseases and traits, which naturally contains moderate TC-specificity. **Results:** We present the first web-based application to quickly assess the cell-type-specificity of genes, named Web-based Cell-type Specific Enrichment Analysis of Genes (WebCSEA, available at <https://bioinfo.uth.edu/webcsea/>). We provide extensive visualization functions such as interactive heatmap, and interactive and static jitter plots to display the cell-type specificity across 1,355 human TCs by human organ system, developmental stage, and top-ranked tissues and cell types. We benchmark the performance of our WebCSEA by using a few ground truth gene lists from cell-type signature genes, disease risk genes with unknown onset tissues, and cell-type-specific pathways. Users could use WebCSEA to explore the TC-specificity of focal gene lists in different scenarios, including 1) cellular context for candidate genes of complex disease, 2) validation for tissue-specific and cell type-specific signatures, 3) differentially expressed genes from bulk transcriptome/epigenome, 4) validation of developmental stage-specific genes, and 5) tissue-cell-type specific gene ontology/KEGG pathway. **Conclusion:** WebCSEA is a user-friendly interactive platform that provides a comprehensive exploration of the TC-specificity of genes among human major TC map. Just like the gene-set enrichment analysis, we expect WebCSEA to serve the broad users in both research and clinical community to explore the biological context of human genes.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3258. 8 Billion Polygenic Risk Score distributions: building unique reference populations based on genetic ancestry to increase the performance of PRS

Authors:

G. Busby, A. Bolli, J. Kintzle, P. Di Domenico, G. Bottà; Allelica, New York, NY

Abstract Body:

The clinical utility of PRS depends on a clear assessment of their predictive performance in the population in which they are intended to be used. Typically, this is achieved by estimating the association (i.e. effect size) of a PRS with disease in a range of different Testing populations. Predicting risk in a new Target individual involves identifying the most appropriate reference distribution from these Testing populations, placing the Target individual within this distribution, and computing the relative increase or decrease in disease risk after applying the appropriate population-specific effect size. Current approaches assume that Target individuals belong to one of a fixed number of a priori defined reference populations. These are typically categorized by (sub-)continental ancestry or ethnicity labels (European, African American etc). While appealing, such categorization enforces boundaries between human populations that are both artificial and potentially misleading. Human genetic diversity is continuous and genetic ancestry is a description of history more than geography.

We present a novel approach to defining a PRS reference distribution that selects the unique set of reference individuals who are most genetically similar to a Target individual without having to assign them to a pre-defined group. The reference individuals are selected from a balanced and deep dataset of individuals from diverse genetic ancestries. Using a combined genetic and clinical dataset of almost 500,000 individuals across the spectrum of global genetic diversity, we test the ability of multiple ancestry-specific PRS to predict risk of breast and prostate cancer in a range of individuals. We construct unique reference distributions using a multivariate approach to identify genetically similar individuals. Comparisons of the performance of these distributions show a significant increase in the accuracy of PRS compared with approaches that place individuals into a limited set of predefined reference populations. Our study shows the potential for large, diverse genetics datasets to power the clinical use of PRS by providing a framework for building personalized PRS reference distributions for the accurate prediction of risk in all 8 billion people on earth.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3259. A Cox proportional hazards whole genome sequencing study of ischemic stroke risk in a Brazilian sickle cell population

Authors:

E. Earley¹, F. Fang², E. Sabino³, C. Dinardo⁴, S. Kelly⁵, B. Custer⁵, G. Page⁶, NHLBI Recipient Epidemiology Donor Evaluation Study (REDS-III) International Component - Brazil; ¹RTI Intl., Durham, NC, ²RTI Intl., Lexington, MA, ³Univ. de São Paulo, São Paulo, Brazil, ⁴Inst. de Med. Tropical, Univ. of São Paulo, São Paulo, Brazil, ⁵Vitalant Res. Inst., San Francisco, CA, ⁶RTI Intl., Atlanta, GA

Abstract Body:

Ischemic stroke is a major complication of sickle cell disease (SCD) and without preventive care affects 11% of children with SCD before the age of 20. Ischemic stroke is more common in children, whereas hemorrhagic stroke is more common in SCD patients who are 20-30 year-olds. Previous genetic studies of stroke in SCD have been either un-replicated or conflicting. To address this gap, we have conducted a genome-wide association study (GWAS) of ischemic stroke within SCD with three novel features: 1) the largest cohort to date at N=1,333 HbSS homozygotes; 2) early-life risk association with a cox proportional hazards model; 3) whole genome sequencing data.

Medical records were abstracted including a history of stroke using standardized definitions across 6 Brazilian sites in Brazil. Cox Proportional Hazards model was run with stroke as the outcome on 1,333 homozygous HbSS individuals (178 Ischemic strokes) using whole genome sequence (MAF > 1%; minimum depth 10x) within the NHLBI Trans-Omics for Precision Medicine (TOPMed), and the top 10 principal components were used as covariates to control for population stratification. Alleles were coded as dosages (0, 1, or 2) across 14M sites.

14 genomic regions were associated with early ischemic stroke at genome wide significance ($P < 5 \times 10^{-8}$). This included variants near two genes which have been previously linked to pediatric or early onset stroke in non-SCD cohorts: *ADAMTS2* (rs147625068, $P = 3.70 \times 10^{-9}$) and *CDK18* (rs12144136, $P = 2.38 \times 10^{-9}$). Individuals in our study harboring multiple risk alleles exhibited increasing rates of stroke at earlier timepoints ($P < 0.001$, Gehan-Wilcoxon). Significant gene enrichment was observed in multiple tissue classes, including upregulated genes in the coronary (8 genes, $P = 0.0005$, FDR), tibial (9, $P = 0.03$, FDR) and aorta arteries (7, $P = 0.03$, FDR), as well as downregulated genes in the spleen (7, $P = 0.005$), pancreas (13 genes, $P = 0.02$), esophagus mucosa (8, $P = 0.03$), hypothalamus (10, $P = 0.03$), and substantia nigra (11, $P = 0.03$). We also observed enrichment within the GWAS Catalog disease category white matter lesion progression (*CNNM2*, *NT5C2*, $P = 0.01$) and the combined phenotype of coronary artery disease or large artery stroke (*CNNM2*, *NT5C2*, $P = 0.04$).

In conclusion, we have conducted the largest GWAS of ischemic stroke in SCD to date. Results have replicated known associations with pediatric and early stroke (<65 years) for genes *ADAMTS2* and *CDK18*, respectively. Pathway analysis suggests gene dysregulation in the arteries (coronary, tibial, and aorta), spleen, pancreas, esophagus mucosa, hypothalamus, and substantia nigra.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3260. A fast Bayesian screen to identify pleiotropic loci and describe pleiotropic profiles.

Authors:

S. Huo¹, L. Wang², S. Lindström², P. Kraft¹; ¹Harvard Univ., Boston, MA, ²Univ. of Washington, Seattle, WA

Abstract Body:

Pleiotropy occurs when a genetic variant is associated with more than one trait. Multi-trait test statistics have been proposed that leverage pleiotropy to increase power in identifying trait-associated loci. However, most of these methods only test the global null that none of the tested traits are associated with a variant but do not provide any information regarding whether the variant is associated with more than one trait, or with which traits the variant is associated. In this paper, we propose a new fast screening framework based on the Bayesian support region to overcome these restrictions. Our approach accounts for correlation among test statistics due to sample overlap and leverages cross-trait heritability. Computation scales linearly in the number of traits. We compare our approach to widely used alternatives (including the likelihood-ratio test, Bonferroni correction for the number of traits, and ASSET, among others) via simulation. Our approach shows both high sensitivity and high specificity and outperforms the alternatives under most scenarios. For example, our approach can correctly detect up to 67.8% more pleiotropic SNPs than Bonferroni correction. We applied our approach to GWAS summary statistics from 12 different cancers and identified 82 independent regions exhibiting pleiotropy, including *TERT*, *ABO*, and the HLA region, each associated with three cancers. We hope that this new method can facilitate biological discoveries in the future.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3261. A fast variational inference approach for powerful genome-wide association of quantitative traits.

Authors:

H. Loya^{1,2}, G. Kalantzis¹, A. Pazokitoroudi³, S. Sankararaman^{3,4,5}, P. Palamara^{1,2}; ¹Dept. of Statistics, Univ. of Oxford, Oxford, United Kingdom, ²Wellcome Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, ³Dept. of Computer Sci., UCLA, Los Angeles, CA, ⁴Dept. of Human Genetics, David Geffen Sch. of Med., UCLA, Los Angeles, CA, ⁵Dept. of Computational Med., David Geffen Sch. of Med., UCLA, Los Angeles, CA

Abstract Body:

The rapid growth of modern biobanks is creating new opportunities for large-scale genome-wide association studies (GWAS) and the analysis of complex traits. The computational cost of performing GWAS in millions of samples, however, often leads to a trade-off between computational efficiency and statistical power, reducing the benefits of data collection efforts. We developed a method, called Quickdraws, that achieves increased association power without sacrificing efficiency. Quickdraws uses a non-infinitesimal model based on a mean-field spike-and-slab prior on phenotypic effects, which is efficiently optimized using stochastic variational inference, GPU matrix operations, and code vectorization for parallel trait analysis.

We tested the statistical power and robustness of Quickdraws and other available GWAS methods by simulating data sets of 50k UK Biobank samples including up to 10% of related individuals, as well as 50 quantitative traits with 2% polygenicity, 40% heritability, and population stratification. Quickdraws, BOLT-LMM-MoG, and FastGWA were robust to relatedness and stratification, while linear regression was inflated in the presence of relatedness and Regenie was inflated when only very close relatives were included. We evaluated association power using average χ^2 test statistics at simulated causal SNPs. Quickdraws and BOLT-LMM-MoG yielded similar average χ^2 statistics (7.38, SE = 0.25 and 7.43, SE = 0.25, respectively), significantly higher ($p < 10^{-25}$) than Regenie (7.06, SE = 0.22), FastGWA (6.76, SE = 0.22), and linear regression (6.72, SE = 0.22).

We applied Quickdraws and other methods to 50 quantitative blood-related traits in 404k self-reported UK Biobank White British samples. We performed variant clumping (Plink p-value threshold = 5×10^{-9} , r^2 threshold = 0.01) to extract sets of approximately independent variants. Consistent with simulations, Quickdraws and BOLT-LMM-MoG detected more associations than other methods ($N_{\text{Quickdraws}} = 14,801$, $N_{\text{BOLT-LMM-MoG}} = 14,796$, $N_{\text{BOLT-LMM-Inf}} = 13,713$, $N_{\text{Regenie}} = 14,149$, $N_{\text{FastGWA}} = 11,287$). We compared computational costs by fitting 458k model SNPs and performing association for 7.4 million imputed SNPs. Quickdraws was 43x faster than BOLT-LMM-MoG and 20% faster than Regenie v3 ($T_{\text{Quickdraws}} = 10.6\text{h fitting} + 22\text{h association testing}$, $T_{\text{BOLT-LMM-MoG}} = 1,150\text{h} + 255\text{h}$, $T_{\text{Regenie}} = 4.7\text{h} + 36.2\text{h}$).

These results highlight the promise of leveraging modern machine learning techniques such as stochastic variational inference and GPU-based acceleration to develop scalable GWAS association methods without sacrificing statistical robustness or power.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3262. A framework to improve the interpretation and prediction of variant effect sizes using non-linear functional models

Authors:

S. Cheng^{1,2}, S. Gazal^{1,2}; ¹Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ²Ctr. for Genetic Epidemiology, Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA

Abstract Body:

Modeling the relationship between the effect sizes of common variants and their functional annotations is critical to accurately characterize the functional genetic architecture of human diseases and complex traits. Existing functional models rely on linear relationships between variants squared effect sizes (or per-variant heritability) and their functional annotations; this linear relationship is methodologically constrained by the fact that we are observing marginal effect sizes from genome-wide association studies (GWAS), and that true effect sizes are unobserved. However, linear models constrained our interpretation of functional architectures, and it is thus unclear if a non-linear structure would improve model prediction. Here, we aim to improve interpretation and prediction of existing functional models by using machine learning approaches.

We developed a framework that 1) leverage the fine-mapping method SuSie (Wang et al. 2020 Journal of the Royal Statistical Society) to estimate the true allele effect sizes of each common variant, and 2) model per-variant heritability directly on these estimated effects (rather than marginal effects) using 96 functional annotations from the baseline-LD model (Gazal et al. 2017 Nat Genet). First, using in 16 independent traits of the UK Biobank, we observed that a simple linear regression on SuSie effects provide similar functional enrichment than the state-of-the-art method S-LDSC (Finucane et al. 2015 Nat Genet) that estimate enrichment directly on marginal effect sizes ($r = 0.93$ across 40 functional annotations), validating the approach and allowing further investigation of non-linear models. Second, we ran decision trees and deep neural networks using a different set of hyper parameters on these 16 traits, and observed that these approaches outperform the linear model when looking at the mean squared error (MSE), and that they are robust to overfitting when using a leave-one-chromosome-out (LOCO) approach. We compared functional enrichment obtained with these approaches and S-LDSC to validate current S-LDSC estimates. Finally, we ran decision trees with only 3 layers on 16 UK Biobank traits, and observed a more interpretable functional architecture of human disease and complex traits.

To summarize, we have developed and validated a framework using non-linear models to investigate functional architectures of human diseases and complex traits. These new predictions could be leveraged to improve priors of fine-mapping and/or polygenic risk score analyses.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3263. A gene-based association test of interactions for maternal-fetal genotypes identifies genes associated with non-syndromic congenital heart defects.

Authors:

M. Huang¹, C. Lyu², N. Liu¹, W. N. Nembhard³, C. A. Hobbs⁴, M. Li¹, The National Birth Defect Prevention Study; ¹Indiana Univ., Bloomington, IN, ²New York Univ. Grossman Sch. of Med., New York, NY, ³Univ. of Arkansas for Med. Sci., Little Rock, AR, ⁴Rady Children's Inst. for Genomic Med., San Diego, CA

Abstract Body:

The risk of congenital heart defects (CHDs) may be influenced by maternal genes, fetal genes, and their interactions. Existing methods commonly test the effect of maternal and fetal variants one-at-a-time, and may have reduced statistical power to detect genetic variants with low minor allele frequencies. In this article, we propose a gene-based association test of interactions for maternal-fetal genotypes (GATI-MFG). GATI-MFG is a region-based method that can integrate the effect of multiple variants within a gene or genomic region, and thus, may detect the associations of rare variants with improved power. It also evaluates the joint effect of maternal and fetal genotypes while allowing for their interactions. In simulation studies, GATI-MFG had improved statistical power over alternative methods, such as the single-locus test and functional data analysis under various disease scenarios. We further applied the method to a two-phase genome-wide association study of CHDs using 947 CHD case mother-infant pairs and 1,306 control mother-infant pairs from the National Birth Defects Prevention Study. After Bonferroni adjustment for 23,035 genes, two genes on chromosome 17, *TMEM107* (p-value=1.64e-06) and *CTCI* (p-value=2.0e-06), were identified for significant association with CHD risk. Gene *TMEM107* encodes a ciliary transition zone protein which regulates ciliogenesis and ciliary protein composition. Previous studies have shown that *TMEM107* is associated with heterotaxy. Gene *CTCI* plays an essential role in protecting telomere from degradation. Previous studies have suggested that telomere length is associated with heart development, function, and disease. Our findings are consistent with existing literature supporting the association between *CTCI* and CHDs.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3264. A generalized framework to decorrelate family or population structure that can easily incorporate epistasis, GxE interaction or multiple-variant analysis in GWAS and NGS studies

Authors:

D. Li, D. McGovern; Cedars-Sinai Med. Ctr., Los Angeles, CA

Abstract Body:

Background: Linear mixed models (LMMs) have been widely used in Genome-wide Association Studies (GWAS) or Next Generation sequencing studies (NGS) to account for population or family structures. Still, most of the proposed algorithms are limited to single-variant analysis and there are intensive needs for novel algorithms to accommodate analyses such as epistasis, GxE interaction or multiple-variant analyses.

Methods: Instead of directly incorporate the Genetic Relation Matrix (GRM) as a random term in the LMM, we utilized an extension of generalized least square (GLS) as a way to “decorrelate” the genetic structure. In brief, a reverse of the Choleski Decomposition of the GRM was multiplied to the original genetic matrix G and phenotype vector Y , which generates the transformed G' and Y' .

It can be proved that with this transformation, direct regression of Y' on G' will yield a Best Linear Unbiased Prediction (BLUP) of the coefficient β . With this approach, all the methods developed for epistasis, GxE interaction or multi-variant analysis can be applied to the decorrelated G' and Y' .

A simulation was performed to evaluate the performance of the proposed GLS approach across different scenarios. Performance was compared to no adjustment for genetic structures (lm_{crude}) under the null hypothesis across different scenarios including single-variant, epistasis, GxE interaction as well multiple causal variants. Performance was also compared to the LMM based approach, when the corresponding algorithm is available (single variant, epistasis, GxE).

With the multi-variant association a SKAT-CommonRare based association was performed in the transformed G' and Y' . In this case the performance of the GLS was not compared to LMM as currently there is no LMM based SKAT algorithm available.

Results: Under the null hypothesis, the proposed GLS approach retained nominal type I error rate across all the scenarios. For example, when the genetic structure was from trio families, with a p-value threshold of 0.05, type I error rate is 0.048 for GLS and 0.092 for the lm_{crude} in the multi-variant analyses.

Under the alternative hypothesis, the proposed GLS approach has comparable power compared to the LMM based analysis. For example, in the single-variant analysis, when there are both family and population structures, the power is 0.865 for GLS and 0.857 for LMM.

Conclusion: With the proposed GLS approach, investigators can easily apply different algorithms with independent hypothesis to GWAS or NGS datasets with complex genetic structures.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3265*. A genetic correlation disease-disease network (gcDDN) for the improved identification of novel phenotype relationships

Authors:

V. Sriram, J. Woerner, Y. Nam, D. Kim; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Many complex disorders share associations with multiple phenotypes. However, our understanding of these disease connections is still severely limited, and it is unclear how much of a role is played by genetics. A disease-disease network (DDN), a graph where nodes represent diseases, can help to identify related disorders. By applying cross-trait linkage disequilibrium score regression to the data generated from a biobank-scaled phenome-wide association study (PheWAS), we can generate a corresponding genetic correlation DDN (gcDDN), where edges represent genetic associations between phenotypes. We hypothesize that a gcDDN can help to better identify novel disease co-occurrences as well as highlight potential genetic contributors to these disease connections, particularly compared to conventional PheWAS-derived DDNs, where edges simply represent the number of phenotype-associated variants that pass a specified significance threshold and are shared across nodes. We constructed a gcDDN for 487 phenotypes that had at least 1000 cases from UK Biobank (UKBB) PheWAS summary data. To demonstrate the gcDDN's clinical significance, we focused on myocardial infarction (MI), a polygenic disease with high heritability, as our index phenotype of interest. Graph-based semi-supervised learning was applied to generate scores of predicted disease co-occurrence between MI and its neighboring phenotypes in the gcDDN. These predictions were validated against ground truth disease co-occurrences derived from UKBB in-patient data. We find that the gcDDN identifies disease connections with MI evinced in the in-patient data more accurately than a conventional DDN, with a 22.8% increase in the area under the receiver operating curve (AUC) from 0.632 to 0.776. This result suggests that a gcDDN can effectively identify relationships between diseases for complex, heritable phenotypes, indicating its relevance in further explorations of personalized medicine and disease comorbidity. Remaining work involves the assessment of gcDDN for additional phenotypes as well as an interpretation of the specific genetic variants that contribute to connections in the network.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3266. A genome-wide search for parent-of-origin effects in 4,505 unrelated individuals identifies the *8q21.3* locus as a novel genetic regulator of Factor V plasma levels.

Authors:

B. Gendre¹, A. Martinez-Perez², M. Kleber³, A. Van Hylckama Vlieg⁴, A. Boland³, N. Saut⁵, J-C. Souto⁶, M. Germain⁷, G. Delgado⁸, J. Deleuze³, M. Winfried⁸, M. Sabater-Lleal⁶, F. Rosendaal⁴, J-M. Soria⁶, P-E. Morange³, D-A. tregouet⁷; ¹Univ. of Bordeaux, Bordeaux, France, ²Hosp Sant Pau, Barcelona, Barcelona, Spain, ³Leiden Univ., Med. Ctr., Leiden, Netherlands, ⁵CRB-HV, Aix-Marseille Univ., Marseille, France, ⁶Res. Inst. Hosp. Sant Pau, Barcelona, Spain, ⁷INSERM, Bordeaux, France, ⁸V Med. Clinic, Univ. of Heidelberg, Mannheim, Germany

Abstract Body:

Factor V (FV) is a key molecular player in the coagulation cascade. FV plasma levels have been associated with several human diseases, including thrombosis, bleeding and diabetic complications. Two genes have been robustly found to contribute in the inter-individual variability of plasma FV levels: structural *F5* gene and *PLXDC2*. Because the later has been observed to be an imprinted gene in some rodent model, we investigated if parent-of-origin effects (POE), a specific epigenetic mechanism linked to imprinting, could participate in the genetic regulation of FV plasma levels. For this, we deployed the QUICKTEST program dedicated to the detection of POE in GWAS datasets of unrelated individuals. QUICKTEST is based on the Brown-Forsythe statistical methodology and suggests the presence of POE at a given polymorphism when the phenotypic variance observed in heterozygotes is significantly higher than that observed in homozygotes. QUICKTEST was applied to 4 independent GWAS studies (LURIC, MARTHA, MEGA and RETROVE) totaling 4,505 participants of European ancestry with measured FV plasma levels. The regression coefficients characterizing the difference of phenotypic variances and obtained in the 4 cohorts were then meta-analyzed using a fixed effect model.

We observed a genome-wide significant association at *SLC7A13/PSKH2* locus, on *chr8q21.3*, where the lead variant, *rs75463553*, associated with a POE regression coefficient of 0.128 ± 0.022 ($p = 8.4 \cdot 10^{-9}$). This effect was very homogeneous across the 4 contributing studies: 0.098 ± 0.033 , 0.181 ± 0.049 , 0.147 ± 0.050 , 0.115 ± 0.055 in LURIC, MARTHA, MEGA and RETROVE, respectively. Of note, the *rs75463553* did not show any association with mean FV plasma levels in the 4,505 individuals ($p = 0.49$). No evidence for POE was detected at the *PLXDC2* locus (minimum p-value = $6.1 \cdot 10^{-3}$). Of importance, other mechanisms than POE, such as haplotype effects, gene x gene or gene x environment interactions, could produce significant QUICKTEST results. Preliminary haplotype and epistasis analyses remained inconclusive and family studies are now underway to confirm that the *chr8q21.3* locus contributes to the genetic regulation of FV plasma levels through POE. The association of the identified locus with methylation marks is also ongoing.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3267. A locus near *ICOS* modifies risk for *Mycobacterium avium* complex airway infection in people with cystic fibrosis.

Authors:

W. Gordon¹, A. Faino², A. Stilp¹, E. Blue^{1,3}, K. Buckingham¹, M. Rosenfeld^{2,1}, P. Qu², J. Collaco⁴, R. Pace⁵, G. Cutting⁴, M. Knowles⁵, M. Bamshad^{1,3,2}, R. Gibson^{2,1}, Cystic Fibrosis Genome Project; ¹Univ. of Washington, Seattle, WA, ²Seattle Children's Res. Inst., Seattle, WA, ³Brotman Baty Inst. for Precision Med., Seattle, WA, ⁴Johns Hopkins Univ., Baltimore, MD, ⁵Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract Body:

Nontuberculous mycobacteria (NTM) is an increasingly common lung disease in people with cystic fibrosis (PwCF) associated with substantial comorbidities and mortality. The two most common NTM complexes in PwCF are *Mycobacterium avium* complex (MAC) and *Mycobacterium abscessus* complex (MABSC). Most incident NTM infections in PwCF do not progress to NTM lung disease and microbiological criteria for NTM disease require ≥ 2 NTM+ sputum cultures. Here we focus on NTM lung infection as an antecedent to NTM lung disease.

To find genetic modifiers of risk for MAC and MABSC infections in PwCF we identified 3,321 pancreatic insufficient PwCF of European ancestry with NTM sputum cultures in the U.S. Cystic Fibrosis Foundation Patient Registry. Controls were defined as PwCF with no NTM+ sputum cultures and ≥ 2 years of NTM culture data. We defined 2 infection outcomes of interest— ≥ 1 (“incident”) or ≥ 2 (“repeated”) positive cultures—for both MAC and MABSC, totaling 4 outcomes. Among the 3,321 PwCF, 1,830 met our uninfected control criteria. For both infection outcomes, we found that MAC infection was more common (353 had ≥ 1 MAC+ cultures, 143 had ≥ 2 MAC+ cultures) than MABSC infection (257 had ≥ 1 MABSC+ cultures, 120 had ≥ 2 MABSC+ cultures).

We conducted a genome-wide association study on whole genome sequencing data (hg38) from the Cystic Fibrosis Genome Project. Cases for each of the 4 outcomes were compared with the shared control set. Relatedness estimates and principal components (PCs) were calculated using PC-Relate and PC-AiR. We fit a generalized linear mixed model with a binomial family using the GMMAT approach, adjusting for sex, birth cohort, PCs 1 and 2, and genetic relatedness as a random effect. Assuming an additive effect, we performed score tests for variants with a minor allele count > 10 and missingness $< 2\%$.

Genomic inflation was absent ($1.00 \leq \lambda \leq 1.01$). An association signal near *ICOS* was observed in both the tests of ≥ 1 MAC+ cultures (chr2:204111791, $p=2.12 \times 10^{-7}$, minor allele frequency (MAF)=0.03) and ≥ 2 MAC+ cultures (chr2:204109873, $p=2.05 \times 10^{-9}$, MAF=0.02). A less significant signal was seen in this region for tests of ≥ 1 MABSC+ cultures (chr2:204076672, $p=1.13 \times 10^{-5}$, MAF=0.02) and ≥ 2 MABSC+ cultures (chr2:204111791, $p=1.12 \times 10^{-3}$, MAF=0.02).

ICOS encodes the inducible T cell costimulator protein. Reduced expression of *ICOS* has been associated with NTM lung disease in people without cystic fibrosis, and our results indicate that in PwCF variants near *ICOS* modify risk for NTM infection. Together, these observations suggest that within the lung *ICOS* plays roles in susceptibility to NTM infection and disease.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3268. A machine-learning approach for disease prediction improves GWAS and polygenic scores

Authors:

L. Eick, M. Cordioli, S. Jukarainen, Z. Yang, FinnGen, A. Ganna; Inst. for Molecular Med. Finland, FIMM, Helsinki, Finland

Abstract Body:

Background/Objectives:

GWAS assume there is a correct classification of subjects into cases and controls. Nonetheless, misclassification is expected, especially in biobanks that rely on administrative health data. Accounting for misclassification and for diseases liability can improve GWAS discovery and development of more predictive polygenic scores.

Methods:

To address these limitations, we implemented a gradient boosted classifier (XGBoost) that we used to predict 6 diseases (type 2 diabetes, ischemic stroke, coronary heart disease (CHD), Alzheimer, Dementia and Breast cancer) by integrating information on 3828 diagnoses, 499 medications, 25 socio-economic variables and other health information. Instead of binary disease labels, the model outputs a continuous liability measure which was used to perform GWAS. The analyses were conducted on N=317,687 individuals from the FinnGen study.

Results:

We predicted disease prevalence with an AUC between 90% (type 2 diabetes) and 97% (Alzheimer's disease). First, we perform a GWAS of the predicted liability values. Results recapitulate a case/control GWAS and in some instances identify new loci (e.g. 3 loci in stroke). Second, we performed a GWAS of disease liability only in the control group. To see if the model was capable of assigning high liability values to those who may be at high risk of developing the disease. The results of the liability GWAS in controls were highly genetically correlated with the case-control GWAS (e.g. $r_g=0.88$, $P\text{-value}=3.7e-7, 6.3e-14$ for stroke and CHD). We also identify established loci, for example APOE among alzheimer controls ($rs429358$, $P\text{-value}=1.97e-29$)

We used MTAG to combine the results from the case-control GWAS with the disease liability GWAS among controls. These combined results improved prediction ability of polygenic scores when tested in UK Biobank, for some but not all traits (e.g. $AUC=0.537$ to $AUC=0.546$, $P\text{-value difference}=2e-16$ for stroke).

Conclusion:

Our method extends the capabilities of GWAS by recovering signals from the control group. The possibility to go beyond a case-control model for GWAS analyses opens new venues to the development of more predictive polygenic scores. However, interpretation of new biological results from this approach remains challenging.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3269. A multi-omics risk score approach to identify individual risk for incident type 2 diabetes

Authors:

H. Ng¹, M. Loh^{1,2,3}, D. Tay¹, L. Lakshmanan¹, F. Tai¹, X. Guan¹, R. van Dam^{4,5}, X. Sim⁶, J. Chambers^{1,7}; ¹Nanyang Technological Univ., Singapore, Singapore, ²Imperial Coll. London, London, United Kingdom, ³Natl. Skin Ctr., Singapore, Singapore, ⁴Natl. Univ. of Singapore, Singapore, Singapore, ⁵Havard Univ., Boston, MA, ⁶Natl. Univ Singapore, Singapore, Singapore, ⁷Imperial Coll. London, London, United Kingdom

Abstract Body:

Introduction: Type-2 diabetes (T2D) is one of the leading causes of mortality and morbidity globally. Environment factors such as diet and physical inactivity as well as genetics are recognized as important factors that contribute to T2D risk, with polygenic risk score (PRS) gaining popularity as a method to estimate an individual's risk of developing T2D. In addition, other omics data such as epigenetic and metabolomics have shown promising results in recent years. However, these omics studies have been carried out predominantly amongst people of European ancestry. To improve risk score prediction, we took a multi-omics approach by combining genetic data (PRS) with DNA methylation data and metabolomics data [methylation risk score (MRS) and metabolomic risk score (MetRS) respectively] and assess the performance of the risk scores with respect to incident T2D prediction. In particular, we investigated the performance of our risk scores in a multi-ethnic South East Asian populations, whereby data are currently lacking globally.

Method: The PRS and MetRS were developed from summary statistic derived from published studies while the MRS was developed from in-house unpublished DNA methylation data measured on Illumina 450K and EPIC array, using baseline samples (4,349 East Asians, Malays and South Asians from United Kingdom and Singapore) collected before onset of T2D. The performance of the multi-omics risk scores were then evaluated in an independent series of 1,819 South Asians (900 incident T2D; 919 controls) with genetic (GSA array), methylation (EPIC array) and metabolomics data (H-NMR, Nightingale).

Results: In a univariate model, relative risks of PRS and MRS were 1.37 [95% confidence interval (CI): 1.24-1.51, P = 2.01E-10] and 1.58 [95% CI:1.43-1.75, P=1.07E-18] per 1 standard deviation increase in risk score respectively. When included in a multivariate model including traditional risk factors such as body mass index (BMI), HbA1c and fasting glucose, in additional to age and gender, the PRS and MRS still remained highly and independently significant, with relative risks of 1.27 [95% CI: 1.14-1.43; P= 3.43E-05] and 1.33 [95% CI: 1.18-1.49; P= 3.16E-06] respectively. The analysis of MetRS is currently ongoing.

Summary: In addition to the traditional risk factors and the more commonly used PRS, molecular risk scores from other omics data such as epigenetics could improve prediction of T2D, thereby optimising stratification of resources for disease surveillance and education to targeted individuals with higher risk for disease, leading to improved disease prevention and treatment.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3270*. A network based approach for fine-scale population structure inference in genetic datasets

Authors:

R. Shemirani¹, S. Cullina¹, C. Gignoux², N. Zaitlen³, G. Belbin⁴, E. Kenny⁵; ¹Icahn Sch. of Med. at Mt. Sinai, New York, NY, ²Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ³UCLA, Los Angeles, CA, ⁴Gencove Inc., New York, NY, ⁵Icahn Sch. of Med. at Mt Sinai, New York, NY

Abstract Body:

Background: The Genetic Relatedness Matrix and Principal Components (PCs) of the genotype data play a prominent role as proxies for population structure in statistical genetics applications such as association studies. They assist with model calibration and correcting for confounding factors including environmental signals not captured in genetic datasets. Inadequacies of these methods in delineating recent fine-scale structure could result in spurious findings. Here, we use segments of the genome inherited from a common ancestor identical-by-descent (IBD) to create a network of samples with shared recent ancestry. We demonstrate that the graphical properties of such networks carry additional structural information that augments the information derived from PCs. **Methods:** First, we propose a novel approach to generate IBD networks by dissecting IBD segments into contemporary groups to preserve temporal structure. Secondly, we introduce a method to extract fine-scale structural data from these networks in the form of spectral representations of nodes as the laplacian eigenvectors. Finally, we demonstrate that these representations carry signals derived from recent population structure via observing the proportion of variance in traits explained by them. **Results:** In the UK Biobank dataset, our method generates 6 networks with ~19 billion edges connecting 350,110 samples with British ancestry. Using easting and northing birth coordinates as examples of traits with solely environmental origins, our spectral representations yield an increase of 66.7%[66.5-66.8] and 27.9%[27.6-28.0] in the proportion of variance explained when compared to the first 40 PCs, respectively. This demonstrates the ability of our method in extracting signals of recent structure, which also results in a 43.8%[43.1-44.6] and 177.3%[156.7-209.7] increase in the proportion of variance explained in the overall health conditions and HDL Cholesterol levels in individuals, respectively. Combining the spectral representation with the PCs results in an increase of 97.3%[55.6-104.3] in proportion of variation explained for the body-mass index, suggesting that these representations can complement PCs in genetic studies. **Conclusions:** Our method could supplement the current approaches and improve the power of genetic studies by accounting for recent fine-scale population structure, especially in diverse and admixed cohorts where PCs provide inadequate environmental data to address confounding issues. We plan to investigate the efficacy of our method in overcoming the biases that stem from unbalanced datasets in the of calibration Polygenic Risk Scores in such cohorts.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3271*. A novel approach identified eight gene x alcohol or gene x smoking interactions that contribute to serum lipids

Authors:

X. Zhu¹, Y. Yang², N. Lorincz-Comi¹, G. Li¹, A. Bentley³, P. de Vries⁴; ¹Case Western Reserve Univ., Cleveland, OH, ²case western reserve Univ., Cleveland, OH, ³NIH, Bethesda, MD, ⁴UTHlth.at Houston, Houston, TX

Abstract Body:

Introduction: Gene-environment (GxE) interactions have been suggested to contribute significantly to the phenotypic variation of complex traits. However, detection of GxE interactions has been challenging even in large sample sizes, as demonstrated in multiple large GxE interaction studies in the CHARGE consortium. **Methods:** We performed a novel analysis to detect gene×smoking and gene×alcohol interactions on serum lipids using trans-ethnic summary statistics from large genome wide association studies of lipid levels: 1) the Global Lipids Genetics Consortium study (GLGC, (Graham et al., 2021, N=1.65M); 2) Studies of GxE interactions of smoking and alcohol from the CHARGE consortium (Bentley et al 2019, de Vries et al. 2019, N=134K). We compared the genetic effect estimated in GLGC with the main effect estimated in the GxE interaction model from CHARGE in a way similar to Mendelian Randomization (MR) analysis, i.e, selecting lipids associated variants as the instruments and considering the genetic effects from CHARGE as exposure effects and that from GLGC as outcome effects. We then searched the genome for significant departures from the causal line ($P < 5E-8$) to identify genetic variants that either have a GxE interaction effect or an effect on the trait that is mediated by the environmental factor. To exclude variants with mediation effects through the environmental factor, we excluded all loci associated with smoking status or alcohol drinking ($P < 5E-7$ in Liu et al, 2019, Jiang et al 2021). We further examined the identified variants by using the interaction test directly and corrected for multiple testing using Bonferroni corrections. **Results:** Through MR analysis, we identified 3, 6 and 4 genome-wide significant independent loci for high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglycerides (TG) with alcohol, and 3, 5 and 5 loci for LDL-C, HDL-C and TG and smoking, respectively. By filtering out the variants using the interaction test, we identified the following gene x alcohol interactions: *APOE* and *LDLR* for LDL-C, *GALNT2* and *CETP* for HDL-C, and *APOE*, *ZPR1* and *RPL30P9 (LPL)* for TG. We identified gene x smoking interactions for *APOE* with LDL-C, *CETP* with HDL-C, and *APOE*, *ZPR1*, *RPL30P9 (LPL)* and *TRIB1* with TG, respectively. **Conclusion:** We identified 8 gene-smoking or gene-alcohol interactions that contribute to serum lipids traits. All the genes/loci have been reported to be associated with lipids levels, but their interaction effects have not been confirmed. Our approach can be useful in efficiently detecting interactions by combining large GWAS with or without incorporating interactions.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3272. A novel computational deconvolution method to estimate sample-level cell-type-specific gene expression profile

Authors:

K. Kang, M. Kellis; Massachusetts Inst. of Technology, Cambridge, MA

Abstract Body:

The relatively high cost of single cell sequencing may render large cohort studies resorting to bulk RNAseq, however, heterogeneity in tissues hampers efforts to probe the role of each cell type. Hence, in bulk data analysis, deconvolution is a key step to extract cell-type-specific information. Nevertheless, most such tools focus on estimating cell type fractions. Only a few are able to infer cell-type-specific gene expression profiles (ctsGEP) among which barely one or two aim at estimating ctsGEP for each individual. Yet, the technical challenges demand more sophisticated approaches for better estimates which are invaluable for various downstream analyses, such as cell-type-specific differential expression inference and eQTL calling, to list but a few.

To this end, we developed a deconvolution method to estimate both sample-level ctsGEPs and the cell-type proportions simultaneously using bulk RNAseq. We modeled the observed expression data using multinomial distributions whose parameters reflect the unknown ctsGEP and sample-specific proportions. Our model incorporated prior information from single cell data and we used a Dirichlet distribution to control the degree of confidence in the priors. Integrating these components, we built a Bayesian model that fully captures the stochastic nature of bulk data. We employed an MCMC procedure for parameter estimation. Importantly, the hyperparameters of the hierarchical model, which is critical in posterior inference, were also being optimized in a principled way.

We benchmarked our method using mixtures of varied but known composition. Specifically, we generated 48 in silico mixture samples using human brain single cell data from 48 individuals. We performed deconvolution using a reference single cell profile from a different study as a prior and evaluated its performance by comparing the estimated sample-level ctsGEPs with the ground truth. We assessed the accuracy of ctsGEP for each individual; and, more importantly, the ability to capture expression variation of genes for each cell type across all samples. For the former, our method was able to achieve around 0.9 in correlation; For the latter, we were able to attain above 0.6 in correlation across samples for around 80% of genes. Our method, in both cases, outperformed existing alternatives with averagely 16% less Root-Mean-Square-Errors on ctsGEP estimations.

Our method holds promise for computationally deciphering complex mixtures of cell types, each with differing expression profiles in different individuals, using RNA-seq data measured in bulk tissue.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3273. A Novel Gene-Based Test for Sequencing Studies Based on a Bayesian Variable Selection of Rare Variants

Authors:

J. Xu¹, N. Wang², L. Briollais³; ¹Lunenfeld-Tanenbaum Res. Inst., Toronto, ON, Canada, ²Univ. of New Brunswick, Fredericton, NB, Canada, ³Mount Sinai Hosp, Toronto, ON, Canada

Abstract Body:

Background: Next Generation Sequencing (NGS) technology provides opportunities to discover rare variants (RVs) associated with complex human diseases. A usual paradigm for detecting RVs in NGS experiments is to perform a gene-based (or region-based) association test followed by single RV tests in the top selected genes. Gene-based tests often select or assign higher weights to coding variants to enhance the power to detect associations but to the cost of limiting the possibility for new RV discoveries. On the other end, the inclusion of all RVs in a gene-based test might reduce its power. As an alternative, we propose a novel strategy that performs a variable selection of the RVs to be considered in the gene-based test as a powerful approach.

Methods: Our novel approach extends our previous region-based test for case-control designs using a Bayes Factor statistic (Xu et al., *Biometrics*, 2020) where the association between a set of RVs in the same region (e.g., a gene) and a disease was assessed. Our novel method is based on generalized linear regression models and its conjugate prior (Chen et al., *Statistica Sinica*, 2003), which can handle outcomes of different types (binary, continuous, count), informative functional annotations and unbalanced designs. A birth-death MCMC algorithm is used to select the most important RVs to be considered in the region-based test.

Results: Through simulations of sequencing data with our R package *sim1000G*, we show that performing a variable selection step of RVs in a gene-based test is a more powerful approach than considering all possible RVs in the gene as proposed originally in SKAT and Burden gene-based tests. Assigning weights to RVs in a gene-based test can enhance its power only when those weights are consistent with the variable selection. Weights can be used as prior in our novel Bayesian approach. Application to UK Biobank whole-exome sequencing (WES) data with lung cancer outcome show that our method leads to new gene and RV discoveries. Conclusion: Our new approach improves existing RV gene-based tests while leveraging new RV discoveries.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3274. A novel haplotype based approach allows for conducting ancestry agnostic genome wide association studies.

Authors:

D. Perry¹, M. Gymrek², K. Frazer³, M. D'Antonio¹; ¹Univ. of California, San Diego, San Diego, CA, ²Univ California San Diego, La Jolla, CA, ³UC San Diego, SAN DIEGO, CA

Abstract Body:

Genome-wide association studies (GWAS) are heavily biased towards analyzing individuals of European descent, and translating findings to diverse or admixed populations has proven challenging. It is thus necessary to make GWAS more informative for individuals from non-European backgrounds and enhance equity in the distribution of benefits from genetics research.

Haplotypes may serve as an ancestry-agnostic tool to study genetic associations and have been proposed as an alternative approach because they reduce redundancy due to LD and require a fewer number of tests when compared to single nucleotide variants (SNVs). Haplotype blocks are traditionally defined using recombination hotspots. However, a major downfall to this approach is that the resolution (~700 bp) of recombination hotspots is low compared to each block size (2-10 kb). Therefore, an empirical method to detect haplotype block boundaries may be more appropriate to exploit haplotypes in genetic association studies.

Here, we present a novel method for constructing haplotype blocks. In a given genome block size, we select the SNVs with MAF $\geq 20\%$, then count the number of observed genotype combinations. If this is greater than a predetermined threshold, the haplotype block is defined by these SNVs. Otherwise, SNVs are added until the observed number of genotype combinations exceeds the determined threshold. Once the haplotype block boundaries are set, haplotypes are defined according to the allele frequency of all the SNVs included in each block.

To demonstrate the utility of this approach, we investigated the APOE locus on chromosome 19, where a haplotype consisting of two SNPs (rs429358 and rs7412) is known to have strong associations with neurological disorders and blood traits. GWAS have also described different associations between LDL in European and African individuals in this region. We conducted a GWAS using SNPs on LDL levels in a 250kb region flanking APOE in both African and European individuals in the UK Biobank. We also conducted a GWAS using haplotypes on LDL levels. We then separately finemapped the results from both methods using SuSiE. We observed the same lead haplotype in both African and European individuals, which comprises the previously described APOE- $\epsilon 2$ (rs429358-T; rs7412-T), known to be associated with atherosclerosis and cholesterol levels. Alternatively, SNP credible sets were discordant between Europeans and Africans and did not include rs429358. Our method of constructing haplotypes is robust to differences in LD structure between different populations, making it relevant for analyzing individuals from admixed and diverse populations.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3275. A novel method for annotation-agnostic differential transcript usage analysis

Authors:

K. Eldjarn Hjoerleifsson¹, P. Melsted², L. Pachter¹; ¹Caltech, Pasadena, CA, ²Univ Iceland, Reykjavik, Iceland

Abstract Body:

The quantification of RNA reads is a key step in most analyses of RNA-Seq data. Current quantification methods rely on annotations of the organisms' transcriptomes, which may be incomplete or nonexistent. This may lead to data being discarded and to erroneous quantifications. In downstream applications such as eQTL analysis, such errors can propagate and result in missed associations. We present a novel method for associating phenotypes with RNA expression, that can identify expression associations resulting from a wide variety of underlying transcriptional and post-transcriptional events, without requiring a prior annotation of the transcriptome. By constructing a *de Bruijn graph* of all the reads mapping to a single gene, and pruning away nodes that are likely to be erroneous, we obtain a representation of the expression of the gene in our cohort. Each expressed isoform constitutes one path through the graph. We then run associations on the expression of each individual node and a phenotype. Should an isoform of the gene associate with the phenotype, there will be a set of nodes in the graph that uniquely identify the isoform, the expression of which also associates with expression of the phenotype. This method enables discovery of novel alternative polyadenylation, exon-skipping, duplications, insertions, deletions, and circular RNA, among other transcriptional and post-transcriptional variations, without prior knowledge of these events. We show that we can reliably reproduce known associations, and detect, *de novo*, phenotypically relevant transcriptional structures.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3276. A novel method for multiple phenotype association studies based on genotype and phenotype network

Authors:

X. Cao, S. Zhang, Q. Sha; Michigan Technological Univ., Houghton, MI

Abstract Body:

Although genome-wide association studies (GWAS) have emerged as a common and powerful tool to detect the complexity of the genotype-phenotype associations, a common limitation of GWAS is that they focus on only a single phenotype at a time. Joint analysis of multiple correlated phenotypes for GWAS can identify and interpret pleiotropic loci which are essential to understanding pleiotropy in diseases and complex traits. Meanwhile, constructing a network based on associations between phenotypes and genotypes provides a new insight to analyze multiple phenotypes, which can explore whether phenotypes and genotypes might be related to each other at a higher level of cellular and organismal organization. Notably, it might broaden the understanding of genetic architecture that exists between diagnoses, genes, and pleiotropy. In this research, we develop a bipartite signed network by linking phenotypes and genotypes into a Genotype and Phenotype Network (GPN). The GPN can be constructed by both quantitative and qualitative phenotypes and is applicable to binary phenotypes that have extremely unbalanced case-control ratios for large-scale biobank datasets. After projecting genotypes into phenotypes, the genetic correlation of phenotypes can be calculated based on the shared associations among all genotypes. Then we propose a novel community detection method to partition phenotypes into disjoint modules based on the genetic correlation. For each module, we can apply multiple phenotype methods to test the association between phenotypes in a module and a SNP. Simulation results show that most multiple phenotype tests based on network modules are much more powerful than those based on all phenotypes, especially when phenotypes are affected by different directions. Furthermore, we apply the proposed method to a set of diseases from the UK Biobank. All multiple phenotype tests based on network modules can identify more potentially SNPs than those without considering network modules.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3277. A pipeline for predicting pathogenicity of monogenic cardiovascular disease genetic variants.

Authors:

M. Ramaker, J. W. Abdulrahim, W. E. Kraus, S. H. Shah; Duke Univ., Durham, NC

Abstract Body:

Large improvements in sequencing capabilities have allowed for increased rare variant detection in large population-based cohorts, yet interpretation of such rare genetic variation has proven difficult. The majority of identified rare variants are of unknown significance even in patients with suspected monogenic diseases. Thus, there is an unmet need for annotation of variants of unknown significance (VUS) to both properly diagnose and better understand monogenic diseases. Furthermore, monogenic cardiovascular diseases encompass a variety of phenotypes that often lead to major cardiac problems including heart failure. Most patients do not harbor a known pathogenic or likely pathogenic variant associated with disease but harbor several VUSs in known monogenic cardiovascular disease genes. Several publicly available in-silico annotation tools predict the pathogenicity of variants across the genome using various statistical and predictive strategies based on measures of conservation, sequence, and protein information amongst others. Additionally, large genomics projects aimed at understanding multi-omic tissue specificity offer a resource for tissue-specific variant interpretation. Here we report the combined use of such resources via machine learning to predict the pathogenicity of genetic variation in 48 monogenic cardiovascular disease genes in a tissue-specific manner. Random forest was trained using six-fold cross validation on ClinVar pathogenic, likely pathogenic, benign, and likely benign variants (n = 26,402) from 48 monogenic cardiovascular disease genes to build a cardiac-specific predictive model of variant pathogenicity utilizing predicted measures of deleteriousness and splicogenicity, local pathogenic variation, cardiac-specific exon expression, and population allele frequency as model predictors. The top-performing model was validated in a held-out subset of variants for which ClinVar clinical significance had changed from benign, likely benign or VUS status to pathogenic or likely pathogenic (or vice versa) from 2014 to 2022 (n = 445) resulting in a ROC AUC of 0.66. Last, we predicted VUS pathogenicity in the CATHeterization GENetics (CATHGEN) cohort. We identified 6 predicted pathogenic VUSs in 6 individuals with hypertrophic cardiomyopathy (HCM). Of those, a highly conserved missense VUS in the myosin heavy chain 7 gene (MYH7) previously reported in HCM patients was coincidentally returned 2 weeks prior via genetic panel testing and is suspected to be the disease-causing variant.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3278. A polygenic framework nominating causal SNPs, their target genes, and cell-types of action

Authors:

A. Kim, S. Gazal; Univ. of Southern California, Los Angeles, CA

Abstract Body:

Genome-wide association studies (GWAS) have identified thousands of disease-associated loci, but they generally do not identify the underlying causal variants, target genes and cell-types, thus limiting our understanding of disease mechanisms. Recent advances in functionally-informed fine-mapping (FIFM) have improved our ability to nominate causal variants by prioritizing variants in disease-relevant functional annotations. However, the current state-of-the-art FIFM method (PolyFun; Weissbrod et al. 2020 Nat Genet) does not leverage the cell-type-specificity of human diseases, thus not optimally maximizing their power. In addition, identifying the function of fine-mapped variants remains a critical challenge as causal variants are predominantly non-coding SNPs that do not necessarily regulate the closest genes.

Here, we propose a novel polygenic framework integrating new cell-type specific regulatory datasets, recent FIFM methods and SNP-to-Gene (S2G) linking strategies to improve nomination of causal variants and directly characterize their function.

First, we improved PolyFun strategy by considering the context specificity of complex traits. Specifically, we optimally select cell-type specific annotations from a set of >1,000 annotations using stratified LD score regression (S-LDSC). Our context-specific strategy refined the fine-mapping results of the default PolyFun, validating 80% of causal (PIP>0.95) SNP-trait pairs and identifying 13% novel SNP-trait relationships (preliminary analysis of 8 UK Biobank traits). Second, we leveraged in vitro and in silico predictions of variant functionality and showed that our strategy identifies a significantly greater ($P < 0.05$) proportion of functional variants directly modulating gene expression. Finally, we built a probabilistic framework estimating the probability for a cell-type to be causal and leveraging SNP-to-gene maps (Gazal et al. 2022 Nat Genet) to compute SNP-gene-cell-type-disease quadruplet scores, thus easily characterizing the function of fine-mapped variants. To illustrate the benefits of our framework, we highlight several examples of previously undetected causal SNPs as well as their predicted target genes and affected cell-types, providing the molecular basis of variant causality.

Our results indicate the relevance of context-specific FIFM and propose a novel polygenic framework to directly characterize the function of fine-mapped variants. This new method will significantly improve biological interpretation of fine-mapped variants and help the translation of GWAS findings into discoveries that will enhance disease treatment.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3279. A polygenic score method including recessive model for improving odds ratio.

Authors:

R. Ota, S. Morishita; The Univ. of Tokyo, Chiba, Japan

Abstract Body:

For studying polygenic diseases, GWAS has been largely contributing to identifying associated single nucleotide variants (SNVs); however, SNVs for polygenic diseases often have low odds ratio (1.1-1.3) and only explain a small proportion of heritability, which is called "missing heritability." To solve this problem, a number of statistical methods to estimate polygenic scores (PGS) have been proposed. Several PGS has been proposed, including LASSO, lassosum, LDpred, and C+T. All of these methods exploited additive models, which do not consider low-frequency recessive SNVs, and are likely to use a large number of SNVs except for LASSO which is capable of regularizing the number of SNVs in PGS. In an attempt to build a more accurate method that requires a smaller number of essential SNVs, we explore the boosting approach. Adaboost, a classic and widely used boosting method for qualitative phenotype, adopts a linear model for better interpretation of the prediction model and is known to be infrequent to be overfitting. However, Adaboost is not good at capturing low-frequency (recessive) SNVs. To take low-frequency SNVs into account properly, we propose an improved boosting method, **GenoBoost**, that considers the genetic inheritance models (additive, dominant, or recessive). GenoBoost demands no information on LD beforehand but can implicitly exclude the effect of LD, which is the unique nature of boosting algorithm. GenoBoost can be interpreted as an "iterative algorithm for multivariate logistic regression" or an extension of conditional analysis. We also developed a program that can process very large GWAS data such as the entire UK Biobank data, which contains about 300 thousand samples each of which has more than 7 million SNVs, in a reasonable amount of time (12 hours) using a main memory of 32GB. GenoBoost showed the best or comparative odds ratio for some polygenic diseases while using fewer SNVs to achieve the odds ratio. For BC, the odds ratio at high-risk proportion of 5% of GenoBoost was 2.1 and the second-best LASSO was 2.05, and GenoBoost used 300 SNVs while LASSO did 430. We also showed that the GenoBoost's PGS with any genetic model was better than GenoBoost's additive model only PGS, especially at smaller proportions like high-risk proportion of 1%. We also found that some recessive SNVs were reported to be associated with the diseases before. In conclusion, we propose a first PGS algorithm that incorporates any genetic models, as far as we know, and outputs better odds ratios with sparser models than state-of-the-art methods in some diseases, which demonstrates that incorporating the recessive SNVs, some of which were reported, improves the accuracy.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3280. A precise, combinatorial approach for the estimation of *de novo* mutation rate from short-read WGS data.

Authors:

M. Kohailan¹, W. Aamer², N. Syed², S. Padmajeya², S. Hussein², A. Sayed², J. Janardhanan², S. Palaniswamy², N. El hajj¹, A. Akil², K. Fakhro²; ¹Hamad Bin Khalifa Univ., Doha, Qatar, ²Sidra Med., Doha, Qatar

Abstract Body:

De novo mutations (DNMs) are variants found in the genome of a child, yet absent in both parents. While DNMs are a critical driver of diversity in any species, they also play a major role in severe genetic disorders. Thus, it is important to understand the rate, distribution and pattern of DNMs. Considering the rarity of DNMs in the genome, detecting such mutations requires prudent approach that can extract them from NGS data. Previous studies followed a single computational approach to call DNMs. In addition, no such large scale genetic analysis was performed in the Middle-East. Here, we applied a combinatorial approach by using three different tools to generate an integrated list of DNMs per individual. A total of 645 individuals, mainly Middle-Eastern with Arab ancestry, were enrolled in this study; generating 353 trio combinations. For these trios, we identified 24,808 *de novo* single-nucleotide variants (SNVs) and 2,405 insertions-deletions (INDELs), with a median of 70 SNVs and 6 INDELs per individual. We calculated a median *de novo* mutation rate of 1.25×10^{-8} and 1.07×10^{-9} per base per generation for SNVs and INDELs, respectively. We determined the parent-of-origin for around 13% of the *de novo* variants. We found a paternal to maternal DNMs ratio of approximately 3.96:1. We then plotted the number of phased DNMs in each individual against parental age at conception. We observed a significant increase of 1.79 DNMS per year of paternal age and 0.34 DNMS per year of Maternal age. However, this correlation substantially differs between families. We also examined the DNM spectra and mutational signature. We found a clear enrichment of transitions over transversions and that the mutations at CpG sites contribute to a large fraction of DNM events. We then questioned whether DNMs at CpGs are correlated with high methylation levels. To answer this question, we compared the mutation rates at CpGs with respect to the level of methylation, and found that highly methylated CpGs are 2.05 times more likely to have mutations than low-methylation sites. This study illustrates the importance of using combinatorial approaches for DNM calling. It also serves as a reference for DNM discovery in multiplex families from the globally under-represented populations of the Middle-East. Future studies with larger cohorts will be required to shed the light on the genetic architecture of the Arab population, with important implications for screening and intervention strategies.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3281. A Rare Missense Variant of Large Effect is Associated with Cataract in Puerto Ricans

Authors:

J. Shi¹, J. O'Connell¹, B. Hicks¹, W. Wang¹, V. Vacic², A. Auton^{1,2}, S. Shringarpure¹; ¹23andMe, Inc., Sunnyvale, CA, ²23andMe, Inc. Therapeutics Div., South San Francisco, CA

Abstract Body:

Founder populations can enable novel genetic discoveries through enrichment of globally-rare variation due to genetic drift. Here, we identify a novel genetic association for cataract in Puerto Ricans by constructing an imputation panel of Puerto Ricans (PR) and performing GWAS on cataract. We found a rare missense variant association on chromosome 2q31.1 with very high OR, and the variant is also linked with significantly earlier age-at-onset for cataract.

Using the research participant base of 23andMe, Inc., we verified that PR had a founder event and are thus enriched for globally rare variations. We selected a panel of 500 unrelated PR samples for whole genome sequencing (WGS), with a mean depth of 22X generated by Broad. We built an imputation panel with 117 million variants, and imputed the sequence variants into 58,694 additional chip-genotyped and phased Puerto Ricans. We ran GWAS on the imputed PR cohort against the self-reported cataract phenotype with 2,951 cases and 42,069 controls. We identified a missense & splice-region variant on chromosome 2q31.1 with significant association (p-value=1.5e-12, OR=12.904, 95% CI=[6.649, 25.042]) to cataract. In addition, the age-at-onset (AAO) for cataract is significantly lower among risk allele carriers than non-carriers, with 62% of carrier cases developing cataracts before 30 years of age, while only 0.67% of non-carrier cases reported age-of-onset < 30 (binomial p-value = 7.11e-11).

The missense variant is population-specific, with a minor allele frequency (MAF) of 0.089% in PR, but 0% in Europeans, African Americans, East and South Asians, 0.0056% in Latinos (gnomAD v2.1.1). Therefore, its association could only be found through the PR GWAS. In fact, except for the Puerto Ricans, we were not able to include this variant in our GWASs for the Europeans, Latinos, East Asians, South Asians and African Americans due to its 0% MAF in those populations. Our results demonstrate that GWAS in founder populations can identify novel genetic associations for globally-rare variation.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3282. A robust whole-genome Mendelian randomization approach for improved estimation and inference of causal effects

Authors:

H. Zhang^{1,2}, Z. Liu³, X. Lin⁴; ¹Harvard T.H. Chan Sch. of Publ. Hlth., Cambridge, MA, ²Natl. Cancer Inst., Rockville, MD, ³The Univ. of Hong Kong, Hong Kong, China, ⁴Harvard T.H. Chan Sch Pub Hlth, Boston, MA

Abstract Body:

Introduction: Mendelian randomization (MR) is a powerful tool to assess the causal association between modifiable risk factors and disease outcomes using genetic variants as instrumental variables (IVs). The method only requires summary-level statistics from genome-wide association studies (GWAS), hence reducing the logistical cost and facilitating its widespread application.

Method: MR makes several strong and empirically unverifiable assumptions, which can be violated in practice and lead to biased estimates and invalid statistical inferences. For example, the most prevalent MR approach, inverse-variance weighted (IVW) method, can be biased due to the use of weak IVs, pleiotropic effects, or sample overlaps. Other cutting-edge approaches, such as MR-Egger, MR-RAPs and MR-PRESSO, aim to reduce bias in the presence of pleiotropic effects. We develop a whole-genome MR method (WMR) that integrates variants across the genome while accounting for linkage disequilibrium (LD) between variants. We make weaker assumptions about pleiotropic effects by assuming an arbitrary distribution with mean of zero and a constant variance. The method enhances the accuracy of MR estimate, while efficiently reducing the bias in the presence of weak IVs, pleiotropic effects and sample overlaps.

Simulations: Our simulation analyses mimic realistic LD patterns using European haplotype data from Phase 3 of the 1000 Genomes Project. We compare WMR with a number of existing methods, including IVW, MR-Egger, MR-Raps, and MR-PRESSO, based on a range of simulation settings, such as different proportions of causal variations, pleiotropic effect levels, and sample overlap proportions. We find that WMR is robust in the presence of numerous weak IVs, various degrees of pleiotropic effects, and sample overlap proportion. In addition, WMR provides an MR estimate with more precision, i.e., a smaller empirical standard error (SE), compared to existing methods.

Data Analyses: We utilize the proposed WMR methods in conjunction with other existing MR methodologies utilizing the GWAS summary-level statistics from several distinct variables, such as body mass index, lipids-related phenotypes, cardiovascular disease, breast cancer, etc. Consistent with models, the WMR estimations are more precise and have smaller SEs than other methodologies, resulting in more significant MR results.

Conclusion: WMR is a robust and powerful MR method that reduces biases in the presence of a large number of weak instruments and pleiotropic effects and provides accurate and effective causal effect estimates.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3283. A scalable framework for robust linear mixed model association testing

Authors:

G. Kalantzis¹, A. Pazokitoroudi², H. Loya¹, H. Chen³, S. Sankararaman⁴, P. Palamara^{1,5}; ¹Univ. of Oxford, Oxford, United Kingdom, ²UCLA, LA, CA, ³The Univ of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ⁴UCLA, Los Angeles, CA, ⁵Wellcome Ctr. for Human Genetics, Oxford, United Kingdom

Abstract Body:

Linear mixed models (LMMs) provide state-of-the-art performance in genome-wide association studies (GWAS) but are computationally demanding when the number of analyzed individuals and traits increases. We developed a LMM algorithm, called FMA, which remains scalable while preserving statistical power and robustness to confounding. FMA relies on moment-based multiple variance component estimation and can be used to analyze multiple quantitative traits in parallel. We performed extensive simulations to compare FMA with several scalable methods, including BOLT-LMM, BOLT-LMM-Inf, Regenie, FastGWA, and linear regression (LinReg). We simulated sets of 50 traits with realistic MAF/LD-dependent architectures, 25% trait heritability, 1% or 5% polygenicity, using 387k genotyped variants (MAF>0.001) and 50k UK Biobank (UKB) samples of varying levels of relatedness or population structure. In unrelated samples of homogeneous ancestry, BOLT-LMM achieved the highest increase in association power over LinReg (+3-7.4% χ^2 on top SNPs), while FMA, REGENIE, and BOLT-LMM-Inf performed similarly (+0.8-2.6% χ^2 on top SNPs). When related individuals were included in the analysis, all methods yielded controlled false positive rates (FPR < 5%) with the exception of LinReg (FPR 5.1-5.2%; $p < 1e-4$). In scenarios involving European samples and stratification, where 5% of phenotypic variance was explained by the top 10 principal components (PCs) of ancestry, FMA and BOLT-LMM were calibrated without the need to use PCs as covariates. Regenie, FastGWA, and LinReg were inflated when PCs were not used (5.9-8.2% FPR; $p < 1e-4$), but calibrated when PCs were included. For model fitting (excluding the calculation of test statistics) FMA was faster than BOLT-LMM (x16, or x6.5 for BOLT-LMM-Inf) in analyzing all traits, but slower than REGENIE (x4.4) and FastGWA (x10, or x2.6 including kinship calculation). Finally, we applied FMA (testing with PLINK, using pgen format) and Regenie (using bgen format) to 20 real phenotypes, 446,050 UKB samples, and 38 million imputed variants. FMA+PLINK required 98.8 hours for model fitting and association testing (82.5 and 16.3 hours, respectively), slower by a factor of x1.62 compared to Regenie (4.7 and 56.4 hours). FMA had slightly improved calibration, measured using attenuation ratio statistics computed using LD score regression ($p = 1e-4$), while both approaches resulted in a similar number of associated loci being replicated in association summary statistics from 13 common Biobank Japan traits. Overall, our work presents a new scalable and robust LMM association algorithm and extensive benchmarks of available GWAS methodology.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3284. A scalable pipeline for robust relatedness inference with application to a study of 1,144,542 individuals

Authors:

S. Gaynor¹, J. Staples¹, R. Panea¹, T. Joseph¹, S. Balasubramanian¹, A. Locke¹, A. Ziyatdinov¹, A. Marcketta¹, J. Backman¹, X. Bai¹, W. Salerno¹, J. Kosmicki¹, J. Emberson², R. Collins², J. Torres², P. Kuri Morales³, R. Tapia-Conyer³, J. Alegre³, J. Berumen³, Regeneron Genetics Center, RGC Research Partners, G. Abecasis¹, J. Marchini¹, T. Thornton¹; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Univ. of Oxford, Oxford, United Kingdom, ³Univ. Natl. Autónoma de México, Mexico City, Mexico

Abstract Body:

To optimize genetic discovery, large-scale studies often include both population-based and family-based studies comprised of closely related individuals. Accounting for relatedness has impact on analyses ranging from principal components analysis (PCA) for population structure inference to genome-wide association and sequencing studies for the mapping of complex traits. For example, inaccurate detection of relationships confounds association analysis and can induce spurious association. However, the identification of familial relationships can fail in the presence of admixture or population structure and poses significant computational costs, particularly for large-scale biobank studies.

We present a cloud-based pipeline for inferring relatedness that can robustly and scalably detect relationships in diverse biobank studies. We implement a comprehensive relatedness pipeline that detects identical-by-descent (IBD) segments using the KING algorithm for relationship estimation and inference, and reconstructs pedigrees using the PRIMUS approach. The pipeline permits complete relationship estimation for additional samples when studies are expanded without study-wide re-estimation required. We use six cohorts, varying in size, population structure, and genotyping and sequencing platform, to demonstrate the performance of the pipeline is robust to the variant set used for IBD segmentation given genome-wide data.

We demonstrate the pipeline to be computationally scalable with on-demand resources in one of the largest genetic collections available. We apply our relatedness pipeline to a large-scale study of 1,144,542 individuals from forty-six cohorts across multiple continents with genome-wide data. The contributing cohorts, including the Mexico City Prospective Study and the UK Biobank, differ in demographic composition and study design, leading to moderate to substantial degrees of relatedness. The analysis was optimized to be fully parallelized and completed in hours. We report over 200,000 first-degree relatives, over 300,000 second-degree relatives within this analysis, including relationships across cohorts and populations. Our results demonstrate the extent to which relatedness can be present in genetic analysis of large-scale studies, and thus the importance of accurately inferring and applying methods for relatedness information.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3285. A scalable statistical framework for genome-wide interaction testing harnessing cross-trait correlations with an application to Alzheimer's Disease

Authors:

S. Bian¹, A. Bass², Y. Liu³, A. Wingo³, T. Wingo³, D. Cutler², M. Epstein²; ¹Dept. of Biostatistics and Bioinformatics, Emory Univ., Atlanta, GA, ²Dept. of Human Genetics, Sch. of Med., Emory Univ., Atlanta, GA, ³Dept.s of Neurology and Human Genetics, Emory Univ., Atlanta, GA

Abstract Body:

Well-powered GWAS have shown that many complex human phenotypes originate from thousands of trait loci. Nevertheless, for many traits, family-based heritability estimates are often considerably larger than the corresponding heritability estimates from GWAS SNP data. For example, the family-based estimates of heritability for Alzheimer's disease (AD) are around 60-80% whereas reported GWAS SNP heritability estimates only range between 6-41%. While even larger GWAS samples may fill this heritability gap, there is interest in exploring whether this missing heritability could also be due to non-additive effects of genetic variation including gene-gene and gene-environment interaction effects. Exploring such non-additive effects is inherently challenging due to many factors including the modest magnitude of interaction effect sizes, the crippling multiple-testing burden, the complications with determining the interacting variable(s), and impractical computational time and cost. However, for a single quantitative phenotype, interaction methods exist that circumvent these issues by assessing whether the variance of the phenotype differs by genotype category. In this work, we propose a scalable method for interaction testing in the spirit of these variance-based approaches but instead leveraging valuable information contained within multiple correlated phenotypes for improved performance. Specifically, we can show that SNPs with interactive effects yield differential correlation patterns among phenotypes per genotype category, which we can formally test. Our proposed test first applies linear regression to assess relationship between SNP genotype and pairwise cross-product term among phenotypes. We then combine the resulting pairwise cross-product regression p-values together in a Cauchy aggregate statistic to form an optimal test. Our method, which we call SCAMPI, is computationally scalable to genome-wide analyses (with similar run times to variance-based interaction methods), can handle large number of phenotypes, and can adjust for confounders. Type I error and power simulations verified that SCAMPI was well calibrated and had ability to detect even sparse interaction effects observed among large group of phenotypes that were modeled. We further applied SCAMPI to a GWAS study of AD pathologies/cognitive trajectory within the ROS/MAP study. Considering variants with $MAF \geq 0.05$, SCAMPI identified interaction effects among these phenotypes within known AD-risk genes (e.g. TP63 and DLGAP2) and novel risk genes (e.g. LINC00382). SCAMPI is also scalable to large biobank data. SCAMPI is implemented as an R package for public use.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3286. A simple distance-based model predicts effector genes at GWAS loci

Authors:

D. Seaton¹, G. Atla¹, H. Tran², C. Robins³, E. Dermitzakis⁴, R. Scott¹, T. Johnson¹; ¹GlaxoSmithKline, Stevenage, United Kingdom, ²GlaxoSmithKline, Collegeville, PA, ³GlaxoSmithKline, Atlanta, GA, ⁴GlaxoSmithKline, Geneva, Switzerland

Abstract Body:

The derivation of causal biological insights from common variant studies in human populations requires mapping disease-associated variants to their effector genes. This has led to considerable investment in *in silico* methods to support variant-to-gene mapping. While the best-performing models integrate multiple sources of information, the distance between variant and gene remains the most strongly predictive feature in state-of-the-art models [1, 2]. This is consistent with observations from QTL studies that the proximity of a variant to a gene is a strong predictor of whether it perturbs the expression and/or function of that gene.

We present a simple model of the distribution of gene-perturbing variants in the genome, which we refer to as the Distance-Based Model (DBM). These variants can be divided into two categories: those that affect gene function via regulatory mechanisms, and those that act through a change in protein sequence. The probability of observing such variants depends on genomic position, and we model this as a function of the distances between variants and genes. This model has 4 free parameters, which we fit by reference to the distribution of pQTL lead variants around transcription start sites. We then use this model to infer effector genes at GWAS loci, given the observed positions of GWAS lead variants.

As expected, this model recapitulates the empirical observation that the nearest gene to the GWAS signal is often the effector gene. More surprisingly, when benchmarked alongside the Open Targets L2G model [1] and Effector Index model [2], the DBM achieves state-of-the-art performance in predicting effector genes. For the Open Targets Genetics GWAS loci (June 2021 release), the Area Under the Precision Recall Curve (AUPRC) of the DBM against the Open Targets gold standard sets was 0.75, 0.54, and 0.31 for the high, medium, and low confidence gold standards, respectively. This is similar to the AUPRCs of 0.76, 0.43, and 0.25 of the Open Targets L2G model. Similarly, the AUPRC of the DBM against the EI gold standard was 0.43, while the EI model had an AUPRC of 0.35. These results refine widely used proximity-based heuristics for interpretation of GWAS data, and suggest directions for future development of models to predict effector genes at GWAS loci.

[1] E. Mountjoy *et al.*, *Nat. Genet.*, 2021, doi: 10.1038/s41588-021-00945-5. [2] V. Forgetta *et al.*, *Hum. Genet.*, 2022, doi: 10.1007/s00439-022-02434-z.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3287. A statistical framework for integrative genetic association analysis.

Authors:

J. Okamoto¹, F. Luca², R. Pique-Regi³, A. Helms¹, L. Wang¹, J. Morrison¹, X. Wen⁴; ¹Univ. of Michigan, Ann Arbor, MI, ²Wayne State Univ., Detroit, MI, ³Wayne State Univ, Detroit, MI, ⁴Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Integrative genetic association data are routinely used in post-genome-wide association studies (GWAS) to prioritize functional units that may affect a complex trait for experimental validation. Transcriptome-wide association studies (TWAS) are a popular class of mechanism-aware methods that have identified many causal genes by testing associations between predicted expression and a complex trait. Here, we introduce a novel statistical framework that seamlessly performs multiple tasks using TWAS and colocalization data: causal gene implication and gene set enrichment analysis. First, we propose an evidence integration procedure, INtegration of TWAS And ColocalizaTion (INTACT), that preserves uncertainty from TWAS and colocalization by systematically combining both gene-level results into a probabilistic quantification of "causality". INTACT is designed to be a flexible tool that can use output from any TWAS or colocalization approach. Then, we describe a gene set enrichment estimation method (INTACT-GSE) that utilizes results from INTACT. Using simulations, we observe that INTACT implicates causal genes with higher power than colocalization-only approaches. Additionally, INTACT maintains proper false discovery rate control, in contrast to TWAS-only approaches. We find that INTACT-GSE offers more-accurate enrichment quantification than a widely-used two-stage approach. We then explore the extent to which we can identify trait-relevant genes by explicitly linking genes to the transcriptome. We apply our methods to identify core genes and pathways for 4 GWAS traits for which much of the underlying molecular biology is already known (serum urate, IGF-1, male testosterone, and female testosterone levels). We integrate the GWAS data with GTEx data to generate TWAS and colocalization results. INTACT implicates 14, 35, 7, and 7 core genes for each trait, respectively. We compare the set of core genes implicated by INTACT to the set that is in close proximity to a GWAS hit (within 100 kb). For each trait, we find that some core genes that are in proximity to a GWAS hit are also implicated by INTACT (33%, 37%, 21%, and 46% of core genes in proximity to a GWAS hit for serum urate, IGF-1, male testosterone, and female testosterone, respectively). Although INTACT identifies fewer core genes across traits, there are cases in which it implicates core genes that the proximity-based method does not. These results may suggest gene expression does not have a significant role in some molecular mechanisms underlying complex traits and that our method may complement a proximity-based approach. INTACT-GSE identifies enrichment in 9 pathway-tissue pairs of interest.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3288. A statistical framework for local genetic correlation across populations using GWAS summary data.

Authors:

C. Zhang¹, Y. Zhang², H. Zhao³, Q. Lu⁴, Y. Ye², Y. Shi², K. Liang⁵; ¹Yale Univ., new haven, CT, ²Yale Univ., New Haven, CT, ³Yale Univ. Sch. of Publ. Hlth., New Haven, CT, ⁴Univ. of Wisconsin-Madison, Madison, WI, ⁵Peking Univ., Beijing, China

Abstract Body:

A growing number of studies have discovered that the genetic architecture of complex phenotypes varies across populations, with genetic differences resulting primarily from differences in linkage disequilibrium (LD), minor allele frequencies (MAF), cross-population correlations of causal SNP effects, and overall heritabilities. As the number of genome-wide association studies (GWASs) in non-European groups has been increasing, a number of methods have been developed for evaluating the *global* genetic correlation between populations to shed light on genetic heterogeneity and improve polygenic risk prediction across populations. However, there is a lack of statistical methods to assess shared genetic associations in a *local* genomic region rather than the entire genome between populations. To fill in this gap, we introduce HYPERGNOVA, a method for estimating local genetic correlations across populations in order to better understand etiological mechanisms shared by multiple populations. To expand the global genetic association across populations, HYPERGNOVA uses summary-level data from GWASs to account for specific genomic regions by identifying local, heterogeneous etiologic sharing among diverse populations. We show through simulations that HYPERGNOVA gives an unbiased estimate for both global and local genetic correlation across populations and that it performs comparably to existing approaches at the global level. In addition, we will show our results in discovering population-specific and population-shared genomic regions among 24 common disorders and traits between East Asians and Europeans.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3289. A transcriptomic signature of late-life depression.

Authors:

S. Matan-Lithwick¹, D. A. Bennett², Y. Wang³, D. Felsky¹; ¹Krembil Ctr. for Neuroinformatics, Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, ²Rush Alzheimer's Disease Ctr., Rush Univ., Chicago, IL, ³Rush Univ., Chicago, IL

Abstract Body:

Background: Depression is a neuropsychiatric illness that affects more than one million Canadians annually. The impact of depression is especially clear among the elderly, as it impacts daily function and may be a risk factor for dementia. While genes associated with mid-life depression have been identified, the transcriptomic signature of late-life depression has not been described. We hypothesized that neocortical gene expression would be associated with symptoms of late-life depression proximal to death, independent of pathologic AD diagnosis and medication status.

Methods: Bulk tissue RNA sequencing data from frontal cortex of 1,001 elderly brain donors were analyzed (mean age at death: 89.7yrs [range: 67-108]). All donors had genotype, demographic, and clinical assessments within one year prior to autopsy, as well as detailed postmortem neuropathological characterization. Differential expression analysis was performed with robust linear modeling including technical and demographic covariates, cell type proportions, educational attainment, APOE e4 genotype, and medication status for AD- and depression-prescribed drugs. We then included main effects of depressive symptom burden proximal to death (CESD-sum) and pathologic diagnosis of AD (NIA-Reagan criteria).

Results: After false discovery rate correction (FDR $q < 0.05$), only one gene, Prader Willi/Angelman region RNA1 (PWAR1), was associated with depressive symptoms ($t=5.2$, $q=0.005$); greater PWAR1 abundance was linked to higher symptom burden. At $q < 0.1$, 14 genes showed association. Notably, many of these genes (e.g. CTDSPL2, ACR2B-AS1, ADGRE2, MRM1, IRF8, COL19A1) have been previously implicated in unipolar depression or bipolar disorder. Gene set enrichment analysis revealed upregulation of processes related to energy metabolism (top: "oxidative phosphorylation", $p=2.1 \times 10^{-8}$) and downregulation of DNA modifying processes (top: "DNA conformation change", $p=2.0 \times 10^{-7}$), among others. Results were not changed when controlling for smoking and alcohol consumption, and no significant associations were observed for depression medication status (top $q > 0.23$).

Conclusions: Here we describe a cortical transcriptomic signature of late-life depressive symptoms. Among our top signals are genes with known links to mid-life depression, as well as biological processes relevant to energy metabolism, synaptic plasticity, DNA modification, antigen presentation, and response to pH. Our work provides a depression-specific expression signature for elderly frontal cortex and lays the foundation for investigation of specific mechanisms toward precision therapeutics.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3290. A two-stage mixed effects model to analyze longitudinal zero-inflated microbiome count data

Authors:

J. Wang, C. Reyes-Gibby, S. Shete; U.T. MD Anderson Cancer Ctr, Houston, TX

Abstract Body:

Technological development has provided valuable resources to investigate human microbiome. There is of great interest to study the microbiome longitudinal changes associated with risk factors and clinical outcomes. Such analysis was challenged because of the zero-inflated microbiome data and correlation of the longitudinal data repeatedly collected within the same patient. The current approaches for analyzing longitudinal zero-inflated microbiome abundance data focused on testing the covariate-taxon associations, where the time-varying microbiome abundance is the dependent variable; but ignored the taxon-outcome associations, where the non-time-varying clinical outcome is the dependent variable in the model. In this study, we proposed a two-stage mixed effects model for analyzing the association between zero-inflated longitudinal count data and clinical outcome. Specifically, the longitudinal microbial abundance count is first modeled as a function of time based on the zero-inflated negative binomial mixed effects model; and then the summaries of the temporal patterns are assessed for their associations with the clinical outcome by using the regression models (e.g., linear regression). We conducted simulation studies to show that the two-stage mixed effects model can provide accurate estimations for the regression coefficients of the association between the longitudinal trend of microbial abundance and the outcome. The approach was applied to the study of longitudinal patterns in oral microbiome and oral mucositis in the patients with squamous cell carcinoma of the head and neck.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3291*. A virtual transcriptome approach identifies drug repurposing candidates for Alzheimer's disease.

Authors:

M. Grabowska¹, Q. Feng², W-Q. Wei¹; ¹Dept. of BioMed. Informatics, Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Div. of Clinical Pharmacology, Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

Background: Treatment options for Alzheimer's disease (AD) remain limited, complicated by high preclinical and clinical failure rates of new AD drugs. Drug repurposing offers several advantages over traditional drug development, including higher success rates, shorter development times, lower costs, and increased assurance of drug safety. We propose a virtual transcriptome approach to identify drug repurposing candidates for AD, with clinical validation of identified drug candidates using de-identified EHRs from Vanderbilt University Medical Center (VUMC). **Methods:** We impute virtual transcriptomic signatures for AD from existing GWAS summary statistics for 762,917 individuals (86,531 cases and 676,386 controls) using S-PrediXcan and S-MultiXcan. S-PrediXcan and S-MultiXcan leverage eQTL data from the Genotype-Tissue Expression (GTEx) project to predict tissue-specific and multi-tissue gene expression from GWAS summary statistics, respectively. We construct three AD virtual transcriptomic signatures using: (1) all available GTEx tissues, (2) GTEx brain tissues, and (3) GTEx tissues with demonstrated relevance to AD (brain, whole blood, spleen, and skin). We identify drug repurposing candidates for AD by querying the imputed AD transcriptomic signatures against drug signatures available in the Connectivity Map (CMap) database to find drugs capable of inducing opposing changes in gene expression. We then validate promising repurposing candidates using clinical data in VUMC's de-identified EHR database to determine whether exposure to a repurposing candidate is associated with a lower prevalence of AD in individuals over 60. We use a 2:1 propensity score matching approach with sex, race, and comorbidities (coronary artery disease and rheumatoid arthritis diagnoses at the time of cohort entry, i.e., at age 60) as covariates in defining a control cohort. **Results:** Aspirin was identified as one of the top ten drugs with a negative connectivity score computed by CMap for the AD transcriptomic signature constructed using all GTEx tissues. In the VUMC EHR, we further validated that AD prevalence is significantly lower in individuals exposed to outpatient aspirin compared to matched controls never exposed to aspirin (OR 0.54, 95% CI 0.47- 0.66, $P = 4.08 \times 10^{-12}$). **Conclusions:** Our findings suggest that aspirin exposure may be associated with lower occurrence of AD in individuals over the age of 60. We plan to expand our clinical validation analyses to investigate other drug repurposing candidates suggested by our virtual transcriptome approach to identify high-priority repurposing candidates for further investigation in clinical trials.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3292*. Accurate and cost-effective imputation of genotypes at whole genome level with Sparse Denoising Autoencoders.

Authors:

S. Loguercio¹, R. Dias², D. M. Rogers³, D. Evans¹, S-F. Chen¹, K-Y. Chen¹, L. Chan¹, A. Torkamani¹; ¹Scripps Res. Translational Inst., La Jolla, CA, ²Univ. of Florida, Gainesville, FL, ³Oak Ridge Natl. Lab., Oak Ridge, TN

Abstract Body:

Genotype imputation is a key component of genetic association studies, where it increases power, facilitates meta-analysis, and aids interpretation of signals.

Current statistical imputation methods require access to high-performance computing environments and large, privacy-sensitive, WGS reference panels. The most common strategy to access state-of-the-art genomic imputation is to submit genotype data to imputation servers, resulting in privacy and scalability concerns. Moreover, the accuracy of current statistical techniques is known to degrade in regions of low and complex linkage disequilibrium.

Artificial neural network-based imputation approaches may overcome these limitations by encoding complex genotype relationships in easily portable inference models. Recently, we demonstrated a general strategy for unphased genotype imputation with sparse denoising autoencoders (DSAE), using the HRC reference panel and limited to human chromosome 22. Our DSAE-based imputation strategy achieved superior accuracy across the allele frequency spectrum and across genomes of diverse ancestry, while delivering at least 4-fold faster inference run time relative to standard imputation tools - and without having to transfer private genotype data to external servers.

Here we extend the pilot study on chr22 to the entire human genome, using the larger and more diverse TOPmed reference panel (97K reference samples) and phased imputation.

As autoencoder training is computationally intensive - shifting the computational burden to model trainers, with performance gains for end-users - we leverage the massive GPU resources of a large supercomputing facility (OLCF Summit HPC) for hyperparameter optimization, training and fine tuning of several thousands of autoencoder models in parallel - one for each genomic segment, with boundaries defined by recombination.

We present preliminary results assessing the performance of our fully trained autoencoders in comparison with modern imputation methods (Minimac4, Beagle5, Impute5) across a range of independent datasets, genotyping array marker sets, minor allele frequency spectra, and diverse ancestry groups.

Moreover, since no minimum training size standards have been determined for deep learning-based imputation in genomics, we also report a comparison of the training accuracy results for imputation with HRC and TOPMED reference panels.

Our work provides an efficient genotype imputation platform at whole genome scale with substantial speed, accuracy and privacy benefits for genome association studies and clinical applications in precision medicine.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3293. Accurate and efficient estimation of local heritability using summary statistics and linkage disequilibrium

Authors:

H. Li¹, R. Mazumder², X. Lin¹; ¹Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ²Massachusetts Inst. of Technology, Boston, MA

Abstract Body:

Heritability, defined as the proportion of variability in the phenotype that is attributable to genetic factors, is a fundamental and important parameter in statistical genetics. Many existing heritability estimation methods have been developed in the framework of the restricted maximum likelihood (REML) method under linear mixed models using data from genome-wide association studies (GWAS), but the requirement to access individual-level genotypic and phenotypic data greatly limits the applicability of these methods. Several have been proposed in the last few years to circumvent the need to use individual-level data by using GWAS summary-statistics to estimate heritability, such as LDSC and GRE. However, these approaches produce less efficient estimators the precision of which is less than individual-level data based REML heritability estimate. We introduce a novel REML-score equation based estimating procedure that yields highly statistically efficient heritability estimators similar to REML, but only requires GWAS summary-statistics and the in-sample linkage disequilibrium (LD) information. The efficiency of our estimator is guaranteed as the estimating equations coincide with the REML score equations constructed using individual-level data. The relative efficiency (RE) of our estimator compared to REML is as high as 92% while other state-of-the-art summary-statistics-based methods, such as LDSC and GRE, have a RE lower than 25%. Furthermore, we propose a sparse representation of the LD matrix using a sum of a low rank and a banded matrix, which not only reduces the storage cost and thus improves the portability of the LD matrix, but also increases the computational efficiency of our estimating algorithm. We demonstrate the statistical efficiency of our estimator and the advantages of the proposed sparse representation of the LD matrix both theoretically and using extensive simulations. We apply our method to the summary-level data from the UK biobank to estimate the local heritabilities of 22 common traits and diseases. The increased precision of heritability estimates based on our algorithm will also enable other more powerful summary-statistics-based analyses, such as the estimation of genetic correlation and the construction of polygenic risk scores.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3294. Accurate *in silico* confirmation of rare copy number variant calls from exome sequencing data using transfer learning.

Authors:

R. Tan¹, Y. Shen^{1,2,3}; ¹Dept. of Systems Biology, Columbia Univ., New York, NY, ²Dept. of BioMed. Informatics, Columbia Univ., New York, NY, ³JP Sulzberger Columbia Genome Ctr., Columbia Univ., New York, NY

Abstract Body:

Exome sequencing is widely used in genetic studies of human diseases and clinical genetic diagnosis. Accurate detection of copy number variants (CNVs) is important to fully utilize exome sequencing data. However, exome data is noisy. None of the existing methods alone can achieve both high precision and recall rate. A common practice is to perform heuristic filtration followed by manual inspection of read depth of putative CNVs. This approach does not scale in large studies. To address this issue, we developed a transfer learning method, CNV-espresso, for *in silico* confirming rare CNVs from exome sequencing data. CNV-espresso encodes candidate CNVs from exome data as images and uses pre-trained convolutional neural network models to classify copy number states. We trained CNV-espresso using an offspring-parents trio exome sequencing dataset, with inherited CNVs as positives and CNVs with Mendelian errors as negatives. We evaluated the performance using additional samples that have both exome and whole-genome sequencing (WGS) data. Assuming the CNVs detected from WGS data as proxy of ground truth, CNV-espresso significantly improves precision while keeping recall almost intact, especially for CNVs that span small number of exons. CNV-espresso can effectively replace manual inspection of CNVs in large-scale exome sequencing studies.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3295. Accurate variant calling of the camouflaged LPA KIV-2 VNTR in whole exome sequencing data

Authors:

S. Schoenherr, S. Di Maio, P. Zöschner, L. Forer, F. Kronenberg, S. Coassin; Med. Univ. of Innsbruck, Innsbruck, Austria

Abstract Body:

Background: Recent studies showed that some of the strongest variant signals for phenotypes can be found in protein-coding variable number tandem repeats (VNTRs). The *LPA* gene locus is a prominent example: It (a) controls >90% of Lipoprotein(a) [Lp(a)], which is a strong genetic risk factor for cardiovascular disease variance and (b) contains a 5.6-kb-large protein-coding VNTR with 10 to 40 Kringle IV-2 (KIV-2) repeats, resulting in >40 isoforms of the apolipoprotein(a) [apo(a)]. Up to 70% of the *LPA* coding sequence can be within the KIV-2 VNTR. The number of the KIV-2 repeats inversely correlates with the Lp(a) concentrations and mutations hidden in the KIV-2 region have been shown to explain a considerable amount of Lp(a) variance. However, mutation detection is difficult as the KIV-2 VNTR is camouflaged in current reference databases by several other highly homologous kringle domains in *LPA*, which lead to paralogous sequence variants, if not properly addressed during read alignment and variant calling. In previous work, we developed a targeted amplicon-sequencing approach for the KIV-2 region and reported a first map of the high genetic variability in the KIV-2 region, including SNPs with very strong effect on the Lp(a) variance.

Results: In the work at hand, we expanded our approach to accurately call variants in the KIV-2 region in whole exome sequencing (WES) data to maximize the utility of publicly available datasets. We evaluated its performance by comparing variants called from WES data to a gold-standard variant calling set derived from a targeted PCR-based sequencing approach, which selectively amplifies only the KIV-2 region and is thus not subject to paralogous variants originating from other kringle domains (8 samples, thereof 2 samples with no KIV-2B subtype, resembling the KIV-2B population frequency). We show that the selected read extraction strategy from WES is critical for the overall performance (F1 score range between 0.40-0.96) and also dependent on the KIV-2 subtype (i.e. specific haplotypes). The overall performance of our approach (F1-Score of 0.96, sensitivity of 0.91) improved by 1.4 to 2.4-fold compared to previous published strategies.

Conclusions: Our approach enables highly sensitive and specific mutation screening in the large *LPA* KIV-2 VNTR and is customizable to other medically relevant camouflaged genes.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3296. Adjusting for genetic confounders leads to reliable detection of causal genes from transcriptome-wide association studies

Authors:

W. Crouse¹, S. Zhao², K. Luo¹, S. Qian¹, M. Stephens¹, X. He¹; ¹Univ. of Chicago, Chicago, IL, ²Dartmouth Coll., Hanover, NH

Abstract Body:

While GWAS have identified many loci associated with complex traits, causal genes in these loci often remain unknown. A number of methods have been developed to leverage expression QTL data to nominate candidate genes, including Transcriptome-wide association studies (TWAS), colocalization analysis, and Mendelian Randomization (MR) methods that use cis-eQTLs as instrumental variables. All these methods, however, suffer from various issues, particularly false positive findings. TWAS may find associations in non-causal genes when their eQTLs are shared with nearby causal genes or are in linkage disequilibrium (LD) with nearby causal variants. Colocalization methods typically assume a single causal variant for both gene and trait, are sensitive to parameters, and still cannot guarantee causality. For MR methods, the key instrumental variable assumption is often violated due to the pleiotropic effects of genetic instruments. This problem is exacerbated by the small number of genetic instruments and their correlations due to LD.

The fundamental problem of existing methods is that, when assessing the role of one gene on the phenotype, nearby variants and nearby genetic components of expression can be correlated with the eQTL(s) of the test gene while also affecting the phenotype directly, thus acting as “genetic confounders”. This motivates our approach, causal TWAS (cTWAS), which jointly models the effects of all nearby gene expression traits and nearby variants, which controls for potential genetic confounders. To make the model identifiable, we use a Bayesian variable selection strategy that learns parameters from genome-wide data. In simulations using data from UK Biobank (UKBB), existing methods (e.g. Fusion, SMR, coloc, FOCUS) all suffered from high false positive (FP) rates. In contrast, cTWAS produced calibrated FP rates while maintaining power.

We applied cTWAS to UKBB GWAS summary statistics of low density lipoprotein (LDL) cholesterol using eQTL from liver tissue. Genes identified by cTWAS are enriched for cholesterol-related gene sets, and cTWAS demonstrates high precision in distinguishing known genes regulating LDL cholesterol from nearby “bystander” genes. cTWAS also identified novel genes and pathways related to LDL cholesterol for further study. In applying cTWAS more broadly, using all tissues from GTEx, to complex traits such as Crohn’s Disease and Blood Pressure, cTWAS identified hundreds of putative causal genes. In conclusion, cTWAS solves a fundamental challenge of eQTL-GWAS analysis, reducing false positives relative to competing methods. cTWAS is available in an efficient and easy-to-use R package.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3297. Admixture mapping for phenome-wide discovery in a diverse health system biobank

Authors:

S. Cullina^{1,2}, R. Shemirani^{1,2}, B. Rowan^{1,2}, N. Abul-Husn^{3,2,4}, S. Asgari^{1,2,4}, E. Kenny^{4,3,2}; ¹Icahn Sch. of Med. at Mt. Sinai, New York, NY, ²Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, ³Dept. of Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, ⁴Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Admixed individuals, who's genomes reflect recent ancestry from two or more continents, comprise over one third of the U.S. population. Despite many common heritable diseases being enriched in admixed populations, most genomic discovery approaches are tailored to more homogenous populations. Admixture mapping is a method that utilizes differences in disease prevalence and allele frequency between ancestral populations of admixed individuals for association testing. We have developed a generalizable pipeline and best practices by which local ancestry (LA) haplotypes can be inferred and utilized for admixture mapping in a diverse biobank in New York City (NYC). Individual-level ancestry proportions were calculated using genotype array data for 53,900 individuals in the BioMe Biobank at Mount Sinai, NYC. Participants were assigned to groups based on self-report race/ethnicity and patterns of recent admixture, enabling definition of population groups and optimizing the selection of global reference panels. Genotype data was phased and RFMIX2 was used to perform two- (African (AFR) and European (EUR) ancestry) or three-way (AFR, EUR, and Native American (NAT) ancestry) LA inference, for defined African/African-American (AA) and Hispanic/Latino(a) (HL) groups, respectively. Admixture mapping is currently being performed using generalized linear models in which LA haplotypes are tested for association with 1,740 phecodes from electronic health records as well as a curated set of over ~200 biomarkers and other quantitative traits. To assess the quality of LA calls, we compare the R2 correlation between the sum of LA haplotypes per individual and their global admixture proportions. We observe strong correlation in both HL (R2: AFR 1, EUR 0.96, NAT 0.96) and AA (R2: AFR 1, EUR 0.99). Variance components using LA haplotypes are being calculated phenome-wide using the HAMSTA method. Population-specific phenome-wide significance thresholds are being estimated using the STEAM method which accounts for population composition and history. We will present results from the association, annotation and finemapping of LA haplotypes, replicating known and showcasing novel associations. Large efforts to increase diversity and representation in genomic research mean that biobanks will become increasingly admixed, motivating the development of generalizable pipelines like this for well-calibrated admixture mapping.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3298. Allele frequency differences of causal variants explain the bulk of low portability of cross-ancestry PRS

Authors:

X. Liu¹, M. Saitou², A. Dahl¹, Q. Wang³; ¹Univ. of Chicago, Chicago, IL, ²Norwegian Univ. of Life Sci., Ås, Norway, ³Osaka Univ., Osaka, Japan

Abstract Body:

Genome-wide association studies (GWAS) are overwhelmingly biased toward European ancestries. Nearly all existing studies agree that transferring genetic predictions from European ancestries to others results in a substantial loss of accuracy. This is known as low portability of polygenic risk scores (PRS) and is one of the most important barriers to clinical deployment of PRS. Yet, it remains unclear how much different genetic factors, such as linkage disequilibrium (LD) differences, allele frequency differences or causal effect differences across ancestries contribute to low portability. In this study, we used gene expression levels in lymphoblastoid cell lines as a simplified model of complex traits in order to dissect how much each genetic factor contributes to PRS portability in a setting with minimal nongenetic differences. In stark contrast to the very low *cis*-genetic correlation estimates between European and African ancestries from previous studies, we found that *cis*-genetic effects between European and Yoruban have nearly perfect genetic correlation ($r=95\%$, S.E.=9%). We demonstrated through theory and realistic simulations that the previously reported low genetic correlations were artifacts of statistical bias. We derived a mathematical model to tease apart the genetic factors that may lead to loss of portability. We found that cross-ancestry LD differences and population-specific effects together explain a small proportion ($<10\%$) of loss of portability. In contrast, allele frequency differences of causal variants have a striking impact on portability. For example, we observed a reduction of more than 40% in portability when the causal *cis*-variant is common in European (training population) but rare in Yoruban (prediction population). Importantly, this affects a sizable fraction (15.6%) of genes on the genome. We conclude that genetic effects are highly similar across ancestries, but that many causal variants with allele frequency differences severely decrease PRS portability. Our study thus provides evidence that large GWAS in diverse non-European populations are required for accurate PRS prediction in non-European population and, thus, for equitable clinical use of PRS.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3299. Allelic imbalance of chromatin accessibility in cancer identifies candidate causal risk variants and their mechanisms.

Authors:

D. Grishin¹, A. Gusev²; ¹Dana-Farber Cancer Inst., Cambridge, MA, ²Dana-Farber Cancer Inst., Boston, MA

Abstract Body:

Understanding the functional impact of non-coding variants remains a major challenge in cancer genetics. In the case of non-coding germline risk associations from GWAS, many associations are in high LD and cannot be resolved by statistical fine-mapping. In the case of non-coding somatic drivers from tumor sequencing, existing datasets are underpowered for discovery with frequency-based methods. For both domains, methods that can infer the functional impact of individual variants within regulatory elements are thus urgently needed. Here we studied allelic imbalance of chromatin accessibility in 409 ATAC-Seq samples across 23 cancer types. To this end we previously developed stratAS, a statistical model of allelic imbalance that aggregates signals across individuals while modeling individual-level somatic copy number variation. We discovered 7,262 germline allele-specific accessibility QTLs (as-aQTLs) and found that they are highly enriched for cancer risk heritability across seven common cancer GWAS (e.g. prostate cancer as-aQTLs with a 145 ± 35.7 ($p=6.3 \times 10^{-5}$) fold enrichment for prostate cancer risk) and are caused by genetic variants that alter transcription factor binding and gene expression. To connect as-aQTLs to putative risk mechanisms, we introduced the Regulome-Wide Associations Study (RWAS). RWAS identified accessible peaks genetically associated with cancer risk at >70% of known breast and prostate loci (compared to <45% for a conventional Transcriptome-Wide Association Study) and discovered novel risk loci in all examined cancer types. Motivated by these findings, we developed siamAS, a predictor of allelic imbalance that uses a novel “Siamese” neural network approach that trains two mirrored networks on allele-specific features. We trained siamAS on our allele specificity data incorporating >7,000 features from multiple variant effect predictors. In held out data, siamAS achieved a classification AUC of 0.92, which substantially outperformed any individual predictive feature (e.g. CADD score with AUC of 0.55). We applied siamAS to variants within regulatory elements and demonstrated a significant enrichment in cancer risk heritability for variants classified by our model as functional (e.g. 42.5 ± 20.8 fold breast cancer risk heritability enrichment at top as-aQTLs predicted by siamAS). Additionally, we applied siamAS to non-coding somatic mutations in TCGA and showed statistically significant classification accuracy (AUC > 0.9 for top 10 most imbalance/balanced somatic mutations). In summary, our results establish cancer as-aQTLs, RWAS and siamAS as powerful tools to study the genetic architecture of cancer risk.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3300. Alpha-2A adrenergic receptor (*ADRA2A*) modulates susceptibility to Raynaud's disease.

Authors:

A. Tervi, V. Lammi, FinnGen, C. A. Heckman, S. E. Jones, H. M. Ollila; Inst. for Molecular Med. Finland, FIMM, HiLIFE, Univ. of Helsinki, Helsinki, Finland

Abstract Body:

The autonomic nervous system controls physiological functions in the body that are not under direct voluntary control and are not typically consciously directed. The targets of the autonomic nervous system include, for example, body temperature, heart rate, respiration, blood pressure regulation, and vascular tone. In some instances, however, the autonomic nervous system can malfunction causing symptoms and diseases of dysautonomia and affecting many different targets of the autonomic nervous system at once. Raynaud's disease is dysautonomia where exposure to cold or stress increases the vascular tone of distal arteries causing vasoconstriction and hypoxia, particularly in fingers and toes. Using genetic and electronic health record data from the UK Biobank and FinnGen data freeze 9 we identified 6,940 individuals with a diagnosis for Raynaud's disease or Raynaud's phenomenon and 816,332 disease-free controls. Genetic analysis identified the same risk locus at the *ADRA2A* gene region independently in both cohorts. A meta-analysis combining UK Biobank and FinnGen identified rs7090046 as the lead variant associated with Raynaud's disease ($P=1.1e-30$). Functional analysis using RNA expression from GTEx indicated that the genetic variants modulate *ADRA2A* expression in a tissue-specific manner in the distal arteries (rs7090046, $P=1.3e-13$, effect size (NES) = 0.305). Our results indicate that *ADRA2A* modulates vascular tone in Raynaud's disease and provides the first functional evidence for understanding the mechanisms in dysautonomia.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3301. AlphaCluster: Coevolutionary driven residue-residue interaction models enable quantifiable clustering analysis of *de novo* variants to enhance predictions of pathogenicity

Authors:

J. Obiajulu, R. Kuang, G. Zhong, W. K. Chung, Y. Shen; Columbia Univ., New York, NY

Abstract Body:

Missense variants have highly variable effects and effect size, which often makes it challenging to distinguish pathogenic and non-pathogenic variants and subsequently implicate new genes for disease association in studies. Importantly, missense variants can be the sole molecular mechanism for some genetic disorders, and so statistical approaches tailored for the analysis of missense variants are critical. Analysis of the clustering of missense variants is a promising approach which leverages the fact that missense variants in protein domains often have similar effects on function. Here we describe a new clustering analysis approach, AlphaCluster, a statistical method which quantifiably analyzes the spatial clustering of *de novo* variants by mapping missense residues onto the protein tertiary structure. We show that our approach can quantify the evidence supporting pathogenic missense variants and increase the power to detect clustering when compared to available genomic clustering tools. Using AlphaCluster, we identify two genes newly implicated at the genome-wide level in autism spectrum disorder, namely *GABBR2* and *SATB2*, and eleven in neurodevelopmental disorders (NDD). We also apply AlphaCluster to protein complexes and detect an association between the gamma aminobutyric acid receptor complex (GABA-A α ;1 β ;2 γ ;2 receptor) and Kv1.2 potassium channel with NDD.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3302. An accelerated, massively scalable variant calling and filtering workflow for the detection of de novo mutations in whole genome sequencing data.

Authors:

E. Dawson^{1,2}, M. Yeager³, M. Dean², D. Karyadi², J. Israeli¹; ¹Nvidia Corp., Santa Clara, CA, ²NCI, Rockville, MD, ³FNLRC/NCI, Rockville, MD

Abstract Body:

De novo mutations are important contributors to human disease and variation. Such variants are usually detected using trio studies, in which an offspring sample and both parents are analyzed in tandem. The three samples are each aligned, variants are called, and then filtered to identify Mendelian inheritance violations and putative de novo mutations (DNMs). Trio studies involve significant computational resources because of the need to analyze all three samples, and often there is need to do them quickly to identify the variant causing the disease in the child. We developed a pipeline implemented in the Workflow Description Language for accelerated analysis of trio data. NVIDIA Clara Parabricks-accelerated versions of DeepVariant and GATK HaplotypeCaller are run concurrently with Strelka2. glNexus is then used for joint genotyping of the HaplotypeCaller and DeepVariant calls. A series of filtering steps, including taking the multi-caller intersection, is performed to generate putative de novo mutations. The number of putative DNMs output by our pipeline is in line with other studies, and manual review of variants called in Genome in a Bottle trios indicated high precision among variants called by all three callers. The native task parallelism of WDL allows our pipeline to run independent tasks within a single trio in parallel. We extended the scalability of the standard Cromwell WDL runner by automating WDL input creation and Cromwell instantiation based on comma-delimited manifests of samples. This additional layer of sample-level parallelism enables scaling further across SLURM clusters or the cloud with minimal setup. The integration of accelerated tools reduces the runtime per trio compared to community versions of the tools, which in turn reduces the computational burden of de novo mutation calling in trios and improves scalability and ease of use.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3303. An accurate and efficient causal gene network inference method that handles many confounding variables.

Authors:

J. Kvamme, A. Fu; Univ. of Idaho, Moscow, ID

Abstract Body:

Learning complex regulatory relationships among genes has been a challenge in biology. Several methods have been developed to conduct causal inference for molecular phenotypes (e.g., gene expression and DNA methylation), using genetic variants as instrumental variables under the principle of Mendelian randomization. However, these methods are often limited to mediation and ignore other possible regulatory relationships, or do not account for confounding variables well. Here, we introduce a new causal network inference method for trios consisting of a genetic variant (denoted by V) and two associated molecular phenotypes (denoted by X and Y). Our method, Mendelian Randomization Genomic Network (MRGN), improves existing methods by i) inferring diverse regulatory relationships for a trio; ii) allowing for the inclusion of many confounding variables; and iii) eliminating the need for a large set of independence tests to infer each causal edge.

Specifically, MRGN performs conditional and marginal tests to detect five basic models: mediation ($V \rightarrow X \rightarrow Y$), v -structure ($V \rightarrow X \leftarrow Y$), conditional independence ($X \leftarrow V \rightarrow Y$), fully connected ($X \leftarrow V \rightarrow Y$ and $X - Y$), and the null model ($V \rightarrow X$; no relationship between X and Y). The conditional tests regress X (or Y) on Y (or X), V , and the confounding variables. Marginal tests are used to test for significant correlation between V and X (or Y). Additionally, if V has a rare variant, MRGN performs a permutation test like that in the GMAC method (Yang et. al., 2017) for improved calculation of the p -values for the regression coefficients in the conditional tests.

We test the performance of MRGN by simulating trios from the five basic models with confounding variables. We investigate four simulation parameters: strength of the effect, noise in the residuals, minor allele frequency, and the number of confounders. We compare our method with GMAC, a method designed to detect only mediation, and MRPC (Badsha and Fu, 2019), our other causal inference method designed for generic network inference. For each trio, confounders are identified from a candidate pool before causal inference. Across different scenarios, MRGN correctly detects substantially more edges than MRPC while incurring a few additional false positive edges (recall 0.944 vs 0.141; precision 0.832 vs 0.989, respectively). Additionally, while having the added flexibility to infer the entire network, MRGN also performs comparably with GMAC in detecting mediation (recall 0.918 vs 0.967; precision 0.789 vs 0.754, respectively). Lastly, our method is nearly 15 times faster than GMAC and 8 times faster than MRPC (0.35s, 5.18s, and 3.04s per trio, respectively).

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3304. An allelic series rare variant association test for candidate gene discovery

Authors:

Z. McCaw, B. Ungen, F. Casale, T. Karaletsos, D. Koller, T. Soare; Insitro, South San Francisco, CA

Abstract Body:

A central task in genetics and in drug discovery is identifying genes that influence a molecular trait or disease process. In genome-wide association studies, common variants (minor allele frequency [MAF] > 1%) are typically linked to the phenotype of interest using variant-level association tests. Rare variants (MAF ≤ 1%), by contrast, are typically analyzed using gene-centric approaches, particularly the burden and sequence kernel association tests (SKAT). We sought to develop a gene-based rare variant association test that targets allelic series: cases where increasingly deleterious alterations of a gene lead to increasingly large effects on the phenotype. The proposed allelic series test operates on rare coding variants. Using the Ensembl Variant Effect Predictor (VEP), these variants are annotated as synonymous, benign missense, deleterious missense, or protein truncating. Synonymous variants and those lacking annotations are removed. Gene-level allelic series scores are constructed by aggregating indicators or counts of benign missense, deleterious missense, or protein truncating variants across a gene. The indicators or counts are scaled by a weight that increases with predicted pathogenicity, reflecting the hypothesis that more deleterious alterations will have larger effects. Different methods of weighting and aggregation lead to different allelic series tests. These are combined into an omnibus test using the Cauchy aggregation method.

Through extensive simulations, we show that the proposed allelic series test maintains the type I error in the absence of association, and improves power when more deleterious variants in fact have larger effect sizes. We apply the allelic series test to perform rare variant analysis of lipid traits (total cholesterol, low and high density lipoprotein, triglycerides) using whole exome sequencing data on 145,735 subjects from the UK Biobank. We compare the proposed allelic series test with standard rare variant association tests, including burden and SKAT, replicating known associations and identifying novel candidates.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3305. An alternative approach to the estimation of biological age

Authors:

K. Fischer¹, N. Taba¹, M. Mändul¹, A. Kolde¹, R. Mägi²; ¹Univ. of Tartu, Tartu, Estonia, ²Inst. of Genomics, Univ Tartu, Tartu, Estonia

Abstract Body:

Given a set of age-related biomarkers, the biological age of an individual is often defined as the average chronological age of all individuals with a similar biomarker profile. Being “biologically older” than one’s chronological age is seen as an indicator of a worsened health status, leading to higher disease risks and shorter-than-average residual life expectancy. Conventionally, various regression approaches are used to estimate biological age based on a model with chronological age as dependent variable and omics markers as covariates. We discuss the plausible underlying causal association structures and study, whether, and under which assumptions the difference between biological and chronological age estimated by conventional methods could be a meaningful summary of the individual’s risk profile. Using the metabolomics profiles for the Estonian Biobank participants, we show that the conventional biological age estimates can easily be affected by time-dependent confounders, that can lead to biased estimates of actual risk level as well as invalid conclusions on the effects of risk factors. We propose an alternative definition and estimation strategy for the biological age, using the concepts of survival analysis. The new estimate is directly related to the underlying risk level and is therefore easily interpretable. Using simulations as well as the example of the Estonian Biobank, we compare our parametric and semiparametric approaches to the biological age estimation with the conventional method. We also show how our approach can be used in the estimation of “heart age”, as a tool for personalized feedback on cardiovascular disease risks. We conclude that when there are time-dependent confounders associated with the biomarker profile, younger or older “biological age”, when estimated as a regression prediction for age, does not necessarily correspond to lower or higher disease or mortality risks. The proposed alternative approach would lead to estimates that are more straightforward to interpret and can be used while communicating the omics-based personalized risk estimates to individuals.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3306. An evaluation of ancestry specific polygenic risk scores (PRS) for uterine fibroids across populations

Authors:

J. Piekos¹, J. N. Hellwege², O. Dikilitas³, I. J. Kullo³, D. J. Schaid³, T. L. Edwards², D. R. Velez Edwards⁴; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Mayo Clinic, Rochester, MN, ⁴Vanderbilt Univ Med Ctr., Nashville, TN

Abstract Body:

Uterine leiomyomata (or fibroid, UF) are common (prevalence up to 75% in women), heritable (approximately 35%), and can negatively interfere with a woman's life if symptomatic and lead to surgical interventions such as hysterectomy or myomectomy. We constructed polygenic risk scores (PRS) using PRScsx and genome wide association summary statistics from FinnGen (FG) and Biobank Japan (BBJ). We computed three scores with PRScsx according to genetic ancestry of source data: European and East Asian, respectively, and a meta score. We applied each PRS to women from the Electronic Medical Records and Genomics Network (eMERGE) (N = 25,973) and BioVU (N=9,759) and performed analysis in three strata of women: cross ancestry (CA) including all regardless of race/ancestry, European ancestry (EA), and African ancestry (AA). Using 10-fold cross validation of logistic regression models, (UF ~ PRS + 10 principal components [PC]) we validated all three PRS as significant predictors of UF status in both datasets for EA and CA populations ($P < 0.00185$, multiple testing correction). Odds Ratios (OR) (95% Confidence Interval [CI]) of FG PRS from the meta-analysis were 3.40 (2.41 - 4.80) in EA and 1.98 (1.49 - 2.64) in CA and BBJ PRS was 2.45 (1.93 - 3.11) in EA and 2.33 (1.90 - 2.87) in CA. No PRS significantly predicted UF status in either AA population. Meta-analysis of area under the receiving operator curve (AUROC) found AUROC (95% CI) for FG PRS in EA to be 0.58 (0.55 - 0.62) and 0.61 (0.58 - 0.64) in CA, BBJ PRS 0.59 (0.56 - 0.63) in EA and 0.63 (0.60 - 0.66) in CA, and meta PRS 0.58 (0.55 - 0.62) in EA and 0.62 (0.59 - 0.65) in CA. Cross validation revealed each PRS had more predictive power when applied to CA compared to EA but PRS has larger effect sizes in EA. PRS was then used as the predictor for phenome wide association studies (PheWAS), adjusted for age, BMI, and 10 PC, in each dataset separately and meta-analyzed. FG PRS significantly associated ($P < 2.78 \times 10^{-5}$) with UF phecode (218.1) when applied to BioVU EA and CA and remained significant in CA meta-analysis. BBJ and meta PRS associated with UF phecode in eMERGE EA and CA and all remained significant in meta-analysis. No PRS significantly associated with UF in any AA populations. Additionally, we observed the phecode "excessive or frequent menstruation" to be moderately associated ($P < 2.78 \times 10^{-4}$) with the BBJ PRS in eMERGE CA and both CA and EA in the meta-analysis. Ovarian cysts were moderately associated with the FG PRS when applied to BioVU CA and EA and remains marginally significant in the CA meta-analysis. We observe ancestry specific PRS perform as well as meta-score, however PRS performs better when applied to CA populations as opposed to a single EA stratum.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3307. An Exomiser-based R pipeline for gene burden testing aids novel gene discovery in rare Mendelian diseases.

Authors:

V. Cipriani, L. Vestito, J. Jacobsen, D. Smedley; Queen Mary Univ. of London, William Harvey Res. Inst., London, United Kingdom

Abstract Body:

Recent efforts, such as the 100,000 Genomes Project (100KGP) in the UK, have paved the way for the availability of large sequencing datasets in rare Mendelian diseases. The scale of such datasets now allows for the application of case-control association methods that have been more typically used to identify gene susceptibility in complex disorders, e.g. rare variant gene-based burden testing for novel Mendelian disease gene discovery. Yet, analytical tools for those analyses at scale are not largely available. In a recently published preliminary report of the rare disease data component of the 100KGP (27,591 rare disease families, a total of affected and unaffected individuals of 57,002 and over 197 different rare disorders available at the time of the analysis), we have developed an Exomiser-based analytical framework for gene burden testing. Exomiser is a phenotype-aware (Human Phenotype Ontology, HPO-based) variant prioritization software for sequencing data that has been largely adopted in the scientific community for the search of Mendelian causative variants. This analysis led to the identification of 3 new disease-gene discoveries that have been recently independently confirmed and 19 new associations. We have now extended our Exomiser-based analytical framework into an R pipeline for gene-burden testing and novel gene discovery in rare Mendelian diseases, beyond the application to the 100KGP data. A generic user with access to any large-scale sequencing datasets is now able, first, to apply Exomiser on any available single samples and/or single families for rare variant filtering and annotation and, after defining tailored case and control sets within the dataset, perform a set of gene burden analyses, either genome-wide or restricted to a gene list of interest, including testing for excess in cases of predicted loss-of-function or to be highly pathogenic variants. The R pipeline is complemented with visualization scripts to produce volcano, Manhattan, variant lollipop and HPO-based plots. By making our analytical framework suitable for generic use beyond the original application to the 100KGP data and available as a GitHub repository, we expect to aid substantially novel gene discovery in rare Mendelian diseases.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3308. An experimentally-based functional gene embedding

Authors:

F. Brechtmann¹, T. Bechtler¹, B. Clarke², C. Mertes¹, O. Stegle², J. Gagneur¹; ¹Technical Univ. of Munich, Garching, Germany, ²German Cancer Res. Ctr. (DKFZ), Heidelberg, Germany

Abstract Body:

Gene embeddings, i.e. numerical representations of gene function, are of high relevance for modeling in genomics. Here we propose a generic gene embedding with distinctive features: Our embedding is based on multiple experimental data modalities excluding text-mining sources, to prevent ascertainment bias towards well-studied genes. Specifically, we integrate gene expression across human tissues (GTEx), protein-protein interactions extracted from STRING, and two recently published datasets: a genome-wide deletion screen (DepMap), and a functional gene embedding derived from an autoencoder trained on hundreds of millions of protein sequences (Elnaggar et al.).

Using our embedding, we first predict curated trait-gene associations. For 9 out of 10 traits, our embedding performed on par or better than using an embedding of the STRING network. Additionally, in 4 out of the 10 traits, our joint embedding outperformed any input modality considered individually. Second, we asked whether our 900-dimensional embedding can predict gene-level GWAS signals. To this end, we trained a gradient boosted tree to predict gene-aggregated GWAS signals obtained using MAGMA (de Leeuw et al.). We obtained a median R² of 6.6% across 25 blood biomarkers. This performance was on par with a dedicated study (Weeks et al.) which was using a 57,543-dimensional feature matrix. As a last task, we considered predicting cancer driver genes and improved the mean precision by 62% in predicting COSMIC cancer driver genes over OncoVar. For all these prediction tasks, we used a single embedding demonstrating its generality.

Overall, our embedding captures different aspects of gene functions and can be easily integrated into diverse prediction tasks that can benefit from a general-purpose gene embedding.

de Leeuw, Christiaan A., et al. "MAGMA: Generalized Gene-Set Analysis of GWAS Data." PLOS Computational Biology, 2015

Elnaggar, Ahmed, et al. "ProtTrans: Towards Cracking the Language of Lifes Code Through Self-Supervised Deep Learning and High Performance Computing." IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021

Weeks, Elle M., et al. Leveraging Polygenic Enrichments of Gene Features to Predict Genes Underlying Complex Traits and Diseases. medRxiv, 2020

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3309. Analysis of follow up data in large biobank cohorts: a review of methodology

Authors:

A. Kolde¹, M. Mändul¹, M. Käärrik¹, P. Joshi², K. Fischer¹; ¹Univ. of Tartu, Tartu, Estonia, ²Univ Edinburgh, Edinburgh, United Kingdom

Abstract Body:

Survival analysis is a field of statistics with a long history and well researched in terms of model-fitting in the context of (randomized) clinical trials. Since large omics-based biobanks utilize different study design than clinical trials, it is necessary to investigate whether a survival model needs to be modified in order to assure validity of the results. Additionally one needs to take into account that the classical Cox PH model is computationally prohibitive and that might be a major shortcoming in GWAS setting.

First, we aim to investigate applicability of Cox PH model for biobank setting by choice of different timescales. We were interested in potential bias of different models as well as ways to improve power when high censoring is present. One of the earlier proposed solutions has been the usage of parental lifespan and offspring genotype instead of e data for participants. So far it has been known that this approach increases power, but potential bias has not yet been investigated. Second, we aim to investigate bounds and limitations of the two step martingale residual (MR) based approach for Cox modeling for high-dimensional omics data as a possible way to reduce computational burden. With the MR approach, the analysis simplifies to fitting only one Cox PH model for technical covariates after which all the association testing in GWAS is reduced to a simple linear regression task, while still providing the hazard ratio (HR) estimates.

For small HRs, we show that the choice of timescale does not have much effect on precision, but for bigger HRs it is important to take left-truncation into account to decrease bias. However, if the aim is to maximize the power for association discovery rather than minimizing biases, using the participant's age as the time scale (ignoring left-truncation) yields the highest power. We also propose to use the two step martingale residuals based approach in GWAS settings, showing acceptable precision and no loss of power in case of small effect sizes. Nonetheless, for predictions and PGRS calculation we recommend recalculating effect sizes using the widely accepted and conventional Cox PH model.

The conclusions are mainly based on simulations, but we illustrate the results also on survival data in the Estonian Biobank cohort.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3310. Analysis of vitamin D receptor-binding variants in susceptibility to pediatric-onset multiple sclerosis.

Authors:

D. Yilmaz^{1,2}, C. Adams^{1,2}, M. Horton^{1,2}, J. S. Graves^{3,4}, H. Quach^{1,2}, D. Quach^{1,2}, G. Aaen⁵, B. Greenberg⁶, T. Lotze⁷, S. Mar⁸, J. Ness⁹, Y. Harris⁹, M. P. Gorman¹⁰, L. Benson¹⁰, B. Weinstock-Guttman¹¹, A. Waldman¹², T. Schreiner¹³, M. Rodriguez¹⁴, J-M. Tillema¹⁴, T. Chitnis¹⁵, L. Krupp¹⁶, A. Belman¹⁶, J. Rose¹⁷, C. Casper¹⁸, M. Rensel¹⁹, J. Hart²⁰, C. Schaefer²¹, E. Waubant⁴, L. Barcellos^{1,2,21}, Network of Pediatric Multiple Sclerosis Centers; ¹Univ. of California, Berkeley, Berkeley, CA, ²Ctr. for Computational Biology, Div. of Computing, Data Sci. and Society, Univ. of California, Berkeley, Berkeley, CA, ³Dept. of NeuroSci.s, Sch. of Med., Univ. of California, San Diego, San Diego, CA, ⁴Dept. of Neurology, Univ. of California, San Francisco, San Francisco, CA, ⁵Pediatric MS Ctr., Loma Linda Univ. Children's Hosp., Loma Linda, CA, ⁶Dept. of Neurology, Univ. of Texas Southwestern, Dallas, TX, ⁷Texas Children's Hosp., Houston, TX, ⁸Pediatric-onset Demyelinating Diseases and Autoimmune Encephalitis Ctr., St. Louis Children's Hosp., Washington Univ. Sch. of Med., St. Louis, MO, ⁹Alabama Ctr. for Pediatric-onset Demyelinating Disease, Children's Hosp. of Alabama, Birmingham, Birmingham, AL, ¹⁰Dept. of Neurology, Boston Children's Hosp., Boston, MA, ¹¹Pediatric Multiple Sclerosis Ctr., Jacobs Neurological Inst., SUNY Buffalo, Buffalo, NY, ¹²Div. of Neurology, Children's Hosp. of Philadelphia, Philadelphia, PA, ¹³Children's Hosp. Colorado, Univ. of Colorado, Aurora, CO, ¹⁴Mayo Clinic's Pediatric Multiple Sclerosis Ctr., Rochester, MN, ¹⁵Mass Gen. Brigham Pediatric Multiple Sclerosis Ctr., Massachusetts Gen. Hosp. for Children, Boston, MA, ¹⁶Lourie Ctr. for Pediatric Multiple Sclerosis, Stony Brook Children's Hosp., Stony Brook, NY, ¹⁷Dept. of Neurology, Univ. of Utah Sch. of Med., Salt Lake City, UT, ¹⁸Dept. of Pediatrics, Univ. of Utah Sch. of Med., Salt Lake City, UT, ¹⁹Mellen Ctr., Cleveland Clinic, Cleveland, OH, ²⁰Regional Pediatric MS Ctr., Neurology, Univ. of California, San Francisco, San Francisco, CA, ²¹Kaiser Permanente Div. of Res., Oakland, CA

Abstract Body:

The genetic basis of multiple sclerosis (MS) in adults has been extensively studied and more than 230 susceptibility variants have been identified, to date, through GWAS; less is known about genetic contributions to MS occurring in children (<18 years, 'pediatric-onset' MS) which comprises ~5% of patients. MS prevalence is higher in regions farther from the equator, supporting the hypothesis that vitamin D exposure has a protective effect on MS risk. Mendelian randomization studies have identified causal associations between lower serum vitamin D and increased risk of MS. Vitamin D is important for many biological processes. After being ingested or absorbed, serum vitamin D is first converted to 25-hydroxyvitamin D [25(OH)D], its more stable form, and then to 1,25-dihydroxyvitamin D [1,25(OH)₂D]. Previous studies have established that 25(OH)D signals through the nuclear vitamin D receptor (VDR), a ligand-regulated transcription factor that modulates vitamin D regulated gene expression. Directly testing for associations between VDR binding and disease phenotypes in large-scale human studies poses many challenges. SNPs associated with genetic variation in VDR binding affinity (VDR binding variants or 'VDR-BVs') have been recently identified using ChIP-exo data from calcitriol-stimulated lymphoblastoid cell-lines followed by Allele-seq. We recently studied these VDR-BVs as genetic IVs in adult-onset MS and identified strong evidence for association with several VDR-BVs. The objective of this study was to study adult-onset MS VDR-BV candidates for a role in pediatric-onset MS. We utilized 524 pedMS cases and 1330 controls of European ancestry from the U.S. Network of Pediatric MS Centers and Kaiser Permanente Northern California. Genotyping was performed using Illumina Infinium 660K OmniExpress and OmniExpressExome BeadChip arrays and imputed against reference haplotypes from Phase 3 of the 1000 Genomes Project using IMPUTE4. After quality control measures, SNPs were excluded if minor

ASHG 2022 Annual Meeting Poster Abstracts

allele frequency (MAF) <5%. This resulted in 10 candidate VDR-BVs for analysis. Associations between VDR-BVs and pediatric-onset MS were tested using logistic regression in *PLINK*, and models were adjusted for the first three genetic principal components. One VDR-BV was associated with pediatric-onset MS at $p < 0.05$ (OR: 1.20, 95% CI: 1.03-1.39): rs2531804. This variant is located on chromosome 6 and is upstream of *TEC*. These findings provide further evidence that at least one known VDR-BV associated with adult-onset MS vitamin D is also implicated in pediatric-onset MS.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3311. Ancestry-specific eQTL associations in HIS/NHW Alzheimer's disease cohorts

Authors:

N. Wheeler^{1,2}, P. Benckek^{1,2}, D. Van Booven³, T. Gu³, P. Whitehead³, K. Hamilton-Nelson³, M. Contreras³, J. Sanchez³, S. Tejada³, L. Adams³, P. Mena³, C. Silva⁴, M. Cornejo-Olivas^{5,6}, M. Illanes-Manrique^{5,6}, M. Cuccaro^{3,7}, J. Vance^{3,7}, B. Feliciano-Astacio⁴, J. Haines^{1,2}, G. Beecham^{3,7}, M. Pericak-Vance³, A. Griswold^{3,7}, W. Bush^{1,2}; ¹Cleveland Inst. of Computational Biology, Case Western Reserve Univ., Cleveland, OH, ²Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, ³John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, ⁴Dept. of Internal Med., Univ. Central Del Caribe, Bayamon, Puerto Rico, ⁵Neurogenetics Res. Ctr., Inst. Natl. de Ciencias Neurologicas, Lima, Peru, ⁶Neurogenetics Res. Unit, Fundación San Marcos, Lima, Peru, ⁷Dr. John T. Macdonald Fdn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL

Abstract Body:

Although the genetic etiology of Alzheimer's Disease (AD) varies among ancestrally diverse populations, the influence of genetic variation on the transcriptome within these populations and its association to AD has not been extensively investigated. Multi-ancestry AD datasets featuring both genetic information and gene expression measurements are needed to support expression quantitative trait loci (eQTL) studies in populations underrepresented in genetic research. In this work, we have compared eQTL associations in case/control AD sample populations of both Hispanic (HIS) and non-Hispanic white (NHW) ancestries. We analyzed cohorts of HIS from Cuba, Puerto Rico, and Peru and NHW ancestries ($N_{his}=295$, $N_{nhw}=241$) with corresponding array-based genotypes and RNA-seq mRNA expression levels from whole blood, along with clinical AD diagnoses ($N_{his_case}=149$, $N_{his_control}=146$, $N_{nhw_case}=121$, $N_{nhw_control}=120$). Cis-regulatory eQTL associations between genetic variants and gene expression levels were assessed via linear regression across 14,894 genes, adjusting for age, sex, experimental factors, and ancestry-based principal components. This was repeated across all sample set combinations of ancestry-AD pairings (ancestry=[HIS, NHW, both] * AD=[case, control, both]).

The number of genome-wide significant (p -value $< 10e-7$) eQTL signals varied by ancestry, with the HIS cohort showing 567 eQTL associations (independent of AD status), the NHW cohort 812, and both ancestries combined 1,367. When further sub-setting sample sets to AD cases only, 225 and 253 significant eQTL signals were observed for HISP and NHW cohorts respectively, which impacted 194 (HIS) and 195 (NHW) genes. Interestingly, these ancestry-aware, case-only eQTLs were remarkably distinct between the two ancestries, with only 4 (~0.84%) of the eQTLs and only 3 (0.82%) of the impacted genes being shared by both ancestral cohorts. Of particular interest is the appearance of the variant '22:26184047:T:C' within *MYO18B* (a gene linked to both paired helical filament tau measurement and neurofibrillary tangles measurement phenotypes in reported GWAS studies), as differentially significant in the HIS cohort ($P_{his_case}=2.73e-7$, $P_{his_control}=0.86$) but not in the NHW cohort ($P_{nhw_case}=0.90$, $P_{nhw_control}=0.91$).

We have observed differences in eQTL signals relative to AD diagnosis status between HIS and NHW sample populations, which may indicate the presence of novel ancestry-specific effects impacting the genetic etiology of AD through changes in gene expression.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3312. Ancestry-Specific Polygenic Risk Scores for Telomere Length in TOPMed

Authors:

J. Jin¹, M. A. Taub¹, T. W. Miller-Fleming², M. P. Conomos³, A. Reiner³, N. J. Cox², N. Chatterjee¹, R. A. Mathias¹; ¹Johns Hopkins Univ., Baltimore, MD, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Univ. of Washington, Seattle, WA

Abstract Body:

Telomeres shorten in replicating somatic cells, and telomere length (TL) is associated with age-related diseases. While prior genome-wide association studies using laboratory assays of TL identified genetic loci associated with TL in European and Asian populations, recently published work from the NHLBI Trans-Omics for Precision Medicine (TOPMed) program has expanded our knowledge of TL genetics to include ancestrally diverse African and Hispanic/Latino groups. Despite little evidence of effect-size heterogeneity across ancestry groups, this study identified differences in genetic determinants of TL across ancestries based on differing allele frequencies and an initial polygenic risk score (PRS) analysis revealed cross-population differences in PRS distributions comparing European and African ancestry individuals. This work expands this prior PRS analysis to further leverage the ancestral diversity of TOPMed. In this follow-up, we used TOPMed whole genome sequencing (WGS) of whole blood for genotyping and the bioinformatic estimation of TL in a trans-ethnic sample, including 46,613 European (EUR), 24,891 African (AFR), 4,650 Asian (mainly East Asian, EAS) and 13,860 Hispanic/Latino (LAT) individuals. Using summary statistics from TL GWAS performed separately on each group, we developed ancestry-specific PRS using three methods: (1) “Top-SNP PRS”, calculated as the weighted sum of the number of effect alleles for SNPs reaching genome-wide significance ($p \leq 5 \times 10^{-8}$) in the European-only GWAS, weighting by the GWAS summary statistics; and two methods that borrow information across ancestries: (2) weighted LDpred2, a weighted sum of the ancestry-specific PRS obtained by the LDpred2 algorithm using ancestry-specific GWAS summary statistics, and (3) MEBayes, a novel LD-based method which jointly models the correlated ancestry-specific GWAS summary statistics. On a validation set of 14,387 held-out multi-ancestry individuals in TOPMed, we saw an explained variance (R^2) of 0.034, 0.046 and 0.047 on EUR; 0.016, 0.022 and 0.027 on AFR; 0.029, 0.048 and 0.046 on EAS; and 0.021, 0.033 and 0.031 on LAT for the Top-SNP, weighted LDpred2 and MEBayes approaches, respectively. We observe that the two multi-ancestry PRS, weighted LDpred2 and MEBayes, show substantially improved prediction accuracy on the non-European populations (AFR, LAT, and EAS) compared to the Top-SNP PRS. We are applying these multi-ancestry PRS to perform a phenome-wide association study on individuals in the BioVU database (72,597 EUR, 15,615 AFR, 2,392 LAT, 884 EAS) to further test the association between our TL PRS and 1,800 clinical traits, including age-related diseases.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3313. Application of the colorectal cancer polygenic risk score to the VA CSP #380 “Prospective Evaluation of Risk Factors for Large (≥ 1 cm) Colonic Adenomas in Asymptomatic Subjects” longitudinal colonoscopy screening cohort

Authors:

X. Qin^{1,2}, B. Sullivan^{2,1}, T. S. Redding IV¹, D. Weiss³, S. Pyarajan⁴, Y. Shi⁴, A. Madison¹, K. J. Sims¹, C. Williams^{1,2}, J. Dominitz⁵, D. Lieberman^{6,7}, E. R. Hauser^{2,1}; ¹Cooperative Studies Program Epidemiology Ctr.-Durham, Durham VA Med. Ctr., Durham VA Hlth.Care System, Durham, NC, ²Duke Univ. Sch. of Med., Durham, NC, ³Perry Point VA Med. Ctr., Perry Point, MD, ⁴Massachusetts Veterans Epidemiology Res. and Information Ctr. (MAVERIC), VA Boston Hlth.care System, Boston, MA, ⁵VA Natl. Gastroenterology and Hepatology Program, Washington, DC, ⁶Portland VA Med. Ctr., Portland, OR, ⁷Oregon Hlth.& Sci. Univ., Portland, OR

Abstract Body:

Background: VA CSP #380 is a longitudinal study of 3121 participants aged 50-75 enrolled for screening colonoscopy from 1994-1997. Follow-up using the VA electronic health records is ongoing. The goal of our analysis was to apply the well-calibrated colorectal cancer (CRC) polygenic risk score (PRS) of Huyghe et al. (2019) and Thomas et al. (2020) to long-term programmatic screening and surveillance colonoscopy outcomes. Method: Advanced neoplasia (AN) cases are those participants with invasive CRC or advanced adenoma (AA) at any time within 10 years of follow up, with AA defined as an adenoma >10 mm, an adenoma with villous histology or high-grade dysplasia on at least 1 exam. Controls are individuals with no adenomas or no advanced adenomas at all exams. A total of 612 individuals were genotyped with the Illumina Infinium Omni2.5-8 v1.3 Beadchip. Imputation using the TopMed panel was performed using Shapeit4 and Minimac4. After QC, 137 of the 140 CRC-risk SNPs were used in current PRS analysis as implemented in PRSice. We evaluated the association of the PRS with AN using the rstanarm joint model incorporating colonoscopy outcomes over time, accounting for age, sex, ancestry principal components and differential surveillance. The PRS was divided into tertiles and the frequency of participants with CRC ever, AA ever, and control was calculated for each tertile. Results: Outcomes were assessed at 1471 colonoscopies occurring across 10 years. Among the 612 with a CRC PRS, 22 were in the CRC group, 167 with AA, and 423 were controls. The CRC PRS was significantly associated with the occurrence of AN ($p=0.0057$), even after accounting for differential surveillance ($p=0.0056$). The lowest PRS tertile included 26.8% controls, 14.5% AA and 9.1% CRC. The highest PRS tertile included 12.8% controls, 15.1% AA and 31.8% CRC. Conclusion: In an independent longitudinal screening colonoscopy cohort, the CRC PRS was significantly associated with advanced neoplasia. More individuals with CRC were in the highest PRS tertile. These results suggest that the CRC PRS may be useful for risk prediction during long-term colorectal cancer screening by helping better target scarce resources to the highest risk individuals most likely to benefit.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3314. Applying human phenomics to electronic health records provides a framework for understanding skin-aging related phenotypes

Authors:

A. Abbasi¹, M. Liu¹, B. Riley-Gillis¹, J. Waring², H. Jacob¹, S. Manson Brown³, T. Cheng³, R. Mehta³, N. Smaoui¹; ¹AbbVie, North Chicago, IL, ²ABBVIE, NORTH CHICAGO, WI, ³AbbVie, Irvine, CA

Abstract Body:

Standard disease classification and defining human phenotypes are incomplete and imprecise because biological mechanisms underpinning human diseases and phenotypes are not yet captured even with different coding systems and methods. Such methods are arbitrary (e.g., hypertension), based on a location or organ (e.g., tumors), or have been used since physicians originally described certain diseases (e.g., Alzheimer's diseases) over a century ago, which cause confusion and uncertainty in diagnosis, drug discovery and medical care. In this analysis, we used the UK Biobank linked to primary care electronic health records (EHR) data to identify and reclassify skin-aging related phenotypes in 502,505 participants. Using EHR-based phenotyping, we improved our phenotype classification and increased sample size up to 40 times more than using any given combinations of ICD10 from disease data curation. For instance, we identified 9,327 individuals with dry skin (an increase from 219 patients using ICD10 codes only) and demonstrated that dry skin is associated with environmental exposures (air pollution, $P=0.002$) and lifestyle factors (trouble falling sleep, $P=2.7 \times 10^{-11}$ and smoking, $P=5.5 \times 10^{-6}$). Given that clinical information spanning several decades is available through EHR repositories, the use of EHR-based deep phenotyping can provide value beyond our current understanding of disease and phenotype classification and clinical research. Future research from this group will explore the value of deep phenotyping by applying machine learning approaches to big data and further develop genome-wide association analysis.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3315. Assessing the safety of lipid-modifying medications among East Asian adolescents using genetics: Evidence from Hong Kong's "Children of 1997" birth cohort

Authors:

S. Luo¹, Y. Chan¹, H. Lam², C. Tang¹, B. He¹, M. Kwok¹, C. Schooling^{1,3}, G. Leung¹, S. Au Yeung¹; ¹The Univ. of Hong Kong, Hong Kong, Hong Kong, ²The Chinese Univ. of Hong Kong, Hong Kong, Hong Kong, ³City Univ. of New York, New York, NY

Abstract Body:

Background 3-hydroxy-3-methylglutaryl coenzyme A (HMGCR) reductase inhibitors (statins) are the first-line pharmacologic treatments for children and adolescents who have elevated low-density lipoprotein cholesterol (LDL-C). However, the long-term safety of statins and other common lipid-modifying medications (proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors) among adolescents in East Asia is unclear. **Methods** This is a genetic study focusing on two known protein-encoding genes (*HMGCR* and *PCSK9*) of lipid-modifying drug targets using the data from participants at age of ~17.5 years in the Biobank Clinical Follow-up (n=3,353) of Hong Kong's "Children of 1997" population-representative Chinese birth cohort. Genetic variants strongly associated with LDL-C (p -value 5×10^{-8}) and in low linkage disequilibrium ($r^2 < 0.1$) within the corresponding gene were selected. We derived a weighted genetic risk score (GRS) using the effect size estimates from East Asian Genome-wide association studies to mimic the effects of each drug. The efficacy outcomes included cholesterol-related serum nuclear magnetic resonance (NMR) metabolomics, whilst other metrics from NMR metabolomics (e.g., amino acid and ketone bodies) and extensive health phenotypes, (e.g., anthropometric and haematological traits, renal, liver and lung functions) were considered as safety outcomes. The association of each GRS of drug target on the outcomes was assessed using multivariable linear regression, adjusted for age at sample collection, sex and the top 6 principal components of ancestry. **Results** Genetic inhibition of HMGCR was associated with a substantial reduction in cholesterol-related NMR metabolomics, e.g., total cholesterol (-0.72 standard deviation (SD) per GRS, 95% confidence interval (CI) -0.38 to -1.06), clinical LDL-C (-0.78 SD, 95% CI -0.43 to -1.12) and apolipoprotein B (-0.77 SD, 95% CI -0.42 to -1.12). Similar efficacy was seen for genetic inhibition of PCSK9. Genetic inhibitions of lipid-modifying drug targets were not associated with any of the safety outcomes assessed in adolescents from the "Children of 1997" birth cohort after correcting for multiple comparisons (p -value < 0.001). **Conclusion** As a proof of concept, this study assessed the efficacy and safety of drug targets among late adolescents. This study provided genetic evidence that the pharmacological actions of common lipid-modifying medications (e.g., statins and PCSK9 inhibitors) in terms of their primary drug target are generally safe in adolescents. Larger studies are needed for replication.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3316. Assessing the value of non-coding variants in 500,000 whole-genome sequences in the discovery of gene-phenotype associations

Authors:

Y. Lo¹, S. Donthi¹, E. Arner², E. Dermitzakis³, J. Davitte¹, P. Gormley⁴, R. Scott², E. Ingelsson³, A. Cortes², **J. Liu¹**; ¹GSK, Upper Providence, PA, ²GSK, Stevenage, United Kingdom, ³GSK, San Francisco, CA, ⁴GSK, Cambridge, MA

Abstract Body:

Whole-genome sequencing (WGS) studies of large cohorts provide almost complete coverage of the human genome, allowing us to explore gene-phenotype associations in genomic regions that are not well captured by targeted sequencing and array-based genotyping methods. While large-scale exome sequencing studies have demonstrated the power of gene-based rare-variant collapsing methods for drug discovery, extending these approaches to non-coding regions is challenging due to smaller WGS sample sizes, the difficulty of aggregating non-coding variants by annotation class, and mapping these annotations to genes. Using the largest WGS cohort to date of ~500,000 participants of the UK Biobank, we investigate the additional value of non-coding variants for the discovery of gene-phenotype associations. We explore various strategies to aggregate rare non-coding variants by function. These methods incorporate information from multi-omics data such as alternative splicing and chromatin accessibility, enabling the mapping of these non-coding regions to the genes they act on. We perform region-based association tests to evaluate the burden of rare non-coding variants on hundreds of quantitative and binary traits.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3317. Assessment of correlated multi-SNP Mendelian Randomization method using brain eQTLs to discover schizophrenia genes as a case study

Authors:

D. Baird¹, J. Robinson², Y. Huang¹, H. Runz¹, C-Y. Chen¹, T. Gaunt², E. Tsai¹; ¹Biogen, Cambridge, MA, ²Univ. of Bristol, Bristol, United Kingdom

Abstract Body:

Mendelian Randomization (MR) analysis on cis acting gene expression quantitative trait loci (eQTLs) is used to nominate genes involved in regulation of disease traits. As MR is typically based on a single SNP, we performed a correlated multi-SNP MR between brain eQTLs and schizophrenia to evaluate the potential of this method in gene discovery.

eQTLs were selected from the MetaBrain cortex dataset and schizophrenia SNPs from the PGC 2014 GWAS, both were European samples. Within the 1Mb regions of interest (sentinel eQTL $p < 5 \times 10^{-5}$), eQTL and GWAS SNPs effects were harmonized and then LD clumped ($r^2 < 0.36$, $p < 5 \times 10^{-5}$). An Inverse Variance Weighted (IVW) fixed effects model was fitted using the pairwise LD matrix for the region, as well as the Egger intercept test and IVW PCA method (Burgess, 2017) to assess robustness of findings. A single SNP Wald ratio (WR) effects analysis based on the independent eQTLs within the region available after LD clumping ($r^2 < 0.01$, $p < 5 \times 10^{-5}$) was also conducted.

Correlated multi-SNP IVW analysis was performed on 10,146 gene regions and the WR analysis on 24,891 independent eQTLs spanning these regions. IVW and WR Z scores agreed closely for genes which had a single SNP only ($r=0.96$) compared to genes which had multiple SNP instruments ($r=0.71$). 50 genes had a Bonferroni corrected IVW $p < 4.93 \times 10^{-6}$ and 58 eQTLs for 59 genes had a Bonferroni corrected WR $p < 2.01 \times 10^{-6}$. 24 of the genes detect by IVW and 21 of the genes detected by WR analysis failed the Egger intercept test ($p < 0.05$) used to indicate bias due to directional pleiotropy. After filtering for the genes which passed the intercept test, 16 genes were Bonferroni significant in both analyses, directionality agreed between effect estimates and the Z scores were highly correlated ($r=0.90$). 10 extra genes were found in the IVW analysis. 3 genes were not associated in the IVW PCA model (at $p < 0.05$) indicating instrument selection could be biasing these IVW estimates. As an illustrative example, *POM121L2* returned a more significant IVW estimate on the correlated SNPs (IVW $B = 0.080$, $SE=0.016$, $p=7.41 \times 10^{-7}$, $n_{snps}=36$) compared to the WR analysis (top WR: $B=0.056$, $SE=0.019$, $p=3.43 \times 10^{-3}$). 22 genes were Bonferroni significant in the WR analysis but not the IVW, and 9 of these genes were also not identified by the IVW PCA method. For example, *ZNF322* was not associated in the multi-SNP analysis (IVW $B = 0.00043$, $SE=0.011$, $p=0.70$, $n_{snps}=77$) but is detected in the WR analysis (top WR: $B=-0.24$, $SE=0.045$, $p=7.56 \times 10^{-8}$).

We have demonstrated that multi-SNP and single SNP MR agree for most genes, but it is possible for the multi-SNP to both find additional genes and identify erroneous genes, particularly in more LD complex regions.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3318. Association Between Brain Structure and Alcohol Use Behaviors in Adults: A Mendelian Randomization and Multiomics Study

Authors:

L. A. Mavromatis¹, D. B. Rosoff^{1,2}, R. B. Cupertino³, H. Garavan³, S. Mackey³, F. W. Lohoff¹; ¹Section on Clinical Genomics and Experimental Therapeutics, Natl. Inst. on Alcohol Abuse and Alcoholism, NIH, Bethesda, MD, ²NIH-Oxford-Cambridge Scholars Program; Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, ³Dept. of Psychiatry, Univ. of Vermont Coll. of Med., Burlington, VT

Abstract Body:

Background: Past studies have identified associations between brain macrostructure and alcohol use behaviors. However, determining the directionality of these associations is impeded by the limitations of observational studies and practical and ethical barriers to randomized control trials.

Methods: In this study, we used summary statistics from genome-wide association studies of European ancestry cohorts ($N_{\text{total}} = 763,874$) to perform bidirectional Mendelian randomization (MR) to find evidence of causal relationships between brain structure and alcohol use. Our primary analyses evaluated structural magnetic resonance imaging phenotypes (global cortical thickness and surface area ($N = 33,709$) and left/right subcortical volumes ($N = 19,629$)) and three alcohol use behaviors (alcoholic drinks per week ($N = 537,349$), binge drinking frequency ($N = 143,685$), and alcohol use disorder ($N = 29,502$)). We performed multivariable MR (MVMR) to evaluate the robustness of our bidirectional MR findings and transcriptome-wide association studies (TWAS) and cell-type enrichment analyses to investigate the biology underlying identified relationships.

Results: Our primary MR analyses identified negative associations between genetically-predicted global cortical thickness and binge drinking ($\beta = -2.52$, 95% CI [-4.13, -0.91]) and drinks per week ($\beta = -0.88$, CI [-1.37, -0.40]) at a false discovery rate (FDR) of 0.05. These associations remained significant in MVMR models that accounted for neuropsychiatric phenotypes, substance use, trauma, and neurodegeneration. TWAS of global cortical thickness and alcohol use behaviors identified five genes at the 17q21.31 locus oppositely associated with cortical thickness and binge drinking and/or drinks per week (FDR = 0.05). Cell-type enrichment analyses implicated glutamatergic cortical neurons in alcohol use behaviors.

Discussion: These findings show that associations between global cortical thickness and alcohol use may reflect a predispositional influence of cortical structure. 17q21.31 genes and glutamatergic cortical neurons may play a role in this association. While replication studies are needed, these findings should enhance the understanding of relationships between brain structure and alcohol use.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3319*. Association of genetically predicted iron homeostasis biomarkers in type 2 diabetes and glycemic traits: A two-sample Mendelian Randomization study

Authors:

Y. Liang¹, S. Luo¹, T. Wong¹, B. He¹, C. Schooling^{1,2}, S. Au Yeung¹; ¹The Univ. of Hong Kong, Hong Kong, Hong Kong, ²City Univ. of New York, New York, NY

Abstract Body:

Introduction: Previous Mendelian Randomization (MR) studies suggested that iron status is positively associated with type 2 diabetes (T2D). However, these studies included highly pleiotropic instruments related to hereditary haemochromatosis (HH) and did not comprehensively assess causation for different iron homeostasis biomarkers. We conducted a MR study to address these research gaps using the most up-to-date genome-wide association study (GWAS) of iron homeostasis biomarkers.

Methods: We extracted instruments from the largest GWAS ($N \leq 246,139$) of iron homeostasis biomarkers (i.e., ferritin, serum iron, total iron-binding capacity (TIBC), and transferrin saturation (TSAT)) and applied them to the largest suitable summary statistics from GWAS of T2D (DIAMANTE $N = 933,970$ and FinnGen $N = 252,162$), glycemic traits (fasting glucose ($N = 209,605$), two-hour glucose ($N = 64,469$), glycated hemoglobin ($N = 149,006$), and fasting insulin ($N = 158,550$)). Inverse-variance weighted (IVW) was the main analysis method, with sensitivity analyses including MR-Egger and weighted median, as well as exclusion of instruments which are highly pleiotropic instruments from *ABO* and HH-related genes. Hemoglobin (Hgb) was used as a positive control outcome.

Results: Using IVW, all four iron homeostasis markers were not associated with T2D ($OR_{\text{ferritin}}: 0.99$, 95%CI 0.89 to 1.10; $OR_{\text{serum iron}}: 1.05$, 95%CI 0.98 to 1.12; $OR_{\text{TIBC}}: 0.96$, 95%CI 0.91 to 1.01; $OR_{\text{TSAT}}: 1.01$, 95%CI 0.96 to 1.06). Ferritin, serum iron, and TSAT were associated with lower HbA1c and higher Hgb, while TIBC had opposite associations. We did not find strong evidence for an effect of iron homeostasis biomarkers on any other glycemic traits considered. These findings were similar to sensitivity analyses. However, when restricting to instruments in HH or with concordant effects which included HH instruments, higher iron status was associated with a higher risk of T2D.

Discussion: This study suggests iron homeostasis biomarkers unlikely plays a role in etiology of T2D in the general population. However, whether the positive association of HH and T2D reflects the consequence of severe iron overload or pleiotropy requires additional investigations.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3320*. Association testing in admixed and multiethnic data sets nominate genetic modifiers of age at onset of Alzheimer's disease

Authors:

E. Blue, J. Broome, D. Xue; Univ. of Washington, Seattle, WA

Abstract Body:

Alzheimer's disease (AD) is a leading cause of dementia and mortality in the United States, exhibiting phenotypic, genetic, and allelic heterogeneity. Although age-at-onset (AAO) of AD is highly heritable, few genome-wide association studies (GWAS) of AD incorporate a survival analysis framework. Our adapted linear mixed model with orthogonally partitioned structure for survival analysis traits yields results highly concordant with a Cox mixed-effect model for GWAS in 0.5% the time ($|r_{\text{beta}}| = 0.97$, $|r_{\text{log10p}}| = 0.93$). We performed GWAS for AAO of AD in two data sets enriched with multiplex families: the Columbia University Study of Caribbean Hispanics with Familial and Sporadic Late Onset Alzheimer's disease (CU-LOAD) and the multiethnic National Institute on Aging Genetics Initiative for Late-Onset AD study (NIA-LOAD). Phenotype and genotype data underwent quality control, genotype imputation, then association testing, adjusting for sex, *APOE* $\epsilon 2$ and $\epsilon 4$ allele counts, principal components capturing population structure, and genetic relatedness. The CU-LOAD GWAS representing 1305 cases and 1734 controls identified a significant association with AAO of AD at 14q24.3 ($p < 5e-08$), while another 15 loci reached suggestive evidence of association ($p < 1e-06$) ($\lambda = 0.99$). The signal at 14q24.3 is explained by the *PSENI* p.G206A allele implicated in early-onset autosomal-dominant AD. This causal variant has half the allele frequency (MAF = 0.5% vs. 1.1%), twice the effect size estimate ($\beta = 0.95$ vs. 0.44), and is 2.5Mb from the GWAS peak. Two loci nearly reached genome-wide significance in the NIA-LOAD GWAS representing 2307 cases and 2241 controls ($p < 7e-08$; 4q31.3 and 19q13.32; $\lambda = 1.02$). Fifteen loci had suggestive evidence of association ($p < 1e-06$), implicating genes linked to familial forms of dementia (*CHMP2B*, *PDGFRB*), their interactors (*VEGFC*), or AD risk (*RBFOX1*). The AAO-associated alleles can vary in frequency by orders of magnitude across populations, but global and local ancestry proportions among carriers do not consistently over-represent the population with the maximum reference allele frequency. Given inconsistent evidence of replication for AD GWAS loci identified in samples representing different populations and the association of local ancestry at *APOE* with risk of AD, future studies may need to match for global or local ancestry and enrichment for families affected by AD to replicate these results.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3321*. Asymmetric non-portability across populations of different ancestry.

Authors:

J. Ballard, L. O'Connor; Broad Inst., Cambridge, MA

Abstract Body:

Polygenic risk scores (PRS) have the potential to improve clinical outcomes by identifying individuals who are genetically predisposed to developing certain diseases (Khera et al. 2018 Nat Genet). However, it has been shown that PRS trained on individuals of European ancestry (EUR) are not portable to individuals of African ancestry (AFR) due to differences in allele frequency (AF) and linkage disequilibrium (LD) patterns (Scutari et al. 2016 PloS Genet, Martin et al. 2019 Nat Genet). Here, we performed simulations to investigate the portability of PRS from African to European individuals as well. We found that non-portability is directionally asymmetric: PRS trained in AFR were more portable to EUR ($r^2_{AFR \rightarrow EUR} = 0.27$ vs. $r^2_{AFR \rightarrow AFR} = 0.34$) than PRS trained in EUR were to AFR ($r^2_{EUR \rightarrow AFR} = 0.15$ vs $r^2_{EUR \rightarrow EUR} = 0.40$).

To find the source of this asymmetry, we performed simulations with population specific LD and population specific AF (but a cross-population genetic correlation $r_{crosspop} = 1$), as well as with identical LD and AF for both populations but $r_{crosspop} < 1$; LD and AF both contributed to asymmetric non-portability. Differences in effect sizes, i.e. cross-population genetic correlation $r_{crosspop} < 1$, contribute to non-portability but not to asymmetric non-portability in these simulations. When we examined asymmetry in pairs of all five populations (EUR, AFR, EAS, SAS, and AMR), we found strong asymmetry due to both AF and LD in population pairs that involved AFR, indicating that training in AFR produces much more accurate predictions in non-AFR populations than training in non-AFR populations and testing in AFR. In non-AFR population pairs, we observed little to no asymmetry. This is likely explained by the fact that AFR individuals have a larger number of common haplotypes and common SNPs, allowing the information captured by training in AFR to sufficiently account for LD and AF in other populations, but not the other way around. We derived a formula for the expected accuracy of an optimal PRS under a Gaussian model; consistent with the simulations, it predicts that asymmetric AF and LD differences cause asymmetric non-portability between EUR and AFR.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3322. Auxiliary classifier generative adversarial network for phenotypic conditional genomic data synthesis.

Authors:

E. U. Lim¹, A. Lim^{1,2,3}, C. S. Fann^{1,2}; ¹Inst. of BioMed. Sci., Academia Sinica, Taipei, Taiwan, ²Taiwan Intl. Graduate Program in Molecular Med., Natl. Yang Ming Chiao Tung Univ. and Academia Sinica, Taipei, Taiwan, ³ASUS Intelligent Cloud Services (AICS), Taipei, Taiwan

Abstract Body:

The utilization and translation of genomic data from large biobanks has revolutionized the field of biomedical research, drug development and precision medicine. Despite the advances in genetic epidemiology research, limited sample size of certain rare diseases and minority population remains a critical issue. As an alternative to collecting more samples, generating realistic synthetic human genomic data by mimicking the population structure can uplift the sample size of disease cohorts or minority groups. Besides, the availability of extensive phenotypic data from biobanks provides an unprecedented opportunity for researchers to make use of phenotypic information in the human genomic data augmentation process. In this regard, we propose an Auxiliary Classifier Generative Adversarial Network (ACGAN) tailored for genomic data synthesis that takes into account the dimensionality structure of haploid genotype (haplotype) data and the categorical phenotypic traits labels. The output haplotype data can optionally be transformed into genotype data, unordered representation of the base pair sequences, for post data augmentation analyses. As an initial proof of concept, we applied the ACGAN on phased haplotype data of individuals with or without type 1 diabetes (T1D) from UK Biobank. The input data consists of a class label representing T1D case or healthy control, as well as a set of sequences of haplotype base pairs with select single nucleotide polymorphisms (SNPs), of which includes 2 known SNPs, rs2476601 and rs6679677, in the *PTPN22* gene associated with increased risk of T1D as reported by prior genome-wide association studies (GWAS). T1D was chosen due to the high heritability and high accuracy of existing polygenic risk score (PRS) models, which enable further evaluation of the generated samples' utility. According to the allele frequency spectrums of the real and generated samples, the ACGAN is capable of capturing the overall allele frequency distribution including low frequency alleles, which was addressed previously as a challenge in other studies. Overall, the results show that our proposed ACGAN provides a novel and promising approach for performing phenotypic conditional human genomic data augmentation in population genetics and genetic epidemiology research applications.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3323. Being overweight in late life can be a causal factor contributing to longevity: a Mendelian Randomization study.

Authors:

H. Duan, R. Holmes, K. Arbeev, D. Wu, A. Yashin, S. Ukraintseva; Biodemography of Aging Res. Unit, Social Sci. Res. Inst., Duke Univ., Durham, NC

Abstract Body:

Introduction: Recent research suggested that being overweight may be optimal for survival in late life. However, biological mechanisms of this phenomenon are poorly understood. Mendelian Randomization (MR) could be a useful tool for clarifying causal relationships between longevity and weight/BMI in human data. The *APOE* $\epsilon 4$ allele is known for its associations with both lower weight/BMI and reduced survival in older adults, so it could be used as an instrumental variable in respective MR study.

Methods: We conducted MR analysis in the Health and Retirement Study data subsample of individuals aged 75+. We used the number of *APOE* $\epsilon 4$ alleles as an *instrumental variable (IVar)*; binary BMI variable (BMI between 25-30 ('overweight') vs. BMI < 25, at ages 75-85) as *exposure*; and two binary 'longevity' variables (**1.** survived age 85+ vs. died before 85; and **2.** survived age 90+ vs. died between 80 and 85) as *outcomes*. The non-stratified samples included 3338 and 1664 participants, respectively. Analyses were done in sex and race strata as well. The SAS software was used to estimate initial models, and then summary statistics from these models were used in the R-package *MendelianRandomization*. F-value > 10 criteria was applied to ensure the strength of *IVar*. The causal effect was evaluated using Inverse-Variance Weighted (IVW) and maximum likelihood (ML) methods.

Results: The MR study revealed that having BMI in 'overweight' range 25-30 at ages 75-85 has significant causative effect on survival 85+ (causal estimate=1.56, IVW *p-value*=2.92E-06, ML *p-value*=0.007) in non-stratified sample, as well as in females (causal estimate=1.58, IVW *p-value*=4.7E-05, ML *p-value*=0.018, N=1877) and whites (causal estimate=1.20, IVW *p-value*=5.1E-05, ML *p-value*=0.006, N=2797). Similarly, being overweight had significant causative effect on living to age 90 and beyond in non-stratified sample (causal estimate=1.70, IVW *p-value*=1.68E-06, ML *p-value*=0.013), as well as in females (causal estimate=1.58, IVW *p-value*=3.25E-05, ML *p-value*=0.021, N=976) and whites (causal estimate=1.25, IVW *p-value*=5.17E-05, ML *p-value*=0.0098, N=1407).

Conclusion: Results of our MR study strongly suggest that being overweight in late life (ages 75-85) is one of the causal factors contributing to extreme longevity. It might reflect a better ability of the body to resist the aging-related physical decline, and/or be a sign of better resilience of overweight people, which warrants further investigation.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3324. Bi-ancestral phenome-wide association of complement component 4 haplotypes in 550,000 individuals

Authors:

S. Venkatesh¹, M. Kim², M. Gandal², G. Voloudakis¹, P. Roussos¹; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract Body:

The major histocompatibility complex (MHC) locus is critical to the immune response, and contains the strongest genetic association with schizophrenia. Of particular interest is complement component 4 (C4), which is involved in synapse elimination, and thus contributes to the “excessive pruning” hypothesis of schizophrenia. Here we investigate how different C4 haplotypes relate to traits across the electronic health record (EHR), with a particular focus on neuropsychiatric disorders. We impute C4 haplotypes in over 550,000 Million Veteran Program (MVP) individuals across European and African ancestries.

We impute C4 haplotypes in the MVP using the C4 reference panel from Kamitaki et al. 2020 that was built in a cohort of diverse ancestry. We perform C4 haplotype-wide association studies for selected phenotypes, C4 phenome-wide association studies, and C4 medication-wide association studies. We perform mediation and variance partition analyses to explore the relationship among C4 haplotypes, genetic liability and EHR features. Finally, we use PRS-CS to calculate polygenic risk scores (PRS) for a variety of disorders to understand their relationship with C4 haplotypes. We exclude the MHC region from these calculations in order to assess genetic liability outside the MHC locus.

We find 156 FDR-significant associations between C4 haplotype BS and phenotypes in Europeans. This includes neuropsychiatric phenotypes such as schizophrenia (confirming established results from Sekar et al. 2016), mood disorders, as well as 146 non-neuropsychiatric phenotypes such as ulcerative colitis and psoriasis. We also find 82 significant associations between C4 haplotype BS and medications in Europeans, including olanzapine, prednisone, budesonide (prescribed for ulcerative colitis) and fluocinonide (prescribed for psoriasis). We find FDR-significant associations in Africans as well, such as the association between C4 haplotype AL and chronic pharyngitis and nasopharyngitis. We also explore the relationship between PRS and C4 haplotypes using mediation and variance partition analyses, and find that schizophrenia PRS accounts for approximately 50% of C4 haplotype BS’s effect on schizophrenia.

Overall, C4 haplotypes have strong associations across the EHR and a complex relationship with overall genetic liability in neuropsychiatric disorders. In particular, C4 haplotype BS has 238 FDR significant associations in Europeans. To follow up on these results, we plan to explore whether there is evidence of evolutionary trade-offs among C4 haplotypes.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3325. Boosting SNP discoverability: How genetic architecture impacts the power of multi-trait GWAS

Authors:

Y. Suzuki¹, H. Ménager², C. Nerin¹, R. Torchet², P. Lechat², L. Troubat¹, H. Aschard¹, H. Julienne^{1,2}; ¹Inst. Pasteur, Université de Paris, Dept. of Computational Biology, Paris, France, ²Inst. Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, Paris, France

Abstract Body:

Despite the ever-increasing sample size of Genome-Wide Association Study (GWAS) summary statistics, many causal SNPs for human complex traits remain undetected by univariate GWASs. To overcome this lack of statistical power, investigators have developed multi-trait GWAS methods. However, multi-trait methods entail challenges, such as data harmonization, computational cost, and missing values. More critically, there is no established strategy to astutely select sets of traits to maximize the statistical power of multi-trait GWAS. To address these practical obstacles and investigate the impact of genetic architecture on SNP discoverability, we implemented a python package *JASS* (Joint Analysis of Summary Statistics) with specific attention to computational efficiency (e.g. on 62 traits and ~7 million SNPs, JASS runs in <20 minutes). We applied JASS on 1620 sets of two to 62 traits randomly selected out of 108 traits curated summary statistics (<https://jass-beta.pasteur.cloud/>), and derived for each set the association gain (i.e. the number of new associations identified) achieved by multi-trait test over the original univariate tests. We assessed the effect of multiple factors on this gain and identified five key features: the polygenicity (the number of causal SNPs), the mean effect size (average genetic effect that causal SNPs have on the trait), the genetic covariance between traits, the environmental covariance between traits, and the orientation of genetic and environmental signals. Together, these five features had a substantial predictive power ($R^2 \sim 0.4$, P value $< 2.2e-16$), providing a strong indicator of the multi-trait performance. Mean effect size was the largest contributor (P value = $3.6e-14$). Moreover, we observed a large contribution of nonlinear interactions between the five features (R^2 increased from 0.25 to over 0.4 by including non-linear interactions). We also found that the contribution of the environmental covariance was more important than the genetic covariance (permutation importance ~ 0.12 , ~ 0.05 , respectively). Further refined investigation might provide investigators a powerful tool to select trait to include in their multi-trait association tests and effectively uncover associations with low discoverability in univariate GWAS.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3327. Broad- and narrow-sense heritability-contributing regions are shared across age, genetic sex and geographical cohorts.

Authors:

D. Udwin¹, S. Zhang¹, S. P. Smith¹, L. Crawford^{1,2}, S. Ramachandran¹; ¹Brown Univ., Providence, RI, ²Microsoft Res. New England, Cambridge, MA

Abstract Body:

Genome-wide association (GWA) summary statistics exhibit heterogeneity when individuals from the same database are partitioned based on certain covariates. The goals of our work are to explore whether this heterogeneity extends to genome-wide narrow-sense heritability and to identify genomic regions that contribute to the broad-sense heritability of complex traits. To address the first aim, we make use of the existing Heritability Estimator of Summary Statistics (HESS) model; in the second aim, we introduce the Epistatic and Additive Variance Estimation (EAVE) framework, which uses Haseman-Elston (HE) regression to estimate the proportion of phenotypic variation explained by nonlinear genetic effects. We utilize a population of 349,411 unrelated white British individuals from the UK Biobank and partition their data according to genetic sex, age (<62, 62-69, 70-74, >74) and geographic locations of birth. We find:

(1) Genetic sexes have different genome-wide narrow-sense heritabilities for triglyceride levels, but the same top-contributing region: Chromosome 11:116383348-117747110, which includes genes *APOA1*, *APOA4* and *APOA5*. Likewise, age-stratified cohorts do not have the same genome-wide narrow-sense heritability for cholesterol, but do share the same top-contributing region: Chromosome 19:9238393-11284028, containing *APOC1*, *APOC2*, *APOC4*, *APOC2*, *APOE* and *LDLR*. Narrow-sense heritability is conserved across birthplace cohorts with the top-contributing region being mutual: Chromosome 20:32813441-34960446, containing *GDF5*, *FAM83C* and *FER1L4*.

(2) The homogeneity of heritability-contributing regions holds from an epistatic perspective. 86 of the top one hundred regions that contribute to triglyceride levels through epistatic effects are shared between the genetic sexes ($p = 5.971e-114$). The number of top one hundred epistatic heritability-contributing regions to cholesterol that are shared between any two age cohorts ranges from 78 ($p = 3.014e-93$) to 82 ($p = 3.551e-103$). Of the 45 possible comparisons between pairs of birthplace cohorts, 19 share a significant number of mutual top heritability-contribution regions to height (here, significance is indicated by p-value below 0.05 / 45, to account for multiple testing). On average, these pairs of cohorts share 17 of the top one hundred epistatic heritability-contributing regions ($p = 3.713919e-05$).

We conclude that heritability-contributing regions with respect to both additive and genetic variation are shared across cohorts for select traits. We add that given precomputed genomic relatedness matrices, EAVE produces estimates in 99 seconds for $n=100K$ using 36 CPUs.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3328. BTS: Bayesian Tissue Significance scores for functional evaluation of GWAS results

Authors:

M. Ionita, **P. P. Kuksa**, Y. Leung, K. Clark, O. Valladares, L-S. Wang; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Summary statistics from genome-wide association studies (GWAS) are often used in downstream analyses, such as fine mapping or determination of functional context (i.e., affected tissues, biological mechanisms and target genes). The increasing sizes of both association studies and functional genomic (FG) and annotation data provide an opportunity to perform unbiased analyses, using fast algorithms to search for associations between the putatively causal GWAS variants and thousands of cell type-specific functional annotations. To this end, we propose BTS (Bayesian Tissue Significance score), an efficient computational framework for joint estimation of causal variants and their enrichment in annotation tracks. BTS is a novel algorithm that combines GWAS summary statistics, linkage disequilibrium and functional annotations into a Bayesian model, with significantly reduced complexity for large-scale evaluation. We connect BTS with FILER (Kuksa et al, 2022), a large-scale functional genomics repository, to allow for systematic annotation of GWAS variants across >1,100 cell types and >190 genomic and molecular functions. BTS evaluates genome-wide models for each functional annotation, then prioritizes the ones with the highest association between likely causal variants and the annotation. Due to careful management of data and intermediate results, BTS is up to 100 times more efficient than existing approaches, which allows for fast, unbiased analysis of data across all available tracks.

To systematically evaluate GWAS results and screen for the relevant functional contexts, we used the harmonized FG and annotation data collection from FILER including >60,000 datasets of tissue-specific regulatory elements, transcription factor binding or chromatin states. We applied BTS to three immune-related GWAS: Rheumatoid arthritis (RA), Lupus and Inflammatory Bowel Disease. We found that BTS prioritized tissues, loci and variants that agree with the known or suspected biological mechanisms of these diseases. For instance, in the RA GWAS, 80% of the annotations prioritized by BTS are in immune tissues or organs, mainly T and B cells. Further, BTS prioritized loci near genes such as MICA/MICB, TAP1, TSBP1 and the HLA genes. 159 of these loci overlap enhancers in immune organs or cell types, and 35 overlap DNase hypersensitivity sites. Among the loci overlapping immune enhancers, 54% are immune-specific, while the others are active across multiple tissue categories. Overall, using BTS for unbiased prioritization of GWAS results and their functional contexts can inform follow-up studies and suggest mechanisms and therapeutic targets for future research.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3329. Building ancestry-informed tools for more inclusive genomics and improved clinical interpretation in admixed populations

Authors:

E. Atkinson¹, T. Tan¹, Y. Wang², J. Mauer³, N. Shah¹, G. Tietz¹, K. Tsuo⁴, M. Kamariza⁵, A. Martin², M. Santoro⁶; ¹Baylor Coll. of Med., Houston, TX, ²Massachusetts Gen. Hosp., Boston, MA, ³Univ.e Federal de Sao Paulo, Sao Paulo, Brazil, ⁴Harvard Univ., Cambridge, MA, ⁵Harvard Univ., Boston, MA, ⁶Federal Univ. of Sao Paulo, Sao Paulo, Brazil

Abstract Body:

Genetic studies offer a promising basis for understanding pathophysiology and identifying new molecular targets for medicine. Recent landmark papers have made major strides in our understanding of the genetic architecture of complex disorders, however, these studies are based overwhelmingly on subjects of European ancestry and their results may not fully generalize to other populations. Due to the paucity of methodological and computational approaches that account for their additional genomic complexity, admixed populations, including African American and Latinx individuals, make up more than a third of the US populace, yet face severe disparities in health research and treatment due to being systematically excluded from statistical genomic studies. To reap full and equitable benefits from existing mixed ancestry cohorts and ongoing large-scale data collection efforts, there is an urgent unmet need for the development of tools facilitating the well-calibrated study of complex psychiatric traits in admixed peoples alongside the recruitment of more diverse study participants.

We recently released a novel local-ancestry aware GWAS model, Tractor, which corrects for fine-scale population structure at the genotype level, better localizes GWAS signal, identifies ancestry-specific loci that would have been missed with standard procedures, and produces ancestry-specific effect size estimates and p values. Building off of this work, we present novel computational strategies for ancestry-informed gene discovery, including a statistical method for rare variant association studies in admixed cohorts, the creation of local ancestry informed polygenic risk scores that are more accurate in their prediction for diverse populations, and a strategy leveraging the naturally varying haplotype structure in admixed cohorts to better understand gene-gene interactions at clinically meaningful loci. We benchmark these new computational tools in simulation and using the large and diverse All of Us Project Dataset. This work fills a gap in existing resources and will improve our understanding of complex diseases across understudied populations. The inclusion of diverse participants in genomics efforts directly improves understanding for these individuals who bear a disproportionate health burden, as well as offering the opportunity to expand and accelerate findings relevant for individuals of all ancestries. Our work will lead to a more complete representation of genetic risk across the allele frequency spectrum and across populations, and strives for more equitable genomics research and health outcomes.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3330. Burden testing of imputed rare variants to inform therapeutic hypotheses

Authors:

Z. Fuller, J. O'Connell, A. Kwong, G. C. Partida, K-H. Lin, A. Auton, X. Wang; 23andMe, Sunnyvale, CA

Abstract Body:

A fundamental goal of human genetics is to identify variants in patients that underlie disease. Towards that end, several recent studies have demonstrated the utility of investigating the association between the burden of rare loss-of-function (LoF) mutations in genes and health-related traits to inform potential therapeutic targets. Because such mutations exist at the rare end of the frequency spectrum, these studies have typically relied on whole exome sequence data to call variants with extremely low minor allele counts. While exome sequencing cohorts now approach sample sizes on the order of hundreds of thousands of individuals, burden testing remains underpowered to detect associations for weak or moderate effect sizes and diseases with low prevalence. Moreover, larger sample sizes are still required to observe compound heterozygote or homozygote LoF carriers in the majority of genes and heterozygous carriers in shorter genes or those under stronger selective constraint. Here, we show we can reliably impute rare variants at frequencies of $1e-4$ or lower in over 6 million samples from the 23andMe, Inc. research cohort. Using these imputed rare variants, we replicate associations between rare predicted LoF mutations and disease traits in gene-based burden tests from previous whole-exome studies and discover novel gene-trait associations. We demonstrate how burden tests can help inform directionality and provide additional evidence for or resolve conflicts from eQTL signals. Together, these results highlight the promise of imputed rare variants in large sample sizes to identify novel gene-trait associations and provide estimates of effect direction to inform therapeutic hypotheses.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3331. Calcium signaling in Alzheimer's disease: Insights from *PTK2B* and *miR146a* interaction analyses.

Authors:

A. Yashin, D. Wu, K. Arbeev, O. Bagley, H. Duan, R. Holmes, S. Ukraintseva; Duke Univ., Durham, NC

Abstract Body:

Background and objective. Several recent studies emphasize an important role of *PTK2B* and *miR142a* genes in Alzheimer's disease (AD). *PTK2B* is also known for participation in calcium induced regulation of ion channels, and *miR146a* for its involvement in regulation of neuroinflammation. These and many other properties of *PTK2B* and *miR146a* resulted from experimental studies with laboratory animals and cellular cultures. We hypothesize that interactions of SNPs from the *PTK2B* gene and *miR146a* gene is associated with AD in humans. The objective of this study is to test this hypothesis using human data. **Data and method.** To test this hypothesis, the data from Cardiovascular Health Study (CHS), Framingham Health Study (FHS) and Health and Retirement Study (HRS) were analyzed using logistic regression model with the interaction term. Linkage disequilibrium among SNPs from corresponding genes was used in clumping procedure ($R^2=0.1$) to reduce the number of tests and increase the Bonferroni correction threshold (BF) in analyses of association of SNP-by-SNP interactions with AD. **Results.** Analyses showed that interaction of SNPs rs751019 from the *PTK2B* gene and rs2910163 from the *miR146a* gene is significantly associated with AD ($\beta= 0.42$, $p=1.16E-02$, $BF=5.00E-02$). Further analyses showed that the interactions of rs751019 SNP from the *PTK2B* gene and two SNPs from the *miR146a* gene: rs2910164 and rs2961920 SNPs are also statistically associated with AD in CHS data ($\beta= 0.43$, $p=1.21E-02$, for rs751019 and rs2910164 pair and $\beta= 0.40$, $p=1.76E-02$ for the rs751019 and rs2961920 SNP-pair). Note that rs2961920 and rs2910163 are in LD with rs2910164 SNP and can serve as proxy for this SNP. To replicate these findings, we performed interaction analysis of these genes using HRS data. This analysis did not find rs751019 among pairs of interacting SNPs significantly associated with AD. However, it detected nominally significant association of interaction of SNPs rs1776858 from the *PTK2B* gene and rs2961920 with AD: $\beta= 0.39$, $p=6.90E-03$. It turns out that SNPs rs751019 and rs1776858 are in LD with $D'=0.70$ and $R^2=0.46$. Analysis of FHS data confirmed statistical significance of association of interactions between rs751019 and each of three SNPs from the *miR146a* gene with AD, however, the association effect is opposite to that obtained in CHS data. **Conclusion.** Obtained results elucidate a mechanism that may link disruptions in calcium signaling with AD in humans. The presence of flip-flop effects in the results of interaction analysis of CHS and FHS data may be explained by the presence of unobserved influential factors that differ in populations of study subjects of these studies.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3332. Calibrated and adaptive prediction interval improves polygenic score (PGS) performance in diverse populations

Authors:

Z. Xu¹, K. Hou¹, Y. Ding², B. Pasaniuc³; ¹Univ. of California, Los Angeles, Los Angeles, CA, ²UCLA, Los Angeles, CA, ³Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract Body:

Accuracy of PGS varies across individuals from different continental ancestral populations (Martin et al. 2019 Nat Genet), and individuals from different covariate groups, such as sex, age, and socioeconomic status even within a single ancestry group of individuals with European ancestries (Mostafavi et al. 2020 eLife). Such variable prediction accuracy stems from several factors, including differences in minor allele frequency, linkage disequilibrium tagging of causal effects, gene-environment interaction, and confounding factors captured in genome-wide association studies. These components lead to the limited portability of PGS to predict phenotypes at an individual level and weaken personalized usages of PGS. Here, we develop a new approach, calibrated polygenic score (CalPGS), to provide calibrated and individualized prediction intervals that account for variable prediction accuracy by covariate. CalPGS leverages a held-out calibration data set where both phenotypes and covariates are measured for each individual to generate prediction intervals for test individuals. In simulations and real data results, we find that CalPGS prediction intervals for all test individuals are well-calibrated (i.e. the true phenotype is contained in the interval at a pre-specified probability level) and each individual's credible interval size varies as a function of its covariates. Of note, CalPGS framework can generate prediction intervals for all existing PGS methods that produce point prediction. We use UK Biobank data to investigate the utility of CalPGS. First, we perform a comprehensive survey of covariate factors impacting PGS prediction accuracy across 110 publicly available PGS within individuals of European ancestries and of diverse ancestries in UK Biobank. We find that PGS accuracy stratifies across covariates encapsulating a wide range of socio-demographic factors (including deprivation index, income, education years) and genetic ancestries. For example, the prediction accuracy of education years can differ by up to 125% between individuals in the highest and lowest quintile group of income. Finally, with these identified covariate factors, we apply CalPGS to individuals from UK Biobank to demonstrate CalPGS provides calibrated and adaptive prediction intervals across diverse populations. CalPGS prediction intervals cover true phenotypes adaptively and consistently across traits and covariate groups at the pre-specified level of 95%, improving upon the under-coverage of baseline methods (as low as 80%).

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3333. CARMA: Novel Bayesian model for fine-mapping with high-dimensional functional data

Authors:

Z. Yang¹, C. Wang¹, A. Khan², B. Vardarajan¹, R. Mayeux¹, K. Kiryluk¹, I. Ionita-Laza¹; ¹Columbia Univ., New York, NY, ²Columbia Univ., NEW YORK, NY

Abstract Body:

We propose a novel Bayesian model for fine-mapping in order to identify putative causal variants at GWAS loci. Relative to existing fine-mapping methods, the proposed model has several appealing features, such as assuming a heavy-tail distribution on effect sizes, joint modeling of summary statistics and large number of functional annotations, and accounting for discrepancies between summary statistics and external linkage disequilibrium values in meta-analysis settings. Using simulations, we compare performance with commonly used fine-mapping methods, including fastPAINTOR and SuSiE, and show that the proposed model has smaller credible sets with high power when using in-sample LD, and lower FDR/higher precision and higher coverage for credible sets when using external LD. We further illustrate our approach by applying it to a meta-analysis of Alzheimer's Disease GWAS data where we prioritize putatively causal variants and genes, including NDUFS2, INPP5D, SORL1, BIN1, CD2AP, CASS4, PICALM, MS4A6A and others.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3334. Causal inference for the relationship between transcription and DNA methylation in breast cancer.

Authors:

B. Karki, J. Kvamme, A. Fu; Univ. of Idaho, Moscow, ID

Abstract Body:

DNA methylation, an epigenetic mechanism, plays an important role in transcription regulation and in complex diseases. Whereas methylation in gene promoters is known to be generally associated with silencing, the relationship between transcription and methylation in other parts of the gene is much less clear. Additionally, substantially different transcription and methylation profiles have been observed among breast cancer subtypes, but it is unclear whether and how these differences are influenced by different relationships between the two processes.

Here, we study the relationships between transcription and methylation in estrogen receptor positive (ER+) and negative (ER-) patients, using data from The Cancer Genome Atlas (TCGA) consortium. We formulated trios, each consisting of the copy number alteration (CNA) of a gene, expression (Exp) of this gene, and methylation (Me) of a site located near or in the same gene. Since CNA is prevalent in cancer, it is a highly effective instrumental variable for this causal inference. In each subtype, we further derived principal components from genomewide expression and methylation data, and identified those that are significantly associated with each trio as potential confounding variables.

We applied MRGN (see our other abstract on this method), a novel causal network inference method that accounts for many confounding variables under the principle of Mendelian randomization, to each of the 310,412 trios in each subtype. In 563 ER+ patients, we identified 39% of the trios to have more than one edge: specifically, 14% of all the trios are mediation ($CNA \rightarrow Exp \rightarrow Me$ or $CNA \rightarrow Me \rightarrow Exp$), 14.5% conditionally independent ($Me \leftarrow CNA \rightarrow Exp$), 3.1% the v-structure ($CNA \rightarrow Exp \leftarrow Me$ or $CNA \rightarrow Me \leftarrow Exp$), and 5.9% fully connected among the three nodes. In 187 ER- patients, since the sample size is much smaller, we identified many fewer trios (21%) with more than one edge: 9.5% are mediation, 8.8% conditionally independent, 1.4% v-structure, and 1.2% fully connected. Our analysis provides a first comprehensive picture of causal relationships between transcription and methylation in the two subtypes.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3335. Cell composition inference and identification of layer-specific transcriptional profiles with POLARIS.

Authors:

J. Chen, T. Luo, M. Jiang, J. Liu, G. P. Gupta, Y. Li; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract Body:

Spatial transcriptomics (ST) technology, which provides spatially resolved transcriptional profiles, facilitates advanced understanding of key biological processes related to health and disease. Sequencing-based ST technologies (exemplary platforms include Spatial Transcriptomics and Visium) are widely adopted because they can provide whole-transcriptome profiles as well as a hematoxylin and eosin (H&E) stained histology image. However, these technologies are limited by their resolution, examining spatial spots of 2-100 micrometer in diameter where each spot can easily contain tens of cells of different cell types. Lack of knowledge in the number of cells at each spot nor the cell type composition can lead to invalid downstream analysis, which is a critical issue recognized in ST data analysis. In addition, existing methods tend to under-utilize histological images, which conceptually provide important and complementary information including anatomical structure and distribution of cells. To fill in the gaps, we have developed POLARIS, Probabilistic-based cell compOSition inference with LAYer infoRmatIon Strategy, a versatile ST analysis method that can perform cell type deconvolution, identify genes differentially expressed across anatomical or functional layers, as well as enhance ST data resolution. POLARIS integrates single-cell RNA-seq reference and ST data with annotated layer information. By examining a histology image and its coordinated expression profiles, one can reasonably infer layers or sub-regions that correspond to different biological functions (e.g. cancer vs non-cancer regions in a tumor biopsy, different layers in brain cortex, ventricle and artery areas in heart). By explicitly allowing and modeling layer-specific expression patterns, POLARIS can not only identify cell type composition with high accuracy, but also detect layer-wise differentially expressed (LDE) genes while simultaneously correcting for differential cell composition. Key advantages of POLARIS include flexibility and ability to leverage histology images, which enables POLARIS to obtain super-resolution inference based on images in areas without transcriptomic measurements, and to predict cellular composition purely from H&E images. We apply POLARIS to mouse cortex, developing human heart, and Her2+ breast cancer samples. POLARIS robustly demonstrates high deconvolution accuracy compared to state-of-the-art deconvolution methods, accurately predicts cell composition solely from images, and identifies LDE genes that are biologically relevant and meaningful.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3336. Characterizing copy number variations and their association with plasma proteins using short and long-read sequencing

Authors:

D. Schmitz¹, Z. Li², N. Rafati³, A. Ameer¹, V. Lo Faro¹, Å. Johansson¹; ¹Dept. of Immunology, Genetics and Pathology, Sci. for Life Lab., Uppsala Univ., Uppsala, Sweden, ²Twelve Bio ApS, Copenhagen, Denmark, ³Dept. of Med. Biochemistry and Microbiol., Natl. Bioinformatics Infrastructure Sweden, Sci. for Life Lab., Uppsala Univ., Uppsala, Sweden

Abstract Body:

Introduction

Structural variations (SVs), including copy number variations (CNVs), affect around 20 million bases in the human genome. CNVs that include regulatory or protein-coding regions could dramatically influence the expression or function of the encoded protein. Traditional methods used for CNV detection are array-based, suffer from poor resolution and lack the ability to accurately detect breakpoints. This study aims to identify CNVs in a Swedish cohort and explore the effect of CNVs on plasma proteins, using whole-genome sequencing (WGS) data for unbiased assessment of CNVs.

Methods

A total of 1,021 individuals were sequenced using Illumina next-generation sequencing (30x coverage), and disease-related 438 plasma proteins were measured. We identified CNVs using CNVnator, and summarized the detected CNVs into a cohort matrix. Associations between polymorphic CNVs, identified in at least three individuals, and protein levels were identified using linear regression analyses. We validated a selection of polymorphic copy number variable regions (CNVRs) in 15 individuals using long-read PacBio SMRT sequencing.

Results

A total of 184,183 non-overlapping polymorphic CNVRs were detected in the cohort. Among 872 individuals with CNV and protein data available, we identified 17 CNVRs to be associated with 17 plasma proteins after Bonferroni correction ($p < 6.19 \times 10^{-10}$). Of these CNVRs, six could be validated using long-read sequencing, including a CNVR within the V-set and transmembrane domain-containing protein 1 (*VSTMI*) gene, which was associated to measurements of Osteoclast-associated immunoglobulin-like receptor (OSCAR). Two other CNVRs appeared to be clusters of many short repetitive elements and one a common complex inversion. The remaining eight CNVRs could not be validated due to lack of evidence or low coverage.

Conclusions

Our findings provide insight into the involvement of CNVs on plasma proteins as well as the application of WGS approaches using both short and long reads for CNV detection.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3337. Characterizing features affecting Local Ancestry Inference in Latin American populations

Authors:

J. Mauer¹, A. Maihofer², C. Nievergelt², S. Belangero¹, M. Santoro¹, E. Atkinson³; ¹Univ.e Federal de Sao Paulo, Sao Paulo, Brazil, ²Univ California San Diego, La Jolla, CA, ³Baylor Coll. of Med., Houston, TX

Abstract Body:

Latin American populations have a complex ancestry makeup resulting from historical admixture events from multiple continental areas. In recent years, there has been significant effort in improving methods for admixed populations using Local Ancestry Inference (LAI), including the development of the gene discovery method Tractor. Reference panels are broadly required for genomic pipelines, including these new ancestry-informed strategies, yet are sorely lacking for many diverse populations, particularly those who have some Amerindigenous (AMR) ancestry. Here, we comprehensively test analytic strategies for LAI using existing resources to observe characteristics that impact its performance with particular attention to three-way admixed populations reflective of Latin America. After simulating LD-informed admixed data under a variety of 2 and 3 way admixed Latinx demographic models, we implement and tune LAI pipelines (using RFMix v1) and quantify their accuracy rates. Specifically, we assessed 1) the impact of reference panel composition, 2) demographic features of the cohorts including proportions of major ancestry groups and number of generations since admixture, 3) genetic data technology (genotyping arrays versus sequencing, the impact of imputation), and 4) the analytic thresholds in pipelines such as window size. For reference panel testing, we tested reference panel combinations comprising individuals from the Human Genome Diversity Project and the Thousand Genomes Project in three settings: 1) a very well matched panel but with low sample size; 2) including admixed individuals in the reference versus restricting to homogeneous individuals; or 3) a large reference that is poorly matched. We benchmarked the compute time and quantified the accuracy of each reference panel combination, simulation model and software setting to assess their respective performance. We observe that Amerindigenous ancestry tracts suffer from notably reduced accuracy as compared to European and African tracts. When miscalls occur, LAI error rates are more frequent in the direction of calling European ancestry in Amerindigenous simulated sites than other error modes. We additionally distribute a pipeline for calculation of the LAI true positive rate that can be used for benchmarking options for reference panel composition across other user-specified demographic models. Though our investigations are directly responsive to realistic admixed Latin American cohort compositions, the trends we characterized will also be broadly useful to inform characteristics that influence local ancestry inference across diverse admixed populations.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3339. Colorectal cancer polygenic risk score is inversely associated with eye phenotypes

Authors:

E. Rosenthal¹, W-Q. Wei², Y. Luo³, B. Namjou-Khales⁴, D. Schaid⁵, E. D. Esplin⁶, L. Kottyan⁷, M. Lape⁸, J. A. Pacheco⁹, C. Weng¹⁰, A. S. Gordon¹¹, I. J. Kullo¹², D. R. Crosslin¹³, M. Thomas¹⁴, U. Peters¹⁴, L. Hsu¹⁴, G. P. Jarvik¹⁵; ¹Univ. of Washington Sch. of Med., Seattle, WA, ²Vanderbilt Univ, Nashville, TN, ³Dept. of Preventive Med., Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, ⁴Ctr. for Autoimmune Genomics and Etiology, Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, ⁵Dept. of Quantitative Hlth.Sci., Mayo Clinic, Rochester, MN, ⁶Invitae, San Francisco, CA, ⁷Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, ⁸Dept. of BioMed. Informatics, Univ. of Cincinnati Coll. of Med., Cincinnati, OH, ⁹Ctr. for Genetic Med., Northwestern Univ., Chicago, IL, ¹⁰Dept. of BioMed. Informatics, Columbia Univ., New York, NY, ¹¹Feinberg Sch. of Med., Northwestern Univ., Chicago, IL, ¹²Mayo Clinic, Rochester, MN, ¹³Div. of BioMed. Informatics and Genomics, John W. Deming Dept. of Med., Tulane Univ. Sch. of Med., New Orleans, LA, ¹⁴Publ. Hlth.Sci. Div., Fred Hutchinson Cancer Ctr., Seattle, WA, ¹⁵Div. of Med. Genetics, Sch. of Med., Univ. of Washington, Seattle, WA

Abstract Body:

The development of polygenic risk scores (PRSs) is an attempt to understand the polygenic component of risk for complex traits. Pairs of complex traits may share environmental and/or genetic components suggesting similar biological mechanisms. Analyzing the association between PRSs and traits in deeply phenotyped biobanks and cohorts may illuminate shared genetic components. We sought to elucidate what phenotypes may be associated with colorectal cancer (CRC) through a shared polygenic component by assessing the association between a CRC PRS and phecodes in the electronic medical records and genomics (eMERGE) phase 3 study (N= 87,271 adults after quality control; 12,034 African genetic ancestry, 2,221 Asian genetic ancestry, 73,016 European genetic ancestry). The PRS contains ~1M SNPs and was developed (N=120,184) and validated (N=72,791) in individuals of European ancestry (PMID: 32758450). As the eMERGE data contains relatives, we used a linear mixed model approach that adjusts for relatedness among the participants and included the covariates age, sex, and first four principal components of ancestry (<https://github.com/eamosenthal/MMPheWas>). We defined significance using a Bonferroni correction: $p < 0.05 / (\text{Number top level phecodes}) = 9.2e-5$. We replicated the positive association with CRC ($p=1.2e-64$), as expected. We detected a negative association between CRC PRS and eye phenotypes, including poor visual acuity ($p=4.4e-17$; myopia subphenotype $p=9.2e-17$), cataract ($p=5.2e-7$), retinal disorders ($p=5.3e-7$) and corneal disorders ($p=4.6e-6$). The associations with eye phenotypes remained for controls without CRC (N=45,020, $p<2.8e-5$), indicating that the association is not due to confounding with CRC, such as earlier death. We replicated this finding in unrelated UK Biobank European ancestry participants (N= 433,767), using a linear model analysis adjusting for age, sex and the first four principal components of ancestry, finding a negative association between the CRC PRS and myopia ($p=0.006$). The finding of the same PRS predicting both CRC and eye phenotypes supports the need for objective discovery approaches.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3340. Combining family history with a polygenic risk score enables highly effective risk stratification for glaucoma

Authors:

S. Macgregor¹, W. HE², Y. Wu¹, D. A. Mackey³, J. E. Craig⁴, A. W. Hewitt⁵, P. Gharahkhani¹; ¹QIMR Berghofer Med. Res. Inst., Herston, Australia, ²QIMR Berghofer Med. Res. Inst., Brisbane, Australia, ³Lions Eye Inst., Perth, Australia, ⁴Flinders Univ., Adelaide, Australia, ⁵Univ. of Tasmania, Hobart, Australia

Abstract Body:

Glaucoma is a leading cause of irreversible blindness. Timely diagnosis is key as current treatments cannot restore lost vision. However, the disease often goes unnoticed in its early stages (e.g. 50% non-diagnosis rate, even in developed countries) and all too frequently diagnosis comes after retinal ganglion cell damage has occurred. As glaucoma only affects around 3% of people over 50 years of age, most people are not at high disease risk and it is currently not feasible to screen the whole population. We sought to explore population subgroups with elevated risk who may benefit most from increased surveillance.

We explored the utility of two predictors of increased disease risk: family history and polygenic risk scores (PRS). To explore family history we used data from a Canadian cohort study (CARTaGENE, N=29222 with GWAS data) where participants reported if they had a sibling affected by glaucoma. For the PRS we computed a previously published score (2020) and a new score derived from a multitrait analysis of a large glaucoma GWAS together with GWAS data on two glaucoma risk factors (intraocular pressure and retinal nerve damage, assessed via a machine learning based approach).

In CARTaGENE 1% of participants reported having a sibling affected by glaucoma. Having an affected sibling was associated with a 6.2 fold increased risk of glaucoma compared to those without an affected sibling (proportion affected 17% vs 2.5%). Within the family history positive group, the new PRS (2022) improved prediction compared to the 2020 PRS. Each standard deviation increase in the new PRS was associated with an odds ratio of 1.9. Taken together, PRS + family history clearly stratified individuals into risk categories: in people with an affected sibling the proportion affected was 32% (SD 7%) in the top 20% of the PRS, but only 5% (SD 4%) in the bottom 20%. Future studies should expand the relative types considered as well as examine reporting biases in glaucoma.

Many guidelines prioritize individuals for glaucoma surveillance based on family history but including PRS would greatly improve the allocation of finite resources. Low PRS individuals with a family history of glaucoma have risk similar to the general population, whilst those with a high PRS are at extremely high risk.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3341. Common and Rare Variants Genetic Association Analysis of Late-Onset Alzheimer's Disease in 500,000 Exomes from UK Biobank

Authors:

Y-C. Hwang, A. Fungtammasan; DNAnexus, Mountain View, CA

Abstract Body:

Late-onset Alzheimer's Disease (AD) has been causing more than 1 in 9 people aged 65 and older with dementia. The lifetime cost of care for a person with Alzheimer's is expensive and there is no effective cure. The disease can be caused by genetic, environmental, or other unknown factors. A well-known genetic variant of a person's risk for AD is the variant on the *APOE* gene. However, the specific gene set that directly causes AD are not found yet.

In this study, we performed a genome-wide association study (GWAS) on AD using UK Biobank (UKB) data to discover variants in genes that are associated with AD. UKB is a large prospective population-based initiative. It has collected extensive phenotypic data of half a million participants, aged 40-69 years, from the UK. The resource also includes various genetic data, including genome-wide genotype array, whole exome sequencing (WES), and whole genome sequencing (WGS) data for each participant.

Genetic variants are called per-individual level as well as at cohort-level of all participants. We used the cohort-level variants of the 500k participants of UKB, and derived a binary trait, as an individual's AD risk by proxy - based on each participant's ICD-10 code, age, and their parental diagnoses and ages. After quality controls of the genotype and phenotype data, we found 46,893 cases (AD) and 287,169 controls (non-AD). We applied REGENIE, a machine learning based method, for fitting the whole-genome regression model and testing for associations in a gene-based manner. Compared with Jansen, et al., we rediscovered variants associated with AD, including variants in *APOE*, and extended to more variants and genes. The entire analysis took less than 6 hours of wall-clock time, which is made possible by both the efficiency of the algorithm, the method, and the scalability of the DNAnexus cloud environment.

We further streamlined this computational-intensive analysis flow as a pipeline, which allows us to scale-up the analysis to higher-dimensions such as WGS cohort, and find associations to dozens of other phenotypes and traits of interest. The entire analysis is performed on the cloud-based UKB Research Analysis Platform and under UKB application number '46926'.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3342. COMPADRE: a robust, cloud-optimized genetic relatedness platform

Authors:

G. Evans¹, L. E. Petty², A. S. Petty², J. T. Baker¹, H-H. Chen², R. J. Bohlender³, C. D. Huff³, J. E. Below²; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ Med Ctr., Nashville, TN, ³Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract Body:

Scalable, robust methods that enable accurate relationship estimation in the absence of known pedigree structure are critical for large-scale genetic research such as disease gene mapping and linkage analysis. Despite their growing importance, available tools fail or exhibit biased results in the presence of population structure (e.g., recent admixture), data heterogeneity (e.g., comprising sequence data of varying read depth, libraries, or multiple arrays), and diversity (e.g., datasets spanning multiple distinct populations). We propose a multifaceted approach to improve and combine three existing programs - ERS (Estimation of Recent Shared Ancestry), a program that estimates pairwise relationships from shared genomic segments, PRIMUS (Pedigree Reconstruction and Identification of the Maximum Unrelated Set), a program that utilizes identity-by-descent (IBD) proportions to reconstruct pedigrees, and PADRE (Pedigree-Aware Distant Relatedness Estimation), a program that estimates distant relatedness across pedigrees using shared segments - into a new platform, COMPADRE (COMBined PADRE). COMPADRE's streamlined infrastructure reduces computational footprint (20%) and reconstruction runtime (35%) and improves the proportion of completed reconstructions previously affected by family network assessment errors (25%) relative to PRIMUS. It supports minimal imputation for cross-platform data harmonization and has increased relationship confidence in pedigree reconstruction (>10%) in early iterations by leveraging shared segments' length and distribution to reduce second- and third-degree relationship estimation ambiguity. Moreover, we have built a robust pipeline to generate simulated genetic data en masse reflecting key difference conditions (admixture, diverse populations, heterogeneity, and varied relatedness degree) to both supplement COMPADRE benchmarking relative to existing tools (Bonsai, DRUID, PRIMUS, PADRE) and serve as a community resource. Leveraging a novel multiprocessing framework and on-demand reference to 1000 Genomes Project data, this pipeline enables parallelized, highly specific, and interoperable genotype data generation that is significantly faster (>10x) and more efficient than prior implementations.

We have validated our findings in Vanderbilt University's biobank (BioVU) and are expanding our characterization analyses to UKBioBank and the National Institutes of Health's All of Us dataset. Public availability of these results will ensure appropriate relatedness utilization from key DNA biobank resources, bias reduction, and disease gene discovery power improvements.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3343. Comparing methods to adjust for fine-scale population structure in rare variant analyses.

Authors:

K. Marker^{1,2}, R. Shemirani³, M. Lin¹, E. Kenny³, G. Belbin³, C. Gignoux^{1,2}; ¹Colorado Ctr. for Personalized Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ²Human Med. Genetics and Genomics Program, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ³Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

When there is a difference in average trait value or disease prevalence between subpopulations, fine-scale genetic differences in rare variants can induce population structure that cannot always be corrected with traditional approaches. This is increasingly a problem as very large sample sizes are needed to detect rare variant associations. Datasets from multiple studies will often be combined or external controls will be used to increase sample size. Fine-scale ancestry differences between cohorts will induce population structure bias that can confound rare variant association studies if the fine-scale population structure is not properly adjusted for. We compare four methods for capturing and adjusting for fine-scale population differences: including identity-by-descent (IBD) clusters, kinship clusters, PCA using only rare variants, and Uniform Manifold Approximation and Projection (UMAP) projections. To accomplish this, we have developed a simulation framework to generate 9 populations from a continental European demographic model with stepping stone migration using msprime. We first simulate base populations and then use those populations to simulate the derived populations with migration between the populations. The ancestry differences within these populations replicate fine-scale ancestry found within the White British population in the UK Biobank. Using a phenotype simulator, APRICOT, we simulate phenotypes under varying geographic conditions to induce fine-scale population structure. Finally, we run rare variant associations using these simulations to compare adjustment by IBD clusters, kinship clusters, rare variant PCA, and UMAP projections.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3344. Comparing the relationships of genetically proxied PCSK9 inhibition with mood disorders, cognition, and dementia between men and women: a drug-target Mendelian randomization study

Authors:

A. Bell¹, D. B. Rosoff^{1,2}, L. A. Mavromatis¹, J. Jung¹, J. Wagner¹, F. W. Lohoff¹; ¹NIH, Bethesda, MD, ²NIH-Oxford-Cambridge Scholars Program; Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom

Abstract Body:

Proprotein convertase subtilisin/kexin 9 (PCSK9) inhibitors (PCSK9i) are important therapeutic options for reducing cardiovascular disease risk; however, questions remain regarding potential differences in the neuropsychiatric impact of long-term PCSK9 inhibition between men and women. While randomized controlled trials (RCTs) investigating these effects may be burdensome and costly, PCSK9 has a well-characterized genomic profile which can be used to assess the genetic basis of its physiological effects. We therefore used single-nucleotide polymorphisms (SNPs) in PCSK9 from predominantly European ancestry-based genome-wide association studies (GWAS) summary-level statistics of low-density lipoprotein cholesterol (LDL-C), sex-specific GWASs of anxiety, depression, cognition, and dementia, and drug-target Mendelian randomization (MR) methods to investigate potential neuropsychiatric consequences of inhibiting PCSK9 in men and women. Broadly, we failed to find genetic evidence that PCSK9 inhibition was related to risk for 13 neuropsychiatric endpoints in either men or women. We observed some isolated, nominally significant effect estimates for certain outcome-sex combinations; however, none passed correction for multiple comparisons. Further, corresponding MR analyses using polygenic instruments for LDL-C, the main biological target of PCSK9i, failed to find an impact of LDL-C on depression risk among women. Other mood-related endpoints were also null, including extreme irritability; self-harm; and panic attacks. Similarly, PCSK9 inhibition was not associated with sleep disturbances, dementia, or cognition. Our results indicate that genetically proxied PCSK9 inhibition displays a neutral neuropsychiatric profile with no major sex-specific differences and underscores the importance of genomic techniques such as MR for investigating potential adverse outcomes without requiring human experimentation. These findings may have important clinical implications for lipid-lowering drug prescribing practices and side effect monitoring.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3345. Comparison of trans-ancestry meta- and mega-analyses of cis-eQTLs in whole blood and LCLs

Authors:

P. Nandakumar¹, E. Bullis¹, T. Cavazos², B. Hicks¹, D. Hinds¹, K. Kukar¹, Y. Liang¹, S. Micheletti¹, M. Moreno¹, J. O'Connell¹, A. Petrakovitz¹, S. Pitts², R. Sauteraud², A. Sugathan², C. Wong¹, the 23andMe Computational Biology Team, the 23andMe Research Team, K. Fletez-Brant², V. Vacic²; ¹23andMe, Inc., Sunnyvale, CA, ²23andMe, Inc. Therapeutics Div., South San Francisco, CA

Abstract Body:

Expression quantitative trait loci (eQTL) studies are useful for variant-to-gene mapping and mechanistic interpretation of signals from genetic association studies. Unfortunately, tissue samples are difficult to obtain and as a result, eQTL studies are almost always underpowered. One way to increase sample sizes is to combine individual eQTL studies; however the choice of framework to use is not obvious. To address this question, we conducted a trans-ancestry fixed-effects meta-analysis and mega-analysis in each of whole blood (n=2,712 across three datasets) and lymphoblastoid cell line (LCL; n=1,164 across three datasets) data, in cohorts of primarily African and European ancestries. For the meta-analyses, we combined cis-eQTL 1Mb window statistics in each study produced using FastQTL for each dataset, with each individual study adjusted for age, sex, genetic principal components (PCs), and 15-35 PEER factors, as recommended by the authors. For the mega-analyses, which, with combination of individual-level data, have the benefits of identifying secondary signals and lower MAF eQTLs, we merged the individual-level genotype and expression datasets, and ran the FastQTL linear model adjusting for age, sex, global genetic PCs, 35 global PEER factors, and a study covariate. Considering the intersecting set of variant-gene pairs for comparison, while both the meta- and mega-analyses demonstrate power improvements compared to the individual datasets, the meta-analyses were better powered than the mega-analyses (LCL eGene counts: 10,744 in meta-analysis, 10,357 in mega-analysis; whole blood eGene counts: 12,619 in meta-analysis, 11,071 in mega-analysis), with similar effect sizes but smaller standard errors. We then investigated if additional PEER factors would improve the results of mega-analysis and computed 50 to 200 PEER factors in increments of 50 for 1000 randomly selected top variant-eGene pairs from the initial mega-analyses, and compared the overall AIC improvement with increasing PEER factors. We observed that 100 PEER factors for LCL data and 200 PEER factors for whole blood appeared optimal among the tested models. The final mega-analyses including the larger numbers of PEER factors produced roughly comparable numbers of eGenes as the meta-analyses (LCL: 11,260; whole blood: 12,627) but at the expense of significantly larger computational cost. Our results suggest that meta-analysis is a cost-efficient way to maximize power in eQTL studies; however, if longer PEER run times are not an issue, mega-analysis allows interrogation of eQTLs with smaller MAFs.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3346. Constructing and benchmarking omnibus polygenic risk scores using GWAS summary statistics

Authors:

Z. Zhao¹, J. Song¹, J. Miao², Z. Sun¹, G. Song¹, Q. Lu¹; ¹Univ. of Wisconsin-Madison, Madison, WI, ²Univ. of Wisconsin-Madison, Madison, WI

Abstract Body:

Recent advancements in genome-wide association studies (GWASs) and polygenic risk score (PRS) methodology exhibit promising improvements in prediction accuracy for many complex traits. However, existing methods for PRS model fine-tuning and performance benchmarking require individual-level genotype and phenotype data independent from training GWAS dataset in order to avoid overfitting. These data can be difficult to obtain in practice due to limited data availability and the equally important need to maximize the training GWAS sample size. Therefore, it has been a long-standing challenge in the field to optimize and benchmark PRS models without individual-level data. To address this challenge, we introduced PUMAS that performs cross-validation to fine-tune PRS models using GWAS summary data. Here, we introduce two important extensions to this framework. First, PUMAS now accounts for linkage disequilibrium (LD) and thus can be applied to almost all PRS models without requiring SNP pruning. Second, we introduce a highly innovative strategy to combine multiple PRS models into an omnibus score using GWAS summary data alone. Through extensive simulations, we show that our summary statistics-based approach (PUMAS) has consistent and robust performance compared to Monte Carlo cross-validation based on individual-level genotype and phenotype data. Applied to 16 complex traits in the UK Biobank, PUMAS can accurately pinpoint the best tuning parameters for LDpred2, PRS-CS, and lassosum, and also closely approximate predictive R^2 acquired from external validation. The omnibus PRS, obtained through aggregating multiple fine-tuned PRS models, outperforms single PRS models without requiring additional data. We demonstrate that our approach achieves highest prediction accuracy for all 16 complex traits in UK Biobank, with an average increase of 7.01% in R^2 compared to the best performing PRS method for each trait. Finally, we apply our method to 85 well-powered GWAS with publicly available summary statistics and provide an atlas of fine-tuned and omnibus PRS. Using PRS-CS as the baseline, our omnibus PRS approach shows an average 22.6% increase in predictive R^2 , showcasing substantial improvements compared to the state of the art. Taken together, PUMAS makes it possible to comprehensively evaluate and optimize PRS prediction accuracy across a plethora of PRS methods for hundreds of GWAS, and will likely have broad and impactful influences over how PRS models are constructed in the future.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3347. Controlled Discovery and Localization of Causal Variants via Bayesian Linear Programming

Authors:

A. Spector¹, L. Janson²; ¹Stanford Univ., Stanford, CA, ²Harvard Univ., Cambridge, MA

Abstract Body:

Identifying genetic variants which cause complex traits is a crucial task in statistical genetics. However, high linkage disequilibrium in genetic loci can make it challenging or impossible to isolate causal variants even when a causal variant clearly exists somewhere within a locus. A popular alternative is to output small regions of the genome, called credible sets, which each contain at least one causal variant with high confidence (Wang et al., 2020). Such analyses aim to (1) discover as many credible sets as possible while (2) minimizing the size of each credible set, all while (3) controlling the false discovery rate (FDR).

With this motivation, we introduce Bayesian Linear Programming (BLiP), a Bayesian method for detecting and localizing causal variants. As an input, BLiP can take nearly any Bayesian fine-mapping model, such as SuSiE, allowing it to incorporate sophisticated models and arbitrary prior information based on (e.g.) functional annotations. Given such a model, BLiP yields a set of credible sets for causal variants which verifiably maximize a resolution-adaptive notion of statistical power, meaning intuitively that BLiP makes the largest number of discoveries at the finest possible resolution while provably controlling the FDR. BLiP is extremely computationally efficient and in simulations, it often dramatically improves the power of the underlying model and can even simultaneously reduce the false discovery rate. Furthermore, applying BLiP can be as easy as running one or two lines of code, as we have released two open-source packages `pyblip` (Python) and `blipr` (R) implementing BLiP.

We applied BLiP to analyze four complex traits (height, cardiovascular disease, LDL cholesterol and HDL cholesterol) on a sample of ~318,000 individuals of British ancestry from the UK Biobank. In particular, we ran BLiP on top of a SuSiE model previously fit by Weissbrod et al. (2020). Despite requiring less than one minute of additional computation per trait, SuSiE + BLiP had 30-50% higher power than SuSiE alone. In particular, SuSiE + BLiP made discoveries at a finer resolution (15-25% finer on average) than SuSiE alone and furthermore made 15-20% more discoveries per trait. Of the new discoveries made by SuSiE + BLiP, we found that 45-65% could be corroborated by a separate study in the NHGRI-EBI GWAS Catalog, giving additional evidence that BLiP meaningfully enhanced SuSiE's power to find real causal variants. Overall, BLiP is a promising method to detect and localize causal variants which can wrap on top of nearly any Bayesian fine-mapping model with minimal computational cost.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3348. Copy-number variants as modulators of common disease susceptibility.

Authors:

C. Auwerx, N. Tesio, C. Clark, A. Reymond, Z. Kutalik; Univ. of Lausanne, Lausanne, Switzerland

Abstract Body:

We previously called copy-number variants (CNVs) in 331'522 UK Biobank participants, identifying 131 associations across 47 quantitative traits. Here, we interrogate whether CNVs also affect susceptibility to 60 common manually curated ICD10-based clinical diagnoses. We adapted our pipeline to accommodate the low number of disease cases and CNV carriers by complementing Firth-fallback logistic regression with various additional tests (i.e., Fisher tests, residual regression). To account for different modes of CNV action, genome-wide associations scans (GWAS) were performed according to mirror, U-shape, duplication-only, and deletion-only models, resulting in 309 independent signals across 221 CNV regions (CNVRs) and 54 diseases. Involved CNVRs were depleted from highly constrained genes ($p_{pLI} = 9.2e-5$; $p_{LOEUF} = 2.7e-4$) and all increased disease risk. We identified 71 signals overlapping a single protein-coding gene, recovering well-known associations, e.g., *BRCA1* and *LDLR* deletions increased the risk of ovarian cancer ($OR > 27$; $p = 1.9e-7$) and ischemic heart disease ($OR > 10$; $p = 1.2e-6$), respectively. Conversely, 73 signals mapped to intergenic or non-coding RNA (ncRNAs) regions, suggesting that CNVs also impact disease risk through disruption of regulatory mechanisms and/or modulation of ncRNAs. Some signals were corroborated by single nucleotide polymorphism (SNP)-GWAS signals for the same disorder, such as a 7p21.3 duplication overlapping a recently published Alzheimer's disease locus ($OR > 81$; $p = 1.7e-7$). Other signals were supported by colocalization with CNV signals for disease-relevant biomarkers. For example, our analysis revealed that altered dosage of 17q12 (*HNF1B*), which we previously found to alter renal biomarkers, increased chronic kidney disease risk ($OR > 9$; $p = 8.7e-9$), additionally exposing a role in hepatic fibrosis ($OR > 35$; $p = 1.4e-6$). We then explored the pleiotropic effect of disease-associated CNVRs. While 76% of the CNVRs associated with a single disease, up to 9 associations were found for the 16p11.2 BP4-BP5 CNVR, including know associations with schizophrenia and anemia, as well as novel ones with cardiac, pulmonary, renal, and immune disorders. Twelve CNVRs associated with the total number of diagnoses (burden), including recurrent deletions of 16p11.2 BP4-BP5, 22q11.2, and 16p12.2, the latter being associated with hypertension ($OR > 2.2$; $p = 1.6e-10$), cardiac conduction disorders ($OR > 2.6$; $p = 1.2e-10$), and ischemic heart disease ($OR > 2.2$; $p = 2.7e-6$), suggesting a crucial role in cardiac health. Together, these results shed light on the understudied role of CNVs in common diseases within the general population.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3349. Credible set determination for multi-ancestry fine-mapping.

Authors:

J. Shen, A. Wang, F. CHEN, C. Haiman, D. Conti; Univ. of Southern California, Los Angeles, CA

Abstract Body:

Trans-ethnic genetic studies may increase power to detect novel risk variants and improve fine-mapping resolution by leveraging evidence from diverse populations and accounting for the difference in linkage disequilibrium (LD) across ethnic groups. Here, we expand upon our previous approach for single and multi-population fine-mapping through Joint Analysis of Marginal SNP Effects (mJAM) to develop an approach for determining credible set SNPs. The approach utilizes mJAM to use population-specific summary statistics to fit a single conditional model that incorporates corresponding ancestry-specific reference linkage disequilibrium (LD), and fixed-effects meta-analysis of joint effects. The mJAM framework can be used to first select index SNPs using any feature selection approach, such as forward selection, Bayesian selection, or regularized regression techniques. Then, given a set of index SNPs within each region, the posterior credible set probability (PCSP) of a SNP is defined as a combination of two probabilities: one models the marginal association between the candidate SNP and the outcome; the other models the mediation effect of the index SNP on the candidate SNP, borrowing from a mediation framework. These PCSPs are then used to construct credible sets. We first illustrate mJAM through two implementations for selection: mJAM-SuSiE (a Bayesian approach) and mJAM-forward selection. We then compare these approaches to fixed-effect meta-analysis, COJO stepwise selection, and MsCAVIAR. When available, we also compare credible set performance. Through simulation studies based on realistic effect sizes and levels of LD, we demonstrate that mJAM performs better than other existing multi-ethnic methods for identifying index SNPs and corresponding credible sets that include the underlying causal variants. In a real data application, we apply this approach to the most recent summary statistics from a cross-ethnic prostate cancer GWAS.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3350. Cross-trait meta-analysis reveals shared genetic architecture between PCOS and chronic inflammation markers

Authors:

G. Brixi^{1,2}, L. Petersen^{1,3,1}, J. Canseco Neri⁴, J. Li¹, J. Hu^{5,6}, X. Han⁷, S. Mahalingaiah^{1,4}, L. Liang¹; ¹Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ²Harvard Univ., Cambridge, MA, ³Los Alamos Natl. Lab, Los Alamos, NM, ⁴Boston Univ., Boston, MA, ⁵Brigham and Women's Hosp., Boston, MA, ⁶Harvard Med. Sch., Boston, MA, ⁷Harvard T. H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract Body:

Polycystic ovarian syndrome (PCOS) is a common hormonal disorder with a prevalence of approximately 10% worldwide, and is the leading cause of female infertility. PCOS is known to be closely related to inflammatory markers, but the genetic and causal basis of this relationship is not yet understood. We report results of the largest-to-date meta-analysis of PCOS using data from UK Biobank, Partners Biobank, and prior studies of FinnGen, Estonian Biobank and additional cohorts (cases = 10,005, controls = 498,227). We further conduct meta-analyses for inflammation markers combining recent proteome and biobank cohorts (N ranges from 38860 to 505,690) to obtain the largest GWAS for a broad panel of 138 inflammation markers. We find statistically significant (FDR<0.05) genetic correlation between PCOS and CRP ($R_g = 0.35$, $P = 1.0 \times 10^{-5}$), leptin ($R_g = 0.4$, $P = 0.33 \times 10^{-5}$), white blood cell count ($R_g = 0.21$, $P = 8.0 \times 10^{-4}$), lymphocyte count ($R_g = 0.19$, $P = 1.1 \times 10^{-3}$), and SLAMF1 ($R_g = 0.52$, $P = 1.3 \times 10^{-3}$). We further find statistically significant genetic correlation between PCOS and SHBG ($R_g = 0.40$, $P = 1.1 \times 10^{-14}$), and BMI ($R_g = 0.34$, $P = 3 \times 10^{-13}$). Our findings highlight the genetic architecture underlying PCOS which is closely related to obesity and chronic inflammation. The results further our understanding of PCOS pathogenesis and enable our downstream analyses for causal inference between inflammation status and PCOS based on Mendelian randomization.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3351. *De novo* mutation hotspots in homologous protein domains identify function-altering mutations in neurodevelopmental disorders.

Authors:

J. Hampstead¹, L. Wiel¹, H. Venselaar¹, L. Vissers¹, H. Brunner¹, R. Pfundt¹, G. Vriend², J. Veltman³, C. Gilissen¹; ¹Radboud Univ. Med. Ctr., Nijmegen, Netherlands, ²Baco Inst. of Protein Sci., Baco, Philippines, ³Newcastle Univ., Newcastle-upon-Tyne, United Kingdom

Abstract Body:

Variant interpretation remains a major challenge in medical genetics. While *de novo* mutations (DNMs) are an established cause of approximately 50% of severe developmental disorders (DDs), the mechanisms by which these mutations cause disease are less well understood. Most genes are currently associated to DDs through an enrichment of protein-truncating variation; these genes are assumed to cause disease through haploinsufficiency, or the functional loss of an essential allele. However, in specific genes we observe that disease-causing missense DNMs cluster in functionally important positions; these mutations are thought to be function-altering, and cause disease through gain-of-function or dominant negative mechanisms. The contribution of function-altering mutations to developmental disorders is not well characterised, and these mutations are very difficult to identify due to their comparatively small mutational target.

To address this challenge, we developed a Meta-Domain HotSpot (MDHS) method that leverages protein domain homology to aggregate missense mutations across conserved protein consensus positions within domain families. This approach increases our statistical power to detect domain consensus positions likely to harbour function-altering variation contributing to DDs. We applied MDHS to a dataset of 45,221 DNMs from 31,058 patients with developmental disorders and identified three significantly enriched missense DNM hotspots in the ion transport protein domain family (PF00520). The 37 unique missense DNMs that drive enrichment affect 25 genes, 19 of which have been previously associated with DDs through protein-truncating variant enrichment. 3D protein structure modelling supports the function-altering effects of these mutations. DD genes with missense mutations at hotspot positions have a unique, brain-specific expression pattern in tissue, and we used this pattern alongside *in silico* predictors and population constraint information to identify candidate DD-associated genes. We also propose a lenient version of our method, which identifies 32 hotspot positions across 16 different protein domain families. These positions are enriched for likely pathogenic variation in clinical databases and DNMs in other genetic disorders. Overall, we provide compelling evidence that the aggregation of mutations over homologous protein domains is a scalable approach towards understanding the contribution of function-altering mutations to DDs.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3352. Deep Learning-based Phenotype Imputation on Biobank-scale Data Increases Genetic Discoveries

Authors:

U. An¹, A. Dahl², N. Cai³, A. Pazoki¹, M. Alvarez¹, S. Bacanu⁴, A. Schork⁵, K. Kendler⁴, P. Pajukanta¹, J. Flint¹, S. Sankararaman¹; ¹UCLA, Los Angeles, CA, ²Univ. of Chicago, Chicago, IL, ³Helmholtz Zentrum München, Neuherberg, Germany, ⁴Virginia Commonwealth Univ., Richmond, VA, ⁵Copenhagen Univ. Hosp., Copenhagen, Denmark

Abstract Body:

Biobanks that collect diverse phenotypic and genomic data across large numbers of individuals have emerged as a key resource for human genetics research. However, phenotypic measurements vary drastically in availability across individuals so that the ability to accurately impute or “fill-in” missing phenotypes in these datasets is critical to harness their power. To address a need for a reliable and fully-featured imputation method, we developed AutoComplete, a deep-learning based imputation method based on an auto-encoder architecture designed for highly incomplete biobank-scale phenotype data. AutoComplete can impute binary and continuous phenotypes simultaneously while scaling to datasets with half a million individuals and millions of entries. To handle the absence of ground truth in incomplete datasets, we developed a realistic procedure to simulate missing values as reflected in the real data by implementing copy-masking which propagates missingness patterns already present in the data, from which AutoComplete learns imputation.

AutoComplete greatly improved imputation accuracy over existing methods across a broad range of simulations of missing data. In comparison to the next-best method (SoftImpute), the average squared Pearson correlation coefficient r^2 of AutoComplete improved by 18% across a group of 230 phenotypes related to cardiometabolic conditions and a group of 370 phenotypes related to psychiatric disorders (11% and 25% within each individual dataset).

We demonstrate the utility of imputing highly missing phenotypes for the purpose of improving association power for genome-wide association studies (GWAS). As a verification of this procedure, we simulated highly observed phenotypes to be 67% missing and applied AutoComplete to recover them for GWAS. In total, 116 of 118 recovered significantly associated SNPs were also strongly associated with the originally observed phenotypes. We then imputed 8 highly missing phenotypes which were clinically relevant to psychiatric disorders. Imputing these phenotypes using AutoComplete lead to a five to ten fold increase in sample size, and we observed an increase in the total number of significantly detected loci across all phenotypes of interest from 4 to 173, with the mean increase being 21 loci (minimum increased 4 and maximum 37). Among phenotypes with notable increases in loci were Medication for MDD (from 1 to 14 loci) and Cannabis consumption frequency (from 0 to 23 loci). AutoComplete can scale to ~300,000 individuals and ~400 phenotypes from the UKBB with ease converging within ~6 hours. Our results illustrate the value of deep-learning based imputation to genomic discovery.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3353. Deep Representational Clustering for Genetic Endotype Discovery from Clinical Data

Authors:

A. Averitt^{1,2}, J. Mower¹, N. Banerjee¹, M. Cantor¹, Regeneron Genetics Center, GHS-RGC DiscovEHR Collaboration; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Columbia Univ., New York, NY

Abstract Body:

Endotypes are distinct, genetically motivated pathophysiological mechanisms of disease that play a critical role in precision medicine. We hypothesize that clinical data are reflective of the mechanism of disease and can be exploited to discover previously unknown endotypes. Typical methods of endotype discovery from clinical data (ex, hierarchical clustering or latent class mixture models) are often unable to accommodate different data modalities and are not suitable for use with high dimensional or noisy data. We propose a method of deep representational clustering of clinical data for endotype discovery. Our method employs a multi-headed autoencoder that (i) accommodates multi-modal data; (ii) denoises clinical features; and (iii) learns efficient representations of patients' states. A density-based algorithm then partitions the learned representations into clusters that correspond to endotypes. To evaluate the model, we applied it to electronic health record (EHR) data from Geisinger Health System to discover endotypes of Type 2 Diabetes (T2D) (N=10,751). To evaluate the learned endotypes, we characterized the clinical and genetic differences between clusters.

Our method identified 3 endotypes of T2D, referred to as #1 (N=7,238), #2 (N=2,271), and #3 (N=1,242). When compared to other T2D patients, #1 is characterized by *poorly controlled T2D* (ex, elevated blood glucose and HbA1c); #2 is characterized by *well-controlled T2D* (ex, normal HbA1c and glucose); and #3 is characterized by *cardiovascular comorbidities* (ex, hypertension and hyperlipidemia). Using case-only GWAS to test for endotype-specific genetic signal, we identified a genome-wide significant association between a common intronic variant within the FER gene and #1 (OR=1.44 [1.26, 1.64], $p=4.54e^{-8}$). FER has been previously associated with T2D and is thought to influence insulin action and glucose metabolism via adiponectin. We also found an association between a variant near the ROBO2 gene and #2 (OR=1.94 [1.54, 2.45], $P = 2.48e^{-8}$). This locus has been associated with T2D via pancreatic beta cell dysfunction. These distinct genetic signals across loci have been previously associated with T2D and provide support for the biological relevance of our EHR-derived endotypes. A review of functional relevance of the clusters' genetics is in progress.

This research presents a deep representation clustering method for endotype discovery from clinical data. When applied to real-world data of T2D patients, the model automatically learns phenotypically distinct clusters that demonstrate differing genetic architecture that is suggestive of differing mechanisms of disease.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3354. Deep sequencing of DNA from urine of kidney allograft recipients to estimate donor/recipient-specific DNA fractions

Authors:

A. Belkadi; Weill Cornell medicine in Qatar, Doha, Qatar

Abstract Body:

Kidney transplantation is the treatment of choice for patients with end-stage kidney failure, but transplanted allograft could be affected by viral and bacterial infections and by immune rejection. The standard test for the diagnosis of acute pathologies in kidney transplants is kidney biopsy. However, noninvasive tests would be desirable. Various methods using different techniques have been developed by the transplantation community. But these methods require improvements. We present here a cost-effective method for kidney rejection diagnosis that estimates donor/recipient-specific DNA fraction in recipient urine by sequencing urinary cell DNA. We hypothesized that in the no-pathology stage, the largest tissue types present in recipient urine are donor kidney cells, and in case of rejection, a larger number of recipient immune cells would be observed. Extensive in-silico simulation was used to tune the sequencing parameters: number of variants and depth of coverage. Sequencing of DNA mixture from 2 healthy individuals showed the method is highly predictive (maximum error < 0.04). We then demonstrated the insignificant impact of familial relationship and ethnicity using an in-house and public database. Lastly, we performed deep DNA sequencing of urinary cell pellets from 32 biopsy-matched samples representing two pathology groups: acute rejection (AR, 11 samples) and acute tubular injury (ATI, 12 samples) and 9 samples with no pathology. We found a significant association between the donor/recipient-specific DNA fraction in the two pathology groups compared to no pathology ($P = 0.0064$ for AR and $P = 0.026$ for ATI). We conclude that deep DNA sequencing of urinary cells from kidney allograft recipients offers a noninvasive means of diagnosing acute pathologies in the human kidney allograft.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3355. DeepPerVar: a multimodal deep learning framework for functional interpretation of genetic variants in personal genome

Authors:

Y. Wang^{1,2}, L. Chen³; ¹Biogen, Cambridge, MA, ²Indiana Univ. Sch. of Med., Indianapolis, IN, ³Univ. of Florida, Gainesville, FL

Abstract Body:

Understanding the functional consequence of genetic variants, especially the noncoding ones, is important but particularly challenging. Genome-wide association studies or quantitative trait locus analyses may be subject to limited statistical power and linkage disequilibrium, and thus are less optimal to pinpoint the causal variants. Moreover, most existing machine learning approaches, which exploit the functional annotations to interpret and prioritize putative causal variants, cannot accommodate the heterogeneity of personal genetic variations and traits in a population study, targeting a specific disease. By leveraging paired whole genome sequencing data and epigenetic functional assays in a population study, we propose a multi-modal deep learning framework to predict genome-wide quantitative epigenetic signals by considering both personal genetic variations and traits. The proposed approach can further evaluate the functional consequence of noncoding variants on an individual level by quantifying the allelic difference of predicted epigenetic signals. By applying the approach to the ROSMAP cohort studying Alzheimer's disease (AD), we demonstrate that the proposed approach can accurately predict quantitative genome-wide epigenetic signals and in key genomic regions of AD causal genes, learn canonical motifs reported to regulate gene expression of AD causal genes, improve the partitioning heritability analysis, and prioritize putative causal variants in a GWAS risk locus. Finally, we release the proposed deep learning model as a stand-alone Python toolkit and a web server.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3356*. Deep-PheWAS.R: an innovative R package for phenome wide association of novel phenotypes.

Authors:

R. Packer¹, A. T. Williams¹, W. Hennah^{2,3}, M. Eisenberg¹, K. A. Fawcett¹, W. Pearson¹, A. Guyatt¹, A. Edris^{1,4}, E. Hollox¹, B. S. Rao⁵, R. J. Bratty⁵, L. V. Wain¹, F. Dudbridge¹, M. Tobin¹; ¹Univ. of Leicester, Leicester, United Kingdom, ²Orion Pharma, Espoo, Finland, ³Univ. of Helsinki, Helsinki, Finland, ⁴Ghent Univ., Ghent, Belgium, ⁵Orion Pharma, Nottingham, United Kingdom

Abstract Body:

Background: Phenome-wide association studies (PheWASs) can be used to better understand the pleiotropic effects of genetic variants and to inform drug development. Current PheWAS approaches have several limitations. The phenotypes generated rely on a single data field or coding ontology and do not take advantage of all available data. Further there are no existing tools to develop new phenotypes. Online PheWAS resources, which allow fast retrieval of pre-existing results, only allow single nucleotide polymorphisms (SNPs) as inputs and retrieve results only for genetic models tested, whilst de novo per variant alternatives are computationally inefficient and can result in inflated type I error. **Methods:** Deep-PheWAS.R (DPR) addresses these limitations through enhanced phenotype development and flexible, efficient association testing. To study phenotypes not well captured by current classification trees, DPR creates novel clinically-curated composite phenotypes (combining primary and secondary care data), quantitative phenotypes from primary care data, disease progression phenotypes and drug response phenotypes. Existing classification structures are exploited by using Phecodes and individual data-field phenotypes. Optimised for use with UK-Biobank data, DPR can be adapted to other data sources and all phenotypes can be edited using in-built functions. Association is optimised for SNP inputs by integrating with PLINK2 and supports additive, dominant, recessive, or genotypic models. NonSNP inputs such as complex structural variants and genetic risk scores (GRS) are analysed through regression models in R. To highlight some of these features we applied DPR to rs7193778 (previously associated with urate levels) and a lung function GRS created using 279-variants. **Results:** DPR created 2246 phenotypes. Rs7193778 showed association (false discovery rate <1%) with 46 phenotypes, 13 at genome-wide significance, 12 of which are not reported on GWAS catalog. The strongest association was with blood sodium, a phenotype derived using measurement data from primary care ($P=6.49 \times 10^{-45}$, $BETA=0.07$). Our lung function GRS showed association with 47 traits including association with exacerbation of COPD and age-of-onset of COPD both of which are unavailable in existing PheWAS resources and have no published GWAS results on GWAS catalog. **Conclusion:** We present Deep-PheWAS.R, an R package that addresses several limitations of existing PheWAS approaches. This includes the ability to analyse more informative composite phenotypes, greater flexibility in the type of genetic variation that can be studied and assessing associations with genetic risk scores.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3357*. Defining longitudinal disease trajectories in 146,000 individuals with hypertension from Penn Medicine Electronic Health Records

Authors:

P. Singhal¹, L. Guare¹, A. Lucas², C. Morse³, M. Byrska-Bishop⁴, M. Guerraty³, D. Kim^{5,6}, M. Ritchie^{1,6,7}, A. Verma^{3,2}; ¹Dept. of Genetics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ²Corporal Michael Crescenz VA Med. Ctr., Philadelphia, PA, ³Dept. of Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ⁴New York Genome Ctr., New York, NY, ⁵Dept. of Biostatistics, Epidemiology, and Informatics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, ⁶Inst. of BioMed. Informatics, Univ. of Pennsylvania, Philadelphia, PA, ⁷Ctr. for Precision Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Many chronic diseases are associated with hypertension (HTN), yet we are currently unable to predict which future disease states are likely in an individual diagnosed with HTN. Interrogation of the comorbidity architecture of HTN using temporal individual-level electronic health record (EHR) data is needed to better predict the disease trajectory individuals will follow after a HTN diagnosis. Currently, there is no clear best way to model longitudinality in the EHR given irregular visits over the course of a patient's disease journey. This leads to the challenge of extracting meaningful temporal patterns at the health system and individual levels. We developed a framework to identify disease trajectories in Penn Medicine EHR data for a cohort of 146,654 individuals with HTN. We first generated a disease co-occurrence network using all ICD codes for each individual appearing after HTN onset (3,642 codes). The Ising Model, a probabilistic graphical model, was leveraged to estimate pairwise occurrence between each condition while taking into account all other conditions. This generated an undirected disease comorbidity network with significant pairs of conditions ($\beta > 0$). Directionality was determined by calculating relative risk (RR) in both directions, yielding 73,729 significant pairs of conditions ($RR > 1$). Trajectories were then assembled step-by-step using an iterative RR framework starting from HTN (I10*) outward to ICD neighbors while masking codes earlier in the trajectory at each step. Each individual was binned into a single trajectory starting from HTN out to the 5th neighbor condition. This uniquely eliminated insignificant disease pairs and identified the conditions that could function as "sign-posts" in a disease path. Our results yielded numerous positive-control trajectories with early symptoms that developed into concrete diagnoses and then chronic disease outcomes. For example, individuals on a "hepatic" path have HTN trajectories that go from "abnormal liver function lab" to "inflammatory liver disease", while individuals on a "renal" path go from "albuminuria" to "chronic kidney disease". Some novel connections between autoimmune and gynecological conditions (stratified analyses), as well as neuro-immunological connections, were identified. Our results demonstrate the potential for stronger characterization of the different disease trajectories and outcomes for patients with HTN. Further, the goal of this work is to train a model to predict comorbid conditions and disease trajectories that may present in the future for individuals with HTN, which may lead to the design of improved precision medicine strategies.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3358*. Deprivation index as a predictor in PheWAS recapitulates associations between socioeconomic status and phenotypes captured by ICD9 codes.

Authors:

M. March¹, F. Mentch², H. Qiu³, P. Sleiman⁴, H. Hakonarson²; ¹Children's Hosp. of Philadelphia, Philadelphia, PA, ²Children's Hosp. of Philadelphia, Philadelphia, PA, ³Children's Hosp. of Philadelphia, Philadelphia, PA, ⁴Rhythm Pharmaceuticals, Boston, MA

Abstract Body:

PheWAS analyses are typically used to examine the impact of a single genetic marker (usually single nucleotide polymorphisms) on an entire catalog of phenotypes derived algorithmically from diagnosis codes. However, other properly encoded predictive variables can be used in place of genotypes, and several examples of PheWAS performed with quantitative traits as the inputs have been published. A method for combining socioeconomic variables from the American Community Survey into a deprivation index based on census tracts was applied to the cohort of patients collected by the Center for Applied Genomics (CAG) at The Children's Hospital of Philadelphia. The resulting deprivation index was run through a PheWAS analysis containing the phecode phenotypes based on ICD9 codes as well as additional quantitative traits derived from lab test results. Strong associations were found for conditions expected to positively correlate with the deprivation index, including acute respiratory illness ($p = 3.2E-31$, OR = 4.2 in African ancestry), atopic dermatitis ($p = 7.2E-18$, OR = 3.2 in African ancestry), and obesity ($p = 8.3E-30$, OR = 19.4 in European ancestry, $p = 5.4E-17$, OR = 3.4 in African ancestry). Marked differences exist between cohorts of European and African ancestry: 141 phenotypes are significantly associated after Bonferroni correction in the African ancestry cohort, 33 in the European ancestry cohort; acute respiratory illness is significant in African Americans, while acute sinusitis is significant in both populations; otitis media is significantly associated in European but not African ancestry. Additionally, associations with conditions that would be predicted to be inversely associated with the deprivation index, like allergic reactions to food, were observed with the expected direction of effect ($p = 9.5E-08$, OR = 0.27). The deprivation index recapitulated associations between socioeconomic status and phenotypes previously described in the literature, while providing a wealth of information for potential data mining. Efforts are underway to understand the contributions of the components of the deprivation index to various phenotypes and to consider the confounding variables of geography and distance.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3359. Detecting haplotype association using synthetic genetic variants

Authors:

Y. Guan; Duke University, Durham, NC

Abstract Body:

Consider any k consecutive genotypes to define a synthetic genetic variant as follows: substituting each genotype with a matrix, with M_0 for homozygous reference allele, M_2 for homozygous variant allele, and $M_1 = 1/2(M_0 + M_2)$ for heterozygous; multiplying the k matrices to obtain a product matrix $M = P_1 \dots P_k$ where P_j equals M_0 or M_1 or M_2 depending on genotypes; computing $\log \text{tr}(M)$ (log of trace of M) and we obtain the dosage of the synthetic genetic variant. We may test for association between the synthetic genetic variant and a phenotype of interest. (The choice of M_0 and M_2 is a matter of design and the synthetic variants works for the simplest choice of M_0 and M_2 .)

The synthetic genetic variants have the following features: 1) they incorporate haplotype information without the need of phasing. To see this, for two adjacent heterozygous, the product matrix $M = 1/4(M_0M_0 + M_0M_2 + M_2M_0 + M_2M_2)$ averages over all four possible phase between two heterozygous SNPs. 2) Missing genotypes can be handled cleanly. Imputation outputs probabilities of the missing genotypes being homozygous reference allele p_0 , or homozygous variant allele p_2 , or heterozygous p_1 . We use $(p_0 M_0 + p_1 M_1 + p_2 M_2)$ in place of either M_0 or M_1 or M_2 to compute matrix multiplication. 3) We can tune k to let synthetic genetic variants reflect different scales of genetic variations. 4) Synthetic genetic variants can serve as surrogates for genetic background of an allele, which allows us to study genetic interactions between genotypes and their genetic backgrounds.

We demonstrate via simulation that synthetic genetic variants have better power to detect association than the single SNP test when there exists allelic heterogeneity. Using a legacy genome wide association dataset we demonstrate the usefulness of the synthetic genetic variants by uncovering association between *CDKN2A* and type 2 diabetes in European samples, a confirmed genetic association in East Asian samples that cannot be detected in European samples via single SNP test.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3360*. Detecting shared and sex-specific causal genetic effects on complex traits

Authors:

W. Zhang, H. Najafabadi, Y. Li; McGill Univ., Montreal, QC, Canada

Abstract Body:

Sex differences in health and diseases are prevalent. Understanding sex-specific biological pathways is important to research and health care. However, identifying sex-specific genetic effects is challenging, because a sex-stratified genome-wide association study (GWAS) or a variant-by-sex interaction test may be under-powered.

We propose SharePro to detect shared and sex-specific genetic effects. With sex-stratified GWAS summary statistics and matched linkage disequilibrium (LD) information, SharePro introduces a sparse projection to group genetic variants in high LD into effects and assess their roles in females and males. Posterior probabilities of genetic effects being shared or sex-specific can be accurately derived by efficient variational inference.

We conducted simulations to assess the accuracy and computational efficiency of SharePro. As expected, when all genetic effects were sex-specific, SharePro achieved comparable performance in identifying true causal effects as fine-mapping sex-stratified GWAS summary statistics separately. When all genetic effects were shared, SharePro was comparable to fine-mapping sex-combined GWAS summary statistics. Importantly, when the genetic effects were partially sex-specific, SharePro outperformed existing approaches. For instance, when 50% of the genetic effects were sex-specific, SharePro achieved an area under the precision-recall curve (AUPRC) of 0.73 while eCAVIAR and fastPAINTOR had an AUPRC of 0.49 and 0.58, whose computation time was 112 times and 215 times longer than SharePro, respectively. Next, we applied SharePro to investigate 20 traits and diseases in large cohort studies. We found that hormonal traits had widespread sex-specific genetic effects, with only 25% of causal genetic effects shared between females and males for testosterone level. Neurological diseases and behavioral traits exhibited moderate sex-specific genetic effects. For example, 56% of causal genetic effects on schizophrenia were shared. For biochemistry traits, genetic effects with a large effect size, which may correspond to core biological pathways, were mostly shared, although sex-specific genetic effects with a weak-to-moderate effect size also existed. For example, 78% of causal genetic effects on lipids were shared and enriched in cholesterol metabolism pathways.

In summary, we have developed SharePro for characterizing shared and sex-specific genetic effects. The accuracy and computational efficiency of SharePro supports its wide utility in profiling sex-dimorphic genetic architecture and identifying potential targets for individualized treatment and intervention.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3361. Detecting Somatic Mosaicism at Tandem Repeats

Authors:

A. Sehgal¹, M. Gymrek²; ¹Univ. of California, San Diego, La Jolla, San Diego, CA, ²Univ California San Diego, La Jolla, CA

Abstract Body:

Somatic mosaicism, in which a mutation occurs post-zygotically, has been implicated in several developmental disorders, diseases and cancers. Short tandem repeats (STRs) consist of repeated sequences of 1-6bp and comprise more than 1 million loci in the human genome. Somatic mosaicism at STRs is known to play a key role in the pathogenicity of loci implicated in repeat expansion disorders, and is highly prevalent in cancers exhibiting microsatellite instability. While a variety of tools have been developed to genotype germline variation at STRs, no method currently exists for systematically identifying mosaic STRs (mSTRs). Here, we introduce mosaicSTR, a novel method for detecting mSTRs from individual next-generation sequencing datasets. Unlike many existing mosaicism detection methods, mosaicSTR does not require a matched control sample as input. We show that mosaicSTR accurately identifies mSTRs in simulated data and demonstrate its feasibility by identifying candidate mosaic STRs in whole genome sequencing (WGS) data for individuals sequenced by the 1000 Genomes Project. MosaicSTR takes as input a vector of observed repeat copy numbers $R = \{r_1, r_2, \dots, r_n\}$, where r_i is the number of copies of the repeat observed in the i th read. For each STR locus, $\langle A, B \rangle$ denotes the diploid germline genotype, f denotes the fraction of chromosome copies harbouring an additional allele C resulting from a mosaic mutation, and Θ represents additional error parameters. mosaicSTR models the observed repeat copy numbers (R) as a mixture, in which $1-f$ copies come from germline alleles A and B , respectively, and f copies come from the mosaic allele C . It then applies an expectation-maximization approach to iteratively infer maximum likelihood values for C and f . Finally, it applies a likelihood ratio test to test the null hypothesis that $f=0$. We developed a framework to simulate observed repeat copy numbers (R) for given values of A , B , C , f , and STR error parameters Θ and for different sequencing coverage parameters. We found that mosaicSTR can accurately infer mosaic fractions (f) down to 2% at $>50x$ coverage. We additionally applied mosaicSTR to WGS data from individuals sequenced by the 1000 Genomes Project and identified candidate STRs with mosaic alleles.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3362. Detecting variable-number tandem repeat (VNTR) variation using Oxford Nanopore sequencing data at population scale.

Authors:

D. Beyter, S. Kristmundsdóttir, M. Hardarson, H. Hauswedell, H. Jónsson, H. Eggertsson, G. Sveinbjornsson, A. Oddsson, G. Holley, O. Stefansson, S. Stacey, S. Gudjonsson, P. Olason, T. Rafnar, G. Masson, U. Thorsteinsdottir, P. Sulem, A. Helgason, D. Gudbjartsson, B. Halldórsson, K. Stefansson; deCODE Genetics, Reykjavik, Iceland

Abstract Body:

Variable number tandem repeats (VNTRs) are associated with a number of diseases and other traits, and are a major source of genetic variation, comprising a large fraction of structural variants. Long-read sequences present an opportunity to greatly improve the discovery and genotyping of VNTRs. While the error rates of long-reads are continually decreasing, current error rates make VNTR allele discovery challenging, especially in low-coverage samples. We analyzed 6,503 individuals long-read sequenced to a coverage of 10-20x. Our large sample size, combined with long-range phasing of the individuals allows us to accurately determine VNTRs, both in motif composition and allele length. We aggregate phased long-reads from a single haplotype shared by multiple individuals to increase read coverage in order to accurately reconstruct VNTR alleles via multiple sequence alignment. This method allowed us to accurately genotype VNTR alleles in 77,586 regions and impute into 173,025 individuals.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3363. *Detection of underdiagnosis of complex diseases due to underrecognition of high polygenic risk*

Authors:

C. Marquez-Luna¹, I. Forrest¹, A. Duffy¹, H. T. Vy¹, R. Do²; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²Icahn Sch. of Med., New York, NY

Abstract Body:

Underdiagnosis is defined as a failure to diagnose a disease correctly. Studies have reported evidence of underdiagnosis of complex diseases in healthcare systems. However, it's difficult to accurately quantify the extent of underdiagnosis in a disease. As a result, the range in which complex diseases are underdiagnosed is still unknown. New statistical approaches are needed to quantify the underdiagnosis rate of common diseases.

A polygenic risk score (PRS) of disease potentially has clinical utility for risk stratification. Its distribution can reflect the true prevalence of inherited polygenic risk to disease. We developed a statistical test that leverages the PRS distribution of diagnosed cases and controls and quantifies the underdiagnosis rate of disease in individuals with high PRS. This test evaluates whether there is an excess proportion of undiagnosed individuals with high PRS compared to a reference PRS distribution under a 0% misclassification rate.

First, we conducted extensive simulations to evaluate the statistical properties of the PRS underdiagnosis test and show that the method has sufficient statistical power under a wide range of disease prevalence and sample size parameters. Next, we estimate the underdiagnosis rate for six complex diseases: atrial fibrillation, depression, coronary artery disease, breast cancer, schizophrenia, type 2 diabetes, and breast cancer. We analyzed on avg 311K samples of British ancestry in UK Biobank and built three PRS sets for each disease: healthy (reference), diagnosed, and undiagnosed. For the PRS underdiagnosis test, we calculated disease-specific unadjusted polygenic risk scores and estimated the excess proportion of undiagnosed individuals at PRS cutoffs of 2.5-, 3- and 4-fold increased inherited genetic risk for each disease. We observed marked excess of undiagnosed individuals for four out of six diseases across all three high PRS cutoffs. For the PRS cutoff of 4-fold risk, we found that across the six diseases, we observed an underdiagnosis rate of 26.2%, 22.3%, 10%, 7.3%, 7.1%, and 1.9% for each disease listed, respectively. We show underdiagnosis at 2.5- and 3-fold increased PRS risk at lower rates. We show the robustness of our findings by performing multiple sensitivity analyses. Furthermore, we validated our underdiagnosis inference using electronic health records and observed higher rates of symptoms and medications in undiagnosed individuals with specific diseases.

In conclusion, we present evidence of marked underdiagnosis of genetically at-risk individuals due to underrecognition of high PRS in a large healthcare system.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3364. Developing an African American population-based transcriptome prediction model from the GENE-FORCAST cohort

Authors:

G. Goodney¹, A. Gaye¹, G. Gibbons²; ¹Natl. Human Genome Res. Inst., Bethesda, MD, ²Natl. Heart, Lung, and Blood Inst., Bethesda, MD

Abstract Body:

African Americans have historically been and remain an underrepresented population in biomedical research. This disparity has translated to genomic research and disease prediction models. By creating reference and prediction models that are trained on individuals of the same ancestry, we will have stronger prediction performance and have insight into gene expression and genotype interactions that may not translate across populations. Using genotype and gene expression data from 505 patients from the Genomics, Environmental Factors and Social Determinants of Cardiovascular Disease in African-Americans Study (GENE-FORECAST) cohort, we have trained a cross-validated elastic net prediction model to be used with the Predixcan suite of software developed by the Im lab. Since most reference and prediction models are currently based on individuals with European ancestry, our goal is to improve prediction performance with a trained model that has similar ancestry to the population being tested. Incorporating diverse populations in reference and prediction models is necessary in genomic research going forward. Initial results show that ~4,558 genes have robust prediction performance after stringent filtering ($R > 0.1$ and $p\text{-val} < 0.05$). Further mixing of PEER and elastic net parameters will improve the strength of this model.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3365. Development and validation of an RNA-seq-based transcriptomic risk score for asthma

Authors:

T. Mersha¹, X. Cao², L. Ding³; ¹Cincinnati Children S Hosp. Med. Ctr., Cincinnati, OH, ²Univ. of Cincinnati, Cincinnati, OH, ³Cincinnati Children s Hosp. Med. Ctr., Cincinnati, OH

Abstract Body:

Most complex diseases are influenced by several loci, each with a small effect on its own, and polygenic approaches that group individual variants collectively influence a phenotypic trait offer a more predictive value than is possible by single variant approaches. The most popular polygenic approach is polygenic risk score (PRS). In this study, we are now adapting PRS approaches to transcriptomics data with the objective of developing and validating an RNA-seq-based transcriptomic risk score (RSRS) for disease risk prediction that can simultaneously accommodate demographic information. We analyzed RNA-seq gene expression data from 441 asthmatic and 254 non-asthmatic samples. Logistic least absolute shrinkage and selection operator (Lasso) regression analysis in the training set identified 73 differentially expressed genes (DEG) to form a weighted RSRS that discriminated asthmatics from healthy subjects with area under the curve (AUC) of 0.80 in the testing set after adjustment for age and gender. The 73-gene RSRS was validated in three independent RNA-seq datasets and achieved AUCs of 0.70, 0.77 and 0.60, respectively. Enrichment pathway analysis found that these genes were significantly ($P < 0.0001$) enriched for DNA replication, recombination, and repair, cell-to-cell signaling and interaction, and eumelanin biosynthesis and developmental disorder. Further in-silico analyses of the 73 genes using connectivity map shows that drugs (mepacrine, dactolisib) and genetic perturbagens (PAK1, GSR, RBM15 and TNFRSF12A) were identified and could potentially be repurposed for treating asthma. These findings show the promise for RNA-seq risk scores to stratify and predict disease risk.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3366. Differential gene expression analysis based on linear mixed model corrects false positive inflation for studying quantitative trait

Authors:

S. Tang¹, A. Buchman², D. Bennett², Y. Wang³, Q. Zheng⁴, J. Yang⁵; ¹Emory Univ., Dept. of Biostatistics and Bioinformatics, Atlanta, GA, ²Rush Alzheimer's Disease Ctr., Rush Univ. Med. Cente, Chicago, IL, ³Rush Univ. Med. Ctr., Chicago, IL, ⁴Univ. of Louisville, Dept. of Bioinformatics and Biostatistics, Louisville, KY, ⁵Emory Univ., Ctr. for Computational and Quantitative Genetics, Dept. of Human Genetics, Atlanta, GA

Abstract Body:

Differential gene expression (DGE) analysis has been widely conducted in biomedical research to identify genes expressed differentially with respect to a trait of interest using RNA sequencing (RNA-Seq) data. Because of expensive RNA-seq cost and difficult accessibility of relevant tissues, most previous studies are limited to dozens of samples and dichotomous traits under general laboratory settings. Most existing DGE methods are developed specifically for handling small sample sizes and dichotomous traits. However, an increasing number of population-based RNA-Seq data have been profiled for cohorts with hundreds of sample sizes and for studying both dichotomous and quantitative traits. Therefore, a DGE method suitable for studying quantitative traits while controls for false positive rate is of pressing need. We propose to use the linear mixed model (LMM) implemented by the Genome-wide Efficient Mixed Model Association (GEMMA) tool to conduct DGE analysis for quantitative traits. We show that the LMM controls false positive rate through our application studies of the RNA-Seq data of dorsolateral prefrontal cortex (DL-PFC) samples (n=632) for studying quantitative Alzheimer's Disease (AD) and Alzheimer's Disease-Related Dementias (ADRD) traits. Q-Q plots showed well control of false positive rates for DGE results by LMM (with genomic control factors ~1) versus the clearly inflated results by standard linear regression model without mixed effects (with genomic control factors >3). LMM identified 39 potentially significant genes with differential expression for AD/ADRD, which had p-values < 0.001 for at least one AD/ADRD trait. Among these 39 potential differentially expressed genes, we identified 2 significant for cognition decline rate, 5 significant for β -amyloid associated genes, 4 significant for tangle density, and 2 significant for global AD pathology burden, which had p-values less than the Bonferroni corrected significance threshold (3.49×10^{-6}). We further conducted DGE analysis in additional RNA-Seq validation data of DL-PFC (n=238), muscle (n=273), and spinal cord (n=236), for both AD/ADRD traits and motor metrics, which validated 13 of these 39 DGEs for AD/ADRD traits and 21 for motor metrics. In conclusion, LMM-based DGE corrects for false positive inflation which is a common issue of the standard linear regression model for studying population-based RNA-Seq data. Our real application results not only showed the effectiveness of the LMM method, but also suggested that motor functions might share gene regulatory mechanisms with AD/ADRD. A pipeline for conducting DGE for quantitative traits by GEMMA will be available on Github.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3367. Differentially Private Release of Genomic Summary Statistics Using Factorization Mechanisms

Authors:

S. Neel¹, L. O'Connor², H. Cho²; ¹Harvard Univ., Boston, MA, ²Broad Inst., Cambridge, MA

Abstract Body:

Summary statistics and allele frequencies reported by genome-wide association studies (GWAS), even large ones, can compromise the privacy of individuals in the dataset (Homer et al., 2008). Over the last decade, differential privacy (Dwork et al., 2006) has become the standard method for preserving privacy in the release of aggregate statistics, by adding a controlled amount of noise to the released data. However, the high-dimensional nature of genomic summary results has thus far prevented a practical application of differential privacy in genomics. Prior work on differentially private GWAS requires an overwhelming level of noise, or focuses on releasing only a small number of top SNPs. In this work, we address the problem of privately releasing full GWAS summary statistics with practical accuracy guarantees, preventing an adversary from inferring the phenotype of a study participant based on their genotype. We investigate two algorithmic approaches to cope with the high dimensionality of the data. First, we consider a matrix factorization-based approach for exploiting the latent structure of the genotype data to minimize the noise required for privacy. While small problem instances can be optimally solved via a semi-definite program, we adopt a more efficient gradient-based optimization for large-scale datasets. Second, we consider a model-based approach based on inferred tree sequences (Kelleher et al., 2019), whereby genealogical patterns in the data inform the construction of more effective differentially private mechanisms. We empirically evaluate the improvement of our approaches upon standard differential privacy mechanisms. Finally, we outline the remaining challenges in developing a fully-differentially private GWAS pipeline.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3368*. Differentiate association and causal effects of cardiovascular disease risk factors on Alzheimer's disease and its age-at-onset.

Authors:

J. La¹, X. Fan^{1,2}, X. Gao^{3,4,5}, E. Martin^{6,7}, Y-J. Li^{1,2}; ¹Duke Molecular Physiology Inst., Duke Univ., Durham, NC, ²Dept. of Biostatistics and Bioinformatics, Duke Univ., Durham, NC, ³Dept. of Ophthalmology and Visual Sci., The Ohio State Univ., Columbus, OH, ⁴Dept. of BioMed. Informatics, The Ohio State Univ., Columbus, OH, ⁵Div. of Human Genetics, The Ohio State Univ., Columbus, OH, ⁶John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miller Sch. of Med., Miami, FL, ⁷John T. MacDonald Fdn. Dept. of Human Genetics, Univ. of Miami, Miami, FL

Abstract Body:

Alzheimer's disease (AD) is one of the most common neuro-degenerative diseases with over 6.2 million people diagnosed in the US. Along with age and genetic risk factors (e.g., *APOE*), growing evidence from epidemiological studies linked various cardiovascular disease (CVD) risk factors to AD. However, whether these associations are causal to AD remain unclear. This study aims to delineate association and causal relationships between CVD risk factors and AD via polygenic risk score (PRS) and Mendelian Randomization (MR). For AD outcomes, we utilized the pooled genome-wide imputed genotype data of 9,219 AD cases and 10,345 controls from 20 cohorts of Alzheimer Disease Genetics Consortium. For exposure factors, genome-wide association study (GWAS) summary statistics of various CVD risk factors, including lifestyle factors were collected from the IEU OpenGWAS project. For each exposure factor, we compiled multiple sets of independent SNPs at different p-value thresholds (i.e., $P < 5 \times 10^{-8}$, 10^{-6} , 10^{-5}) and filtered by LD (clumping $r^2 < 0.001$) to construct optimal CVD-PRSs in the AD dataset with the best model fit. Generalized linear mixed models (GLMM) and linear mixed models (LMM) were used to test for CVD-PRS association with AD and age-at-onset (AAO) of AD, respectively, adjusting for sex, 10 principal components, and a random intercept by cohort. Causal effects were evaluated by the two-sample MR approach, where GWAS summary statistics for AD phenotypes were generated from the same GLMM and LMM models above. Markers meeting $p < 5 \times 10^{-8}$ from the exposure GWAS and not associated with AD outcomes ($p > 0.01$) were selected as instrumental variables (IV). We tested the validity and pleiotropy of IV by a heterogeneity test, and evaluated causal effects by the MR inverse variance weighted method, and then followed by sensitivity analysis using weighted median, MR-Egger, and MR-PRESSO. Our PRS analysis identified four shared CVD risk factors (coronary heart disease, hypertension, LDL cholesterol and total lipids in small LDL) associated with increasing AD risk and decreasing AAO of AD (e.g., LDL: OR (95%CI)=1.08 (1.05, 1.11), $P=1.78 \times 10^{-6}$; beta (SE)=-0.36 (0.07), $P=6.25 \times 10^{-7}$), and three additional ones showed nominal significant association with AAO of AD only. Among them, causal effects were presented in type-2 diabetes, hypertension, LDL cholesterol and total lipids in small LDL on AAO of AD, but only hypertension on AD risk. Further analysis of the expanded list of CVD biomarkers is ongoing. The findings of several CVD risk factors with causal effect on AAO of AD are novel and valuable to future development of lifestyle intervention to delay the AAO of AD and reduce AD risk.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3369. Discovery of clinically relevant gene-phenotype associations from 454,712 whole exomes in the UK Biobank

Authors:

P. Fizev¹, J. McRae², A. Metwally², T. Hamp³, Y. Yang¹, Z. Ni⁴, J. Schraiber¹, F. Aguet¹, H. Gao⁵, Y. FIELD², K. Farh⁶; ¹Illumina, Inc, Foster city, CA, ²Illumina, Inc, Foster City, CA, ³Illumina, Inc, Cambridge, United Kingdom, ⁴Dept. of Statistics, UW Madison, Madison, WI, ⁵Illumina, inc., Foster City, CA, ⁶Illumina, Foster city, CA

Abstract Body:

To what extent is a person's genetic susceptibility to common diseases, such as diabetes and cardiovascular disease, explained by a few rare genetic mutations with severe effects compared to a large collection of common variants with modest effect sizes? We employ PrimateAI-3D, a novel machine learning method for variant pathogenicity prediction with best-in-class performance, in order to identify rare deleterious variants that cause penetrant, early-onset disease across 454,712 exomes from the UK Biobank. Compared to other methods, our pipeline finds 20% more gene-phenotype associations with lower false discovery rate. Furthermore, PrimateAI-3D scores of rare missense variants in relevant genes are predictive of quantitative biomarkers and age of onset of dyslipidemia more than other pathogenicity scoring methods. We find that the majority of the population carries one or more rare penetrant variants, and that for healthy members of the general population, personal genome sequencing can reveal actionable genetic findings that strongly influence their resistance or susceptibility to disease.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3370. Discovery of rare variants associated with alcohol problems improved by leveraging machine learning phenotype prediction and empirical functional variant weighting.

Authors:

M. Ahangari¹, A. Gentry¹, R. Kirkpatrick², T-H. Nguyen¹, K. Kendler², S-A. Bacanu², R. Peterson², B. Riley², B. Webb³; ¹Virginia Commonwealth Univ., Richmond, VA, ²Virginia Commonwealth Univ, Richmond, VA, ³RTI Intl., Research Triangle Park, NC

Abstract Body:

Alcohol use disorder (AUD) is moderately heritable with significant social and economic impact. Genome-wide association studies (GWAS) have identified common variants associated with AUD. However, rare variant investigations have yet to achieve well-powered samples sizes. One strategy to increase power is to predict risk in individuals not directly measured using machine learning methods. Power may also be increased if the impact of variation on functional elements can be empirically estimated for the outcome of interest. In this study, we conducted an interval-based exome-wide analysis of Alcohol Use Disorder Identification Test Problems subscale (AUDIT-P) using both predicted risk and empirical functional weights in the UK Biobank (UKB). This research has been conducted using the UKB (application 30782). AUDIT-P was directly measured in 157,162 UKB participants and predicted using MAGIC-LASSO in the remaining 345,374 participants. Limiting the 200k exome release to unrelated individuals of European ancestry resulted in a sample of 147,386 individuals with 51,357 observed and 96,029 unmeasured but predicted AUDIT-P for exome analysis. Sequence Kernel Association Test (SKAT) was used for rare variant (MAF < 0.01) interval analyses using default and empirical functional weights. Empirical functional weights were constructed using annotations found significant by stratified LD Score Regression analysis of predicted AUDIT-P GWAS. Thus, providing prior functional weights specific to AUDIT-P. Based on the correlation between measured and predicted AUDIT-P, the effective sample size was increased by 56%. There was no evidence of inflation when using default vs empirical weights. ADH1C and THRA gene intervals were significant (FDR $q < 0.05$) using default and empirical weights. The most significant association was found using predicted AUDIT-P and empirical weights in the ADH1C gene (SKAT-O $P_{\text{Default}} = 1.06 \times 10^{-9}$ and $P_{\text{Empirical weight}} = 6.25 \times 10^{-11}$). This represents significant improvement over associations using directly measured AUDIT-P and default SKAT weights with the majority of performance increase coming from the increase in effective sample size. These findings highlight the successful leveraging of machine learning to increase effective sample size for underpowered phenotypes, and prior empirical functional weights based on common variant GWAS data to refine and increase the statistical significance in interval-based tests compared to default SKAT weights. Future directions include improving the performance of empirical functional weights in interval-based tests of rare variants.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3371. Disentangling the common genetic architecture and causality of rheumatoid arthritis and systemic lupus erythematosus with COVID-19 outcomes: genome-wide cross trait analysis and bi-directional Mendelian randomization study

Authors:

J. Zhao¹, M. Yao¹, X. Huang¹, Z. Liu²; ¹The Univ. of Hong Kong, Hong Kong, Hong Kong, ²Univ. of Hong Kong, Hong Kong, China

Abstract Body:

COVID-19 may cause a dysregulation of the immune system and has complex relationships with multiple autoimmune diseases, including rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). However, little is known about their common genetic architecture. We analysed summary-level genetic data from the latest COVID-19 host genetics consortium and international consortia to examine the shared genetic etiology and causal relationship between COVID-19 and RA/SLE. The cross-trait meta-analysis identified 46, 47, and 19 shared genetic loci for severe COVID-19, COVID-19 hospitalization, and SARS-CoV-2 infection with RA, and 19, 24, and 11 shared loci with SLE, respectively. The shared genes were enriched in the spleen, lung, whole blood, and small intestine, and are involved in immune function, inflammation and coagulation. Co-localization analysis identified eight shared loci in *TYK2*, *IKZF3*, *COL11A2*, *PSORS1C1*, *MANEAL* and *COG6* genes for COVID-19 with RA, and four in *CRHR1*, *FUT2* and *NXPE3* genes for COVID-19 with SLE. Bi-directional Mendelian randomization analysis suggested RA is associated with higher risk of COVID-19 hospitalization. Our novel findings can help better understand the common biological processes of COVID-19 and autoimmune diseases.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3372. Disentangling the genetic overlap and causal relationship between primary open-angle glaucoma, brain morphology, and major neurodegenerative disorders.

Authors:

P. Gharahkhani¹, S. Diaz-Torres^{1,2}, W. HE¹, J. ONG¹, S. Macgregor¹; ¹QIMR Berghofer Med. Res. Inst., Brisbane, Australia, ²Faculty of Med., Univ. of Queensland, Brisbane, Australia

Abstract Body:

Primary open angle glaucoma (POAG) is an optic neuropathy characterized by progressive degeneration of the optic nerve that leads to irreversible visual impairment. POAG endophenotypes such as intraocular pressure (IOP) and vertical cup-to-disc ratio (VCDR) of the optic disc are the major risk factors for POAG. Multiple epidemiological studies suggest an association between POAG or its endophenotypes and major neurodegenerative disorders (i.e. Alzheimer's disease, amyotrophic lateral sclerosis, frontotemporal dementia, and Parkinson's disease). However, the nature of the overlap between these neurodegenerative disorders, brain morphology and glaucoma remains inconclusive.

In this study, we performed a comprehensive assessment of the genetic and causal relationship between POAG, brain morphology, and the major neurodegenerative disorders specified above. Leveraging genome-wide association data from magnetic resonance imaging of more than 600 brain regions, POAG and its endophenotypes, and the four major neurodegenerative disorders, this study found a causal relationship between POAG or its endophenotypes and brain morphology in 22 regions using a Mendelian randomization (MR) framework. We also identified genetic overlap between POAG or its endophenotypes and neurodegenerative disorders using genomic structural equation modeling. However, we did not find strong evidence of causal association between the major neurodegenerative disorders and POAG or its endophenotypes using MR and genetic colocalization analyses. Similarly, the brain regions that were causally associated with the neurodegenerative disorders did not overlap with those causally associated with POAG or its endophenotypes.

To the best of our knowledge, this is the first study that comprehensively examined the genetic overlap and causal relationship between POAG, brain morphology and neurodegenerative diseases. Our findings indicate distinctive and likely independent neurodegenerative processes for POAG although several regions of the genome contribute to neurodegeneration of both the optic nerve and neurodegenerative disorders.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3373*. Distinct explanations underlie gene-environment interactions in the UK Biobank

Authors:

A. Durvasula¹, A. L. Price²; ¹Harvard Med. Sch., Boston, MA, ²Harvard Sch Pub Hlth, Boston, MA

Abstract Body:

Gene-environment interactions have been widely hypothesized to impact the genetic architecture of complex diseases and traits (Li et al. 2019 Cell). While interactions have been identified at individual loci, little is known about the overall contribution of gene-environment interactions to complex trait architectures. Here, we leverage the large size of the UK Biobank to quantify the genome-wide effects of gene-environment interactions across 49 complex traits and 8 environmental variables. We specifically assess the interaction between polygenic scores (PS) (trained in N=337K British samples; Weissbrod et al. 2022 Nat Genet) and each environmental variable (E) in N=47K non-British European validation samples, in a model that includes PS, E, and PSxE. We identified 36 significant interactions (FDR < 5%) spanning 17 traits and 7 environmental variables, although the interactions explained an average of <0.01% of trait variance (implying a <1% contribution to trait variance after summing across E variables and correcting for the partial accuracy of PS). Below, we highlight 3 examples with distinct underlying explanations. First, we identified a significant interaction between BMI PS and physical activity ($p=5E-5$); we determined that both the variance of the PS and the SNP-heritability vary as a function of physical activity, implying that the interaction is driven by changes in genetic variance. Second, we identified a significant interaction between Type 2 Diabetes PS and time spent napping ($p=5E-12$); we determined that the SNP-heritability varies as a function of time spent napping, but the variance of the PS does not, implying that the interaction is driven by changes in total environmental variance rather than genetic variance. Third, we identified a significant interaction between All Autoimmune Disease PS and alcohol consumption ($p=4E-4$); we determined that the variance of the PS varies as a function of alcohol consumption, but the SNP-heritability does not, implying proportionate effects on both genetic and total environmental variance. We identified additional interactions corresponding to each of these explanations. Overall, our results provide insights into subtle interactions between genetic and environmental variables, as well as the distinct explanations underlying the interactions.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3374. Diversity is key for POAG: the proportion of variance for primary open-angle glaucoma explained by covariates and genetic risk scores varies by ancestral group in the Million Veteran Program

Authors:

A. R. Waksunski^{1,2}, T. G. Kinzy^{1,2,3}, L. A. Cruz^{1,2}, C. L. Nealon⁴, C. W. Halladay⁵, P. Simpson⁴, R. L. Canania⁴, S. A. Anthony⁴, D. P. Roncone⁴, L. Sawicki Rogers⁶, J. N. Leber⁶, J. M. Dougherty⁶, P. B. Greenberg^{7,8}, J. M. Sullivan^{6,9}, W-C. Wu¹⁰, S. K. Iyengar^{1,2,3}, D. C. Crawford^{1,2,3}, N. S. Peachey^{3,11,12}, J. N. Cooke Bailey^{1,2,3}, VA Million Veteran Program; ¹Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, ²Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, ³Res. Service, VA Northeast Ohio Hlth.care System, Cleveland, OH, ⁴Eye Clinic, VA Northeast Ohio Hlth.care System, Cleveland, OH, ⁵Ctr. of Innovation in Long Term Services and Supports, Providence VA Med. Ctr., Providence, RI, ⁶Ophthalmology Section, VA Western NY Hlth.care System, Buffalo, NY, ⁷Ophthalmology Section, Providence VA Med. Ctr., Providence, RI, ⁸Div. of Ophthalmology, Alpert Med. Sch., Brown Univ., Providence, RI, ⁹Res. Service, VA Western NY Hlth.care System, Buffalo, NY, ¹⁰Cardiology Section, Med. Service, Providence VA Med. Ctr., Providence, RI, ¹¹Cole Eye Inst., Cleveland Clinic Fndn., Cleveland, OH, ¹²Dept. of Ophthalmology, Cleveland Clinic Lerner Coll. of Med. of Case Western Reserve Univ., Cleveland, OH

Abstract Body:

Primary open-angle glaucoma (POAG) is an age-related eye disease that is the leading cause of irreversible blindness worldwide. Early POAG intervention is essential for mitigating severe outcomes. POAG is highly heritable, yet approximately 90% of the genetic component has yet to be discerned. The largest-to-date cross-ancestry POAG genome-wide association study identified 127 POAG-associated variants. Aggregating these variants via a genetic risk score (GRS) may be useful for risk stratification and informing recommendations for clinical POAG screening. We calculated 127-variant GRS for European-descent (EUR), African-descent (AFR), and Hispanic (HIS) Veterans in the Million Veteran Program. GRSs were unweighted or weighted by published cross-ancestry effect estimates. Models were either unadjusted or adjusted for age, sex, and 10 sample-specific principal components (PCs). To evaluate GRS performance, we constructed receiver operating characteristic curves and estimated the area under the curve (AUC). AUC estimates in EUR and HIS Veterans were higher than in AFR Veterans in unadjusted models; this trend was reversed in adjusted models. Consequently, we explored the estimated proportion of POAG variance explained by: (i) age and sex, (ii) age, sex, and 10 PCs, and (iii) age, sex, 10 PCs, and each GRS (unweighted and cross-ancestry meta-weighted) in our models. We calculated coefficients of determination (R^2) on the observed scale (Nagelkerke's) and the liability scale using a fixed disease prevalence of 2.4% as well as increases in R^2 with the addition of each variable to the model. We found that covariates alone (age, sex, and 10 PCs) explained a higher proportion of POAG variance in AFR Veterans (Nagelkerke's $R^2=0.103$; liability $R^2=0.063$) than in EUR and HIS Veterans (Nagelkerke's $R^2=0.002$ and 0.0343 ; liability $R^2=0.0023$ and 0.030 , respectively). Adding the unweighted GRS resulted in a larger increase in R^2 in EUR and HIS Veterans than in AFR Veterans (Nagelkerke's R^2 increase= 0.045 , 0.051 , and 0.024 ; liability R^2 increase= 0.052 , 0.046 , and 0.016 , respectively). This trend was consistent with the addition of weighted GRS (Nagelkerke's R^2 increase= 0.056 , 0.062 , and 0.03 ; liability R^2 increase= 0.065 , 0.056 , and 0.019 in EUR, HIS, and AFR Veterans, respectively). These observations suggest that, in our adjusted models, covariates alone were more informative for AFR Veterans while a GRS was more informative for EUR and HIS Veterans. Thus, these findings highlight the need for increased inclusion of diverse populations, specifically of African descent, in POAG genetics to inform the development of more equitable GRS-based tools.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3375. DNARecords: An extensible sparse format for petabyte scale genomics analysis

Authors:

A. Mañas Mañas¹, A. Dixit², L. Seninge³; ¹Coral Genomics, Almería, Spain, ²Coral Genomics, San Carlos, CA, ³Coral Genomics, San Francisco, CA

Abstract Body:

Recent growth in population scale sequencing initiatives involve both cohort scale and proportion of genome surveyed, with a transition from genotyping arrays to broader genome sequencing approaches. The resulting petabyte scale datasets can be challenging to analyze. Here we introduce DNARecords a novel sparse format for large scale genetic data. The structure enables integration of complex data types such as medical images and chemical structures towards the development of machine learning methods to predict disease risk and drug response. We demonstrate its speed and memory advantages for various genetics analyses including PCA and prioritizing drug targets. These performance advantages will become more pronounced as it becomes feasible to analyze variants of lower population allele frequencies. We provide an open source software plugin, built on top of Hail, to allow researchers to write and read such records as well as a set of examples for how to use them.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3376. DosaCNV: A deep multiple instance learning framework for jointly predicting copy number variation pathogenicity and gene dosage sensitivity.

Authors:

Z. Liu, Y. Huang; Pennsylvania State Univ., State College, PA

Abstract Body:

Since copy number variation (CNV) is a major contributor to human genetic diversity, being able to correctly predict CNV phenotypes would drastically improve our ability to understand the genetic basis of evolution and disease development. It has been shown that pathogenicity of large CNVs may be attributable to reduced or gained gene product of dosage sensitive genes. Adopting this idea, existing methods predict CNV pathogenicity mainly from two separate perspectives: (1) estimate gene-level dosage sensitive scores based on clinically curated CNVs, then use simple mathematical functions to aggregate those scores based on overlapped genes of novel CNV to make variant-level pathogenicity prediction, or (2) estimate CNV pathogenicity directly by using variant-level features such as variant length and how many dosage-sensitive genes are overlapped. However, these two perspectives should be modeled as one continuous biological process since CNV pathogenicity is a joint effect of gene dosage sensitivities. Here we present a deep multiple instance learning model, DosaCNV, which jointly models gene-level dosage sensitivity with CNV-level pathogenicity in a biologically coherent manner to improve predictions at both gene and variant level. In addition, DosaCNV uses over 40 gene-level features (genomic, epigenomic, evolutionary, functional) to capture gene dosage sensitivities. Compared to previous methods, DosaCNV shows unmatched performance in (a) predicting pathogenicity of clinically curated copy number deletions, (b) prioritizing pathogenic copy number deletions from case-control data of neurodevelopmental disorders, and (c) predicting ClinGen haploinsufficient genes. Taken together, DosaCNV has the potential to leverage large-scale CNV data to discover disease-specific essential genes and relate those genes to CNV pathogenicity, which makes DosaCNV an accurate and practical computational method for both basic and translational research.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3377. DrFARM: Identification and inference for master regulator variants in multi-trait GWAS

Authors:

L. Chan¹, G. Li¹, E. B. Fauman², M. Laakso³, M. Boehnke¹, P. X. K. Song¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Pfizer, Cambridge, MA, ³Univ. of Eastern Finland, Kuopio, Finland

Abstract Body:

To understand the mechanism of pleiotropy in the metabolism pathway, we develop a new statistical methodology to identify master variants that involve multi-channels of genetic regulations to simultaneously influence several different traits. Advances in technologies have significantly accelerated the availability of various multi-omics data types, an unprecedented opportunity arises in the discovery of master regulator variants, such as pleiotropic variants. Existing statistical methods for this type of analysis are based mostly on testing for marginal associations of single variants with one-dimensional traits, which may lose statistical power when traits are correlated. In addition, carrying over false positives from the single-variant analyses to the pleiotropic traits analyses is problematic, leading to the finding of spurious master regulator variants. Our proposed statistical approach, termed **Debiased-Regularized Factor Analysis Regression Model (DrFARM)** aims to overcome these technical shortcomings. Built upon a joint regression model, DrFARM allows both high-dimensional genetic variants and multilevel dependencies so that several types of correlations may be incorporated into the model specification, such as kinship for relatedness and population structure. Being a one-stage analysis strategy, this joint modeling approach gives rise to the capacity of controlling an overall error and carrying out adequate uncertainty quantification for valid inferences. Taking both strengths from the debiasing technique in the high-dimensional inference and the Cauchy combination test with correlated p-values, we establish a valid post-variable selection inference on the discovery of master regulator variants. Through extensive simulations, we show DrFARM enjoys an appropriate control of the false discovery rate and exhibits desirable statistical power. In an application of the **Metabolic Syndrome in Men (METSIM)** cohort data analysis, we identified 386 potential master regulator variants at $p < 7.2 \times 10^{-11}$. Among them, two variants, rs138640017 and rs11131799, are linked to putative causal genes, GLDC and ADA, respectively. Our findings suggest vertical pleiotropy of glycine and multiple glycine adducts at GLDC and anti-correlated horizontal pleiotropy of N-acetylglucosamine and aspartate at ADA, which is consistent with their known biology.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3378. Effect of type 2 diabetes and its genetic susceptibility on severity and mortality of COVID-19 in UK Biobank

Authors:

W. Chung^{1,2}, A. Lee², Y. Cho², J. Li¹, T. park³, L. Liang¹; ¹Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ²Soongsil Univ., Seoul, Korea, Republic of, ³Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract Body:

Although Type 2 diabetes (T2D) and its related-traits have been known as one of the important risk factors for the severity and mortality of COVID-19, the effect of T2D, T2D-related traits and their genetic susceptibility on COVID-19 are largely unknown. We analyzed the population-based cohort data of 459,188 individuals from UK Biobank along with COVID-19 test results and information on individuals' hospitalization and death-related records during the period from March 11, 2020 to December 20, 2021. First, we investigated the association of T2D, T2D-related traits and their genetic susceptibility of T2D with severity of COVID-19 using multivariable logistic regression and cumulative logit model adjusted for potential confounders. To capture overall genetic susceptibility for T2D, we computed polygenic risk scores (PRS) based on summary statistics from UK Biobank, representing an individual's overall genetic risk for T2D trait. In the multivariable logistic models, we found that the odds ratio (OR) of T2D for COVID-19 infection was 1.555 ($P=3.49 \times 10^{-86}$) and OR of PRS for T2D with one-unit (=standard deviation) increase in PRS was 1.064 ($P=3.11 \times 10^{-12}$) when adjusted for the potential confounders including age, gender, genotyping array and four genotype PCs, indicating the roles of T2D-related genetics in the pathogenesis of COVID-19 infection. Next, we performed multivariable Cox proportional hazard models to investigate the effect of T2D patients infected with COVID-19 on the survival times. The estimated survival curves and pairwise log-rank tests showed that the estimated hazard for COVID-19 infected T2D patients were 4.67 times ($P=9.88 \times 10^{-324}$) and 2.58 times ($P=6.20 \times 10^{-231}$) higher than individuals without COVID-19 infection and T2D, respectively and the hazard ratio (HR) of PRS for T2D with one-unit increase in PRS was 1.088 ($P=4.76 \times 10^{-14}$). Furthermore, we found the mortality of COVID-19 infected T2D patients was dramatically increased compared to T2D patients not infected with COVID-19 and the mortality of individuals with high genetic susceptibility for T2D was increased as well.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3379. Efficient multivariable Mendelian randomization for confounder adjustment using public GWAS databases

Authors:

J. Morrison; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Mendelian randomization (MR) is a form of instrumental variable (IV) analysis in which genetic variants are used as instruments to test for and estimate the causal effect of an exposure on an outcome. Recently, MR has become increasingly popular, in part because it can be performed using on summary statistics from genome-wide association studies (GWAS) which are readily and publicly available. However, genetic variants often display horizontal pleiotropy, a violation of standard MR assumptions in which variants affect the outcome through pathways not mediated by the exposure of interest. Horizontal pleiotropy mediated by heritable confounders is particularly problematic because it leads to systematic correlation in variant-exposure and variant-outcome associations that is not reflective of a causal effect. If heritable confounders are known or suspected, multivariable MR (MVMR) can be used to adjust for the confounding bias. As the number of GWAS with available summary statistics increases, this strategy becomes a promising option for obtaining accurate causal estimates. However, existing MVMR methods perform poorly for moderate or large numbers of traits, suffering either from bias that increases dramatically with the number of included traits or untenable computational demands. We propose an alternative MVMR method using a fast variational algorithm to optimize the joint likelihood, called Empirical Shrinkage Mendelian Randomization (ESMR). We show that ESMR has low bias regardless of the number of traits included and has computational time scaling linearly in the number of adjusted traits. In addition to performing well in the multivariable setting, ESMR is unbiased in the univariable case, even when the exposure GWAS is under-powered, a setting where alternative methods have large weak-instrument bias. Additionally, we propose a semi-automated algorithm for identifying heritable confounders in public GWAS databases, which can then be adjusted for using ESMR. This creates a pipeline that allows researchers to inform their MR analyses using existing wealth of GWAS results.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3380. Efficient tests of marginal epistasis reveal the impact of interactions on complex traits

Authors:

A. Pazokitoroudi, B. Fu, S. Sankararaman; Univ. of California, Los Angeles, Los Angeles, CA

Abstract Body:

The contribution of epistasis (gene-gene interaction) to human complex trait variation remains poorly understood. Methods that aim to explicitly identify pairs of SNPs associated with a trait suffer from low power due to the large number of hypotheses tested. An alternate approach involves testing whether a single SNP modulates variation in a trait against the background of all (or a subset of) SNPs across the genome. While overcoming the limitation of low power, tests of marginal epistasis are infeasible on Biobank-scale data where hundreds of thousands of individuals are genotyped over millions of SNPs. We present a scalable method to estimate the marginal epistatic effect. Our algorithm is a streaming randomized method-of-moments estimator that has a runtime sub-linear in the size of the genotype matrix while using a constant amount of memory thereby able to test for epistasis of a single SNP with a background of half a million SNPs across ~300K individuals in a few hours. Across extensive simulations, we find that our method accurately estimates additive and epistatic effects across a range of genetic architectures while also controlling false positive rates.

We applied our method to 16914 trait-loci pairs corresponding to 10846 loci associated with 57 traits ($p < 5e-8$) in GWAS to discover 346 trait-loci pairs (254 loci across 34 traits) with significant evidence for marginal epistasis ($p < 5e-8$). We validated 183/346 pairs as remaining significant by testing on imputed genotypes and as being robust to population stratification.

We then partitioned significant marginal epistatic effects with respect to the same vs distinct chromosomal regions to find 16/346 (on the same chromosome) and 76/346 (on the distinct chromosomes) trait-loci pairs with statistically significant epistasis effects. Notable examples of marginal epistasis include loci for mean platelet distribution width and blood monocytes phenotypes are located on *TUBB1* and *FLT3* genes respectively ($p < 1e-50$). Our results provide evidence for epistatic effects underlying complex traits that can now be interrogated in large sample sizes.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3381*. Elucidate trans-ancestry cross trait genetic architecture for tobacco and alcohol use phenotypes using 3 million Individuals.

Authors:

X. Wang, GWAS and Sequencing Consortium of Alcohol and Nicotine Use; Penn State Coll. of Med., Hershey, PA

Abstract Body:

Genome-wide association meta-analysis (GWAMA) is a useful approach to improve the power of detecting associations using summary statistics. Recently, GWAMA starts to include non-European samples to further improve power, pinpoint causal variants, and elucidate genetic architecture. Genetic effects may differ among ancestries, due to genetic and non-genetic influences. Modeling the trans-ancestry genetic effect distribution will improve the power for association analyses, quantify the extent of genetic heterogeneity, and elucidate cross-trait genetic architecture. Borrowing ideas from analysis of variance, we propose a mixed effect meta-regression model MEMO to partition the genetic effects into components that remain constant across ancestries, vary with ancestries, and vary independently of ancestries. We use principal components of allele frequencies from each study as a proxy for ancestry to capture genetic effects variation. The intercept of the model captures shared effect across ancestries. We further use a random effect to capture how genetic effects vary independently from ancestries. By imposing a Dirichlet-Multinomial prior on regression coefficients, MEMO borrows strength across variants, learns the genetic trans-ancestry effect distribution, and fine-maps causal variants. We apply our method to 21 cohorts of tobacco and alcohol use traits, including 23andMe, UK BioBank, Biobank Japan and China Kadoorie Biobank, with 3 million samples (79% European, 9.2% East Asian, 8.8% Admixed American, 3% African American), and discover 722 loci ($p < 5e-9$) across five traits, where 124 loci are novel compared to other methods. Our method fine-maps 23% of loci to less than 6 variants, significantly improving over ancestry-specific results. On average, 78% of loci show a homogenous effect. 14% of the loci are best supported by a model with distinct effects between European and Asian ancestry. In MEMO model, by using components that remain constant across ancestries, defined as partial residual, we observe 0.1 increase ($p = 3e-20$, t-test) in genetic correlation estimates. Moreover, within-trait genetic correlation increased from 0.61 to 0.78 ($p = 5e-15$) with across-traits from 0.36 to 0.45 ($p = 1e-14$). Additionally, genetic correlations rise for studies with different ancestry (0.38 vs 0.53, $p = 2e-28$), while preserve for the same ancestry studies (0.45, $p = 0.75$). Our results indicate partial residual can better capture genetic correlation, especially for studies with different ancestries. Together, our results represent a significant step forward in understanding the genetic architecture of tobacco and alcohol use in trans-ancestry samples.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3382. Estimating heritability explained by local ancestry and stratification bias in admixture mapping from summary statistics

Authors:

T. Chan¹, X. Rui¹, C. Haiman¹, D. Conti¹, L. Le Marchand², G. Wojcik³, M. Graff⁴, H. Highland⁴, K. North⁴, R. Tao⁵, J. Haessler⁶, C. Kooperberg⁶, C. Gignoux⁷, S. Jung⁸, M. Fornage⁹, E. Kenny¹⁰, PAGE Consortium, C. Chiang^{11,12}, N. Mancuso^{11,12}; ¹Ctr. for Genetic Epidemiology, Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ²Epidemiology Program, Univ. of Hawaii Cancer Ctr., Honolulu, HI, ³Dept. of Epidemiology, Bloomberg Sch. of Publ. Hlth., Johns Hopkins Univ., Baltimore, MD, ⁴Dept. of Epidemiology, Gillings Sch. of Publ. Hlth., Univ. of North Carolina, Chapel Hill, NC, ⁵Dept. of Biostatistics, Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁶Div. of Publ. Hlth.Sci., Fred Hutchinson Cancer Res. Ctr., Seattle, WA, ⁷Colorado Ctr. for Personalized Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, ⁸Translational Sci. Section, Sch. of Nursing, Jonsson Comprehensive Cancer Ctr., Univ. of California, Los Angeles, Los Angeles, CA, ⁹Brown Fndn. Inst. for Molecular Med., The Univ. of Texas Hlth.Sci. Ctr., Houston, TX, ¹⁰Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, ¹¹Ctr. for Genetic Epidemiology, Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Los Angeles, CA, ¹²Dept. of Quantitative and Computational Biology, Univ. of Southern California, Los Angeles, CA

Abstract Body:

Admixture mapping is a powerful alternative to genome-wide association studies. The heritability explained by local ancestry, h_A^2 , which serves as a proxy measure for the narrow-sense heritability h^2 , also provides crucial insight into the genetic architecture of a disease or trait. However, current approaches to estimate h_A^2 are computationally limiting for large-scale datasets, and are susceptible to biases due to stratification in ancestral populations. Although existing summary statistics-based methods developed for GWAS distinguish biases from heritability, they fail to fully capture the extensive correlations between markers. Here, we present a novel method, Heritability estimation from Admixture Mapping Summary Statistics (HAMSTA), which uses admixture mapping summary statistics to infer variance components due to local ancestry effects and to distinguish biases due to ancestral stratification. HAMSTA takes into account the long-range nature of the local ancestry correlations, which is crucial for accurate inference. Through extensive simulations, we first demonstrate that HAMSTA computes unbiased estimates of h_A^2 in presence of ancestral stratification, in contrast to estimates obtained from GREML and LDSC. Second, across a range of trait architectures, HAMSTA provides unbiased estimates of biases due to ancestral stratification in observed test statistics with greater precision than LDSC, resulting in $\sim 10\times$ more statistical power on average. Lastly, we find that using a significance threshold corrected by HAMSTA estimates achieves a family-wise error rate of $\sim 5\%$, unlike existing approaches for FWER estimation in admixture mapping. Having demonstrated that HAMSTA is unbiased and more powerful compared to individual-level and other summary-data based approaches, we applied HAMSTA to 26 phenotypes of 17,299 African American individuals in Population Architecture using Genomics and Epidemiology (PAGE) study. We observed significant h_A^2 in 15 phenotypes with an average $h_A^2 = 0.016$ (SD = 0.013), which translates to an average $h^2 = 0.49$ (SD = 0.39). Estimates of h_A^2 are consistent with previous work, where the derived h^2 of height and BMI are estimated to be 0.86 (SE = 0.083) and $h^2 = 0.34$ (SE = 0.095) respectively, and white blood cell count shows highest h_A^2 due to selection at causal variants. The estimated inflation factors have an average of 1.00 (SD = 0.02), suggesting little evidence of bias due to ancestral population stratification in current admixture mapping studies. Overall, HAMSTA provides a fast and powerful approach to estimate genome-wide heritability and biases using summary statistics from admixture mapping studies.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3383. Estimating indirect maternal genetic effects on children's autism spectrum disorder risk

Authors:

Y. Wu, Z. Sun, Q. Lu; Univ. of Wisconsin-Madison, Madison, WI

Abstract Body:

The etiology of autism spectrum disorder (ASD) involves substantial genetic and environmental components as well as their complex interplay. One particular type of gene-environment interplay that caught many geneticists' attention is "genetic nurture": parents' genotypes could influence parental behavior and family environment which in turn affects children's phenotypes. Due to the correlation between parents' and children's genotypes, genetic effect estimates in a regular genome-wide association study (GWAS) are mixtures of direct effects (one's genotypes affecting one's phenotypes) and indirect genetic effects (one's genotypes affecting their children's phenotypes). Careful dissection of direct and indirect genetic effects has revealed strong evidence of maternal indirect effects on birth weight and cognition and have fundamentally changed our understanding of the genetic basis of these traits. Here, we employed an unconventional study design to search for evidence of indirect maternal effects on ASD risk, leveraging a large number of trios of ASD probands and healthy parents from SPARK and Simons Simplex Collection cohorts. We performed a GWAS where the mothers and fathers of ASD probands were coded as cases and controls, respectively ($N = 10,496$). Conditional on children being ASD cases, and assuming no indirect paternal genetic effect, this GWAS estimates the indirect maternal genetic effect on children's ASD risk. We found a suggestive association ($rs79371708$; $P = 2.4E-7$) at the *TAF4* locus which is a known causal gene for ASD. Complementary to the maternal effect GWAS, we then performed a transmission disequilibrium test GWAS using the same ASD trios, which estimates the effects of transmitted variants alone and thus quantifies the contribution of direct genetic effects. Direct and maternal ASD effects had a moderate genetic correlation of 0.45 ($SE = 0.2$), suggesting a substantially different genetic basis. Genetic correlations between direct/maternal ASD effects and other complex traits also showed divergent patterns. We found significant correlations between maternal ASD effects and major depression, cigarette smoking, and a younger age at first birth. Instead, direct genetic effects on ASD showed significant correlations with cognition and educational attainment. These results showed compelling evidence for indirect maternal genetic effect on ASD and provided critical new insights into how self and maternal genotypes jointly influence ASD risk in children.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3384. eVai's "Suggested Diagnosis" feature: a new AI-based method to increase diagnostic yield in Rare Disease Patients

Authors:

G. Nicora¹, F. De Paoli¹, **I. Limongelli¹**, E. Rizzo¹, R. Bellazzi², P. Magni², S. Zucca¹; ¹enGenome srl, Pavia, Italy, ²Dept. of Electrical, Computer and BioMed. Engineering, Univ. of Pavia, Pavia, Italy

Abstract Body:

Introduction

Variant interpretation (VI) is a complex process whose aim is to identify pathogenic events. The ACMG/AMP guidelines define a widely adopted five-tier system to interpret genomic variants. Despite the importance of these guidelines, class-based systems may not be effective in clinical settings, when thousands of variants per-proband need to be examined to detect few causatives. eVai (www.engenome.com) is a SaaS platform that classifies variants according to ACMG/AMP guidelines. To further boost VI, eVai is equipped with a new feature to Suggest the Diagnosis through a machine learning approach that prioritizes proband's variants according to their clinical score.

Materials and Methods

We developed a "Suggested Diagnosis" model, that assigns a clinical score to each variant based on 3 features sets accounting for: 1) variant pathogenicity including the eVai score proportional to the ACMG/AMP levels of evidence; 2) phenotypic similarity based on Human Phenotype Ontology Terms provided by the user; 3) family segregation information. Outputted clinical scores are exploited to rank probands' variants and to suggest putative causatives. The model was trained on a large in-house dataset of bona-fide variants from diagnosed patients and their family members. We applied this model within the Rare Genome Project CAGI 6 (Critical Assessment of Genome Interpretation) challenge, led by experts at the Broad Institute of MIT and Harvard.

We selected 30 challenging samples from the "Deciphering Developmental Disorder" (DDD) study (PMID:25533962) for which 43% of the causative variants were VUS/LB with a PS <5. We compared this clinical score with eVai's PS.

Results

Our model resulted as a best performer among worldwide solutions on CAGI probands, and it enabled the diagnosis of 2 unsolved cases, increasing the diagnostic yield by 12.5%. On 30 DDD probands, the causative variants were ranked in the top 20 positions by their clinical score in 28 cases, while the eVai score ranked the causative in the top 20 positions in 19 cases.

Conclusion

We developed a Suggested Diagnosis methodology combining guidelines, phenotypic and family information. As demonstrated by the international CAGI 6 RGP challenge, our method ranks causatives in top positions, enhancing diagnostic yield in Rare Patients. On sampled DDD cases, the Suggested Diagnosis boosts eVai's capability to pinpoint the correct diagnosis in the very first positions by 33%.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3385. Evaluating differential expression of imputed miRNA expression in the subgenual anterior cingulate cortex.

Authors:

J. Drake^{1,2}, Z. Taylor^{1,2}, A. Denham^{1,2}, N. Gillespie³, S-A. Bacanu³, J. Shin⁴, T. Hyde⁴, V. Vladimirov^{1,2,4}, ¹Texas A&M Univ., College Station, TX, ²Dept. of Psychiatry, Coll. of Med., Univ. of Arizona, Phoenix, AZ, ³Virginia Commonwealth Univ., Richmond, VA, ⁴Lieber Inst. for Brain Devel, Baltimore, MD

Abstract Body:

Background: As the disease-relevant organ in psychiatric disorders is the brain, postmortem brain gene expression (GE) studies have been conducted to elucidate the neuropathological changes associated with mental disorders. MiRNAs, a class of small non-coding RNAs, have been shown to play important role in regulating gene expression. However, a major challenge using postmortem brain tissues is the presence of a myriad of potential confounds. Thus, to mitigate these issues, several GE imputation algorithms (such as PrediXcan) were developed that impute GE using genomic and expression data in unrelated subjects. In this study, we use PrediXcan to impute miRNA expression in brain and incorporate neuronal cell fractions to improve accuracy. **Methods:** We used miRNA expression data from a large postmortem brain sample and identified 947 miRNAs in 280 subjects (142 MD cases and 138 controls), on whom we also have available GWAS data. For increased robustness, we applied a 10-fold cross-validated elastic net model for each miRNA. To increase the accuracy of our models, we included cell brain fractions, which were derived using single cell GE data in CIBERSORT. Models with an average coefficient of determination (R^2) > 0.01 and a Z-score p-value < 0.05 were retained and used to impute the miRNA expression from the genotype information alone. To validate our ability to impute expressions for disease relevant miRNAs, a differential expression (DE) analysis on the imputed miRNAs was done and compared to the measured DE analysis. **Results:** In the model containing genomic and expression data only, 89 miRNAs were retained (average $R^2 \geq 0.16$), of which 9 miRNA were found to be DE and 5 overlapping with the measured DE analysis. The inclusion of cell fractions in the model led to a substantial increase in the number of miRNAs with imputed expressions, i.e., 518 miRNAs were retained (average $R^2 \geq 0.23$) with 41 DE and 8 overlapping with the measured DE analysis. We observed the inclusion of cell fractions to has a significantly more pronounced effect on the accuracy and number of imputed miRNA. We further observed that the impact of cell fractions on the accuracy of miRNA GE is more pronounced for miRNA whose expression is in the top quantile (TMM log2 mean expression of 9.27) compared to the bottom quantile (TMM log2 mean expression of -0.17), i.e., R^2 of 0.26 vs 0.12, respectively; Welch Two Sample t-test p-value = 4.3×10^{-12} . **Conclusion:** Here we've expanded on previous gene imputation efforts by modeling miRNA in the subgenual anterior cingulate cortex. Additionally, the inclusion of cell fractions highlights how tissue heterogeneity and expression levels impact imputation accuracy.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3386. Evaluating human genetic evidence for existing drug targets in the UK Biobank

Authors:

G. McInnes, D. Jakubosky, A. Leverentz, Z. Romer, B. Gonzalez, N. Karra, B. Cajés, S. Bruse, O. Gottesman; Empirico Inc., San Diego, CA

Abstract Body:

Approximately 90% of drugs that enter clinical trials fail due to insufficient efficacy or concerns regarding safety. Previous studies have shown that drug targets with supportive human genetic evidence for the target-indication pair are up to twice as likely to proceed from phase 1 to approval as those without, providing a strong rationale for incorporating human genetics into the drug discovery process at the earliest possible stage. Large-scale biobanks linked with rich phenotypic information enable the evaluation of associations between sequence variants and thousands of clinically and therapeutically-relevant phenotypes. Analyses focused on genetic variation that is predicted to modulate the function or abundance of a given protein can provide insight into the phenotypic effects of targeting that protein pharmacologically. Rare nonsynonymous variants, particularly loss-of-function variants (LoFs), are among the most powerful naturally occurring tools for understanding the phenotypic effects of pharmacological modulation of a target, however, previous studies investigating genetic support for existing drug targets have largely focused on evaluating associations with common variants in the target gene locus. We utilized data from up to 452,401 individuals in the UK Biobank to evaluate associations between rare nonsynonymous variants in approximately 700 drug target genes (representing over 1,500 drugs), their approved indications and known adverse effects. Indications and adverse/side effects for the approved drugs were derived from Open Targets and mapped to relevant phenotypes in the UK Biobank. Rare predicted-deleterious variants in drug target genes and gene burdens (aggregating carriers of those variants) were tested for association with more than 3,800 paired approved indications using whole genome regression analysis. We find that a substantial proportion of genes targeted by approved drugs have supportive genetic evidence for their approved indication or a related phenotype. Additionally, by conducting phenome-wide association analyses across thousands of binary and quantitative phenotypes, we find an enrichment for phenotypes representing known adverse and side-effects. We further delineate results by mechanism of action (i.e., agonist vs. antagonist), therapeutic area, and target gene conservation. Overall, this study highlights the prevalence of genetic evidence for approved drugs, and the utility of rare nonsynonymous variants as powerful tools for drug discovery and development. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3387*. Evaluating strategies to infer continental ancestry using single-cell RNA-seq datasets: an application to the Human Cell Atlas

Authors:

J. Yao¹, S. Gazal^{2,3}; ¹Biostatistics, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ²Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ³Ctr. for Genetic Epidemiology, Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA

Abstract Body:

Characterizing the ancestry of samples in genetic and genomic studies is critical to ensure homogeneity of the datasets or to depict ancestry-specific patterns. Single-cell RNA-seq (scRNA-seq) has emerged as a powerful technique to investigate genome regulation at the cell-type level. However, the information on the ancestry of the samples is often not available. Although it should be feasible to infer ancestry using genetic polymorphisms detected on the reads of scRNA-seq datasets, the low number of polymorphisms present in RNA and genes expressed in scRNA-seq make it unclear how accurate such an approach would be.

Here, we developed and evaluated a pipeline to detect genetic polymorphisms (SNPs) present in scRNA-seq datasets and assessed different strategies to infer the continental ancestry of a sample. First, we developed a pipeline using the recommended GATK pipeline on RNA-seq datasets and a quality control procedure optimally selecting SNPs and samples. Second, we applied this pipeline to detect polymorphisms on four large human scRNA-seq datasets of the Human Cell Atlas (HCA; 217 samples from four different tissues). Third, we compared the accuracy of two probabilistic methods inferring the ancestry of an individual. We considered an estimator based on principal component analysis (PCA) and supervised analyses from the admixture software (Alexander et al. 2009). We used 3,481 sequenced individuals from 67 populations of 6 continental groups from the 1000 Genomes and HGDP projects (1000G+HGDP) to define our reference ancestry datasets. To evaluate these methods, we compared ancestry predictions on each population at a time after removing it from the reference ancestry datasets during each analysis.

We detected between 7,126 and 19,762 common SNPs in the 4 HCA datasets (between 3,272 and 7,320 after pruning to reduce linkage disequilibrium). With these sets of pruned SNPs, PCA and Admixture yielded an average prediction error rate of 0.84% and 0.20%, respectively (vs. 0.23% and 0.06% when using all common sequenced variants) within 1000G+HGDP. When applying Admixture (error rates lower than PCA) to the 4 HCA datasets, we consistently observed a high fraction of European ancestry within samples (mean = 88%, median = 96%), indicating the lack of ancestry diversity across datasets. To summarize, our results highlight that continental ancestry could be estimated as accurately when using genetic polymorphisms from scRNA-seq datasets as when using all present polymorphisms. The application to the HCA also demonstrates the lack of ancestry diversity in publicly available scRNA-seq datasets.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3388. Evaluating the accuracy genotype imputation in selected African populations.

Authors:

R. Nanjala; ICIPE, Nairobi, Kenya

Abstract Body:

The Human Leukocyte Antigen (HLA) region plays an important role in autoimmune and infectious diseases. HLA is a highly polymorphic region and thus difficult to impute. We, therefore, sought to evaluate HLA imputation accuracy, especially in an African population due to their high genetic diversity, and this has not been extensively studied. The study sets were selected from the Gambian individuals within the Gambian Genome Variation Project (GGVP) datasets. The Illumina Omni 2.5 array and H3Africa array data were masked from the GGVP datasets using matching markers. The reference datasets were chosen from the 1000 Genomes population (1kg-All), 1000 Genomes African subpopulation (1kg-Afr), 1000 Genomes Gambian subpopulation (1kg-Gwd), Human Hereditary and Health in Africa (H3Africa) population and the HLA multiethnic population via the Michigan imputation server. HLA-A, HLA-B, and HLA-C alleles were imputed using HIBAG, SNP2HLA, CookHLA and Minimac4. The assessment metric was concordance rate. The best performing tool was HIBAG while the best performing reference panel was the H3Africa reference. The H3Africa array and Illumina Omni 2.5 array performance were comparable showing that genotyping arrays have less influence on HLA imputation in African populations. These findings show that using a larger population-specific reference and HIBAG tool improves the accuracy of HLA imputation in African populations.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3389. Evaluating the impact of parallel mutations on the accuracy of statistical phasing

Authors:

A. Beck¹, Y. Si², H. Kang³, S. Zollner⁴; ¹Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, ²Univ. of Michigan, Ann Arbor, MI, ³Univ Michigan, Ann Arbor, Ann Arbor, MI, ⁴Univ Michigan, Ann Arbor, MI

Abstract Body:

With the advent of large studies such as TOPMed and UKBB, sequencing data for hundreds of thousands of diverse individuals can allow for the interrogation of rare variants and their functional associations with complex traits. However, many analyses require phased sequencing data, and modern statistical phasing algorithms routinely generate thousands of switch errors per genome. These errors limit the accuracy in which genotypes can be imputed and also impact our ability to identify high-interest variants amongst the myriad of rare variants discovered in large sequencing studies. A potential contributor to such phasing errors is recurrent mutation, which have recently been observed to occur more commonly along the genome than previously thought. Such parallel mutations lead to variants identical by state to appear in multiple haplotype backgrounds.

To assess the performance of phasing algorithms and the influence of recurrent mutation, we simulate diploids with known phasing by pairing male X chromosomes from the 1000 Genomes Project deep-sequencing data. We then infer the phase using SHAPEIT4, EAGLE2, and BEAGLE5 and compare the results to the known phase to generate distributions of errors. On average BEAGLE5 introduces total errors at a higher rate (0.0159 errors per heterozygous site) than both EAGLE2 (0.0148) and SHAPEIT4 (0.0142). In contrast, SHAPEIT4 introduces flips, the occurrence of consecutive switches at heterozygous sites, at the highest rate among the methods, with 0.0058 flips per heterozygous sites compared to EAGLE2 and BEAGLE5's rates of 0.0053 and 0.0052, respectively. The location of phasing errors is highly correlated between methods. To evaluate the impact of parallel mutation, we assess the frequency of CpGs at switch errors and flips, as these sites have the highest mutation rate and typically show evidence of parallel mutations. We find that across all three methods, around 15% of switches and flip errors occur at CpGs, while CpGs constitute 14% of all heterozygous sites, a 1.08-fold enrichment. We also assess the enrichment of other highly mutable subtypes among the locations of phasing errors. Our results show a mild enrichment for CpGs at flips and switches, suggesting that parallel mutations are only a minor contributor to errors in statistical phasing at the site of the mutation. To ensure that this result is not specific to CpGs, other genomic contexts shown to be associated with recurrent mutation should be assessed for enrichment among phasing errors. Further analysis is required to understand if modeling of parallel mutation may improve the statistical inference of phase.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3390. Evaluation of GENESIS, SAIGE and REGENIE for genome-wide association study of binary traits in correlated data

Authors:

A. Gurinovich¹, A. Leshchik², M. Li², H. Bae³, Z. Song⁴, T. Perls⁵, P. Sebastiani¹; ¹Tufts Med. Ctr., Boston, MA, ²Boston Univ., Boston, MA, ³Oregon State Univ., Corvallis, OR, ⁴Boston Univ. Sch. of Publ. Hlth., Allston, MA, ⁵Boston Univ. Sch. of Med., Boston, MA

Abstract Body:

Performing a genome-wide association study (GWAS) with a binary phenotype using family data is a challenging task. Using linear mixed effects models is typically unsuitable for binary traits, and numerical approximations of the likelihood function may not work well with rare genetic variants with small counts. Additionally, imbalance in the case-control ratios poses challenges as traditional statistical methods such as the Score test or Wald test perform poorly in this setting. In the last couple of years, several methods have been proposed to better approximate the likelihood function of a mixed effects logistic regression model that uses Saddle Point Approximation (SPA). SPA is implemented in GENESIS, SAIGE and REGENIE software: three increasingly popular tools that were developed to perform GWAS of binary traits in correlated data. We compared Score and SPA tests using real family data to evaluate computational efficiency and the agreement of the results. Additionally, we compared various ways to adjust for family relatedness, such as sparse and full genetic relationship matrices (GRM) as implemented in both GENESIS and SAIGE, and polygenic effect estimates as implemented in REGENIE. We used the New England Centenarian Study imputed genotype data and the binary phenotype of human extreme longevity to compare the agreement of the results and tools' computational performance. The evaluation suggests that all softwares produce similar, although not identical, results, with SPA adjustment performing better than Score tests. Our evaluation also demonstrates the importance of adjusting by full GRM, especially in small datasets with family data.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3391. Evaluation of germline variants in pediatric low grade glioma using non-negative matrix factorization

Authors:

L. Chapman Hannah^{1,2}, J. Kim¹, D. Stewart¹, Z. Boukouvalas²; ¹Natl. Cancer Inst., Rockville, MD, ²American Univ., Washington, DC

Abstract Body:

Approximately 8-10% of children with cancer harbor rare, pathogenic or likely pathogenic (P/LP) germline variants in known tumor-predisposition genes. Common variants have smaller effect sizes and collectively may affect risk. Most children with brain tumors, however, do not harbor known risk variants. To identify additional risk alleles, we used non-negative matrix factorization (NMF), a feature-reduction algorithm commonly used in somatic mutation analysis, to evaluate common and rare germline variants in a set of low grade gliomas (LGG) from the Pediatric Cancer Genome Project (PCGP) and the St. Jude LIFE (SJLIFE) Study. NMF is an unsupervised method that reduces high dimensional feature spaces to a latent feature space. In this pilot study, variants from chromosomes 17 (harboring known tumor-predisposition genes TP53, NF1 and BRCA1) and 18 (control chromosome) were evaluated. Whole genome sequencing (WGS) data from 51 LGG patients and 44 non-cancer controls from PCGP and SJLIFE were downloaded from St. Jude Cloud. Single nucleotide variants (SNVs) from WGS were called using an ensemble-based pipeline (GEMSCAN) and then filtered based on a set of quality metrics. In the following study NMF was used to identify a latent set of features (variants) within chromosomes 17 and 18, and as a result we are able to evaluate underlying groups within the data. We hypothesized that distinct clusters of cases and controls can be identified within chromosome 17. The analysis included 610,536 germline variants from chromosome 17 for NMF analysis, and 637,822 germline variants for chromosome 18. In this study, NMF is used to extract sets of germline variants from a high dimensional feature space from patients and non-cancer controls. As hypothesized, there was a higher frequency of variants in cancer susceptibility genes in cases versus controls within chromosome 17. NMF-based clustering showed the greatest separation between cases and controls in two regions of chromosome 17 : chr17: 29361553-33291117, which contains NF1 (known to increase risk for brain tumors) and chr17:60189-121767 which contains lincRNA RP11-1228E12.2. The results of these studies could help identify putative germline variants that collectively contribute to cancer risk.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3392. Exome sequencing of 628,388 individuals identifies common and rare variant associations with clonal hematopoiesis phenotypes

Authors:

M. Kessler¹, A. Damask¹, S. O'Keeffe¹, N. Banerjee¹, D. Li¹, K. Watanabe¹, A. Marcketta¹, M. Van Meter², S. Semrau², J. Horowitz¹, J. Tang¹, J. Kosmicki³, V. Rajagopal¹, Y. Zou⁴, Y. Houvras², A. Ghosh⁵, C. Gillies⁶, J. Mbatchou⁷, R. White¹, N. Verweij⁸, J. Bovijn⁹, N. Parikshak¹, M. LeBlanc¹⁰, M. Jones¹¹, Regeneron Genetics Center, GHS-RGC DiscovEHR Collaboration, D. Glass¹¹, L. Lotta¹¹, M. Cantor¹, G. Atwal¹, A. Locke¹², M. Ferreira¹, R. Deering¹¹, C. Paulding¹¹, A. Shuldiner¹³, G. Thurston¹¹, W. Salerno¹⁴, J. Reid¹, J. Overton¹, J. Marchini¹, H. Kang¹, A. Baras¹, G. Abecasis², E. Jorgenson¹; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Regeneron Pharmaceuticals, Tarrytown, NY, ³Regeneron, White Plains, NY, ⁴Regeneron, Chicago, IL, ⁵Regeneron Genetics Ctr., Jersey City, NJ, ⁶Univ MICHIGAN, Ann Arbor, MI, ⁷Regeneron Genetics Ctr., TARRYTOWN, NY, ⁸Regeneron Genetics Ctr., New York, NY, ⁹Regeneron Genetics Ctr., Oxford, United Kingdom, ¹⁰Regeneron, Bronx, NY, ¹¹Regeneron, Tarrytown, NY, ¹²Regeneron Genetics Ctr., Saint Louis, MO, ¹³Regeneron Genetics Ctr., tarrytown, NY, ¹⁴Regeneron, New York, NY

Abstract Body:

Clonal hematopoiesis (CH) refers to the expansion of certain blood cell lineages and has been associated with aging and adverse health outcomes. Here, we use exome sequence data on 628,388 individuals to identify 40,208 carriers of clonal hematopoiesis of indeterminate potential (CHIP). Using genome-wide and exome-wide association analyses, we identify 24 loci (20 novel) where germline genetic variation influences CHIP predisposition, including missense variants in the DNA-repair gene *PARP1* and the lymphocytic antigen coding gene *LY75* that are associated with reduced incidence of CHIP. We also identify novel rare variant associations with CH and telomere length. Analysis of 5,041 health traits from the UK Biobank (UKB) found relationships between CHIP and severe COVID outcomes, cardiovascular disease, hematologic traits, malignancy, smoking, obesity, infection, and all-cause mortality. Longitudinal and Mendelian Randomization analyses revealed that CHIP is associated with solid cancers, including non-melanoma skin cancer and lung cancer, and that *DNMT3A*-CHIP is associated with the subsequent development of myeloid but not lymphoid leukemias. Our findings demonstrate that CHIP represents a complex set of heterogeneous phenotypes with shared and unique germline genetic causes and varied clinical implications.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3393. Exploiting the mediating role of the metabolome to unravel transcript-to-phenotype associations

Authors:

E. Porcu^{1,2,3}, C. Auwerx^{1,2,3,4}, M. Sadler^{2,3}, A. Reymond¹, Z. Kutalik^{2,3,4}, ¹Ctr. for Integrative Genomics, Univ. of Lausanne, Lausanne, Switzerland, ²Swiss Inst. of Bioinformatics, Lausanne, Switzerland, ³Univ. Ctr. for Primary Care and Publ. Hlth., Lausanne, Switzerland, ⁴Dept. of Computational Biology, Univ. of Lausanne, Lausanne, Switzerland

Abstract Body:

Despite the success of genome-wide association studies (GWASs) in identifying genetic variants associated with complex traits, understanding the mechanisms behind these statistical associations remains challenging. Several methods that integrate methylation, gene expression, and protein quantitative trait loci (metQTLs, eQTLs and pQTLs, respectively) with GWAS data have been proposed to determine their causal role in the path from genotype to phenotype. However, the role of metabolic QTLs (mQTLs) is still underexplored. Here, we developed and applied a multi-omics Mendelian randomization (MR) framework to study how metabolites mediate the effect of gene expression on complex traits. In particular, we combined GWAS, eQTL, and mQTL data in an integrative MR analysis consisting of three steps. First, we map the transcriptome to the metabolome by identifying causal associations between transcripts and metabolites. Next, we screen the metabolites for downstream causal effects on 28 complex phenotypes, resulting in the identification of gene expression -> metabolite -> phenotype cascades. In parallel, we prioritize trait-associated genes by testing the association of transcripts with phenotypes. Third, for transcripts identified in at least one of the previous steps we test whether the identified target genes exert their effect on the phenotype through the metabolite using multivariable MR. In total, we identified 206 transcript-metabolite-trait causal triplets. Sixty-seven of these associations were missed by classical transcriptome-wide MR, which only uses gene expression and GWAS data. Among these, we identify biologically relevant pathways, such as between *ANKH* and calcium levels mediated by citrate and *SLC6A12* and serum creatinine through modulation of the levels of the renal osmolyte betaine. We show that the signals missed by transcriptome-wide MR are found thanks to the gain in power allowed by integrating multiple omics-layer. Furthermore, through extensive power analyses we show that with larger molecular QTL studies and in case of mediated effects, our multi-omics MR framework outperforms classical MR approaches designed to detect causal relationships between single molecular traits and complex phenotypes.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3394. Exploring germline genetics of *in situ* and invasive cutaneous melanoma

Authors:

N. Ingold^{1,2}, M. Seviiri¹, J. ONG¹, D. R. Nyholt², R. E. Neale¹, N. Pandeya¹, D. C. Whiteman¹, C. M. Olsen¹, N. G. Martin¹, D. Duffy¹, K. Khosrotehrani³, G. Montgomery⁴, N. Hayward¹, S. Macgregor¹, M. H. Law¹; ¹QIMR Berghofer, Brisbane, Australia, ²Queensland Univ. of Technology, Brisbane, Australia, ³The Univ. of Queensland Diamantina Inst., Woolloongabba, Australia, ⁴Univ. of Queensland, St Lucia, Australia

Abstract Body:

While incidence rates of cutaneous melanoma have risen dramatically over recent decades, mortality rates have remained stable. This is partly due to a sharp increase in diagnosis of *in situ* melanoma over the same time period. Risk of cutaneous melanoma has a large germline genetic component relative to other cancers, but it is unclear if there are specific genetic risks for invasive vs. *in situ* melanoma. Detecting differences in germline risk of *in situ* vs invasive melanoma may provide insights that benefit melanoma screening and surveillance programs.

We performed a series of GWAS meta-analysis (including UK Biobank, QSkin Sun and Health Study, Queensland Study of Melanoma (Q-MEGA), and FinnGen) of invasive vs controls, *in situ* vs controls and invasive vs *in situ* to identify genetic loci associated with a diagnosis of invasive melanoma. The effect estimates generated from the invasive vs *in situ* GWAS were then used to generate a polygenic risk score (PRS) to assess associations between genetic risk *in situ* vs invasive status and invasive melanoma.

The GWAS of invasive (n= 7039) vs controls (n= 649048) identified 14 loci previously associated with melanoma, the *in situ* (n= 2961) vs controls GWAS identified 5 loci, of which the *IRF4* and the *GPR98* loci appeared only for *in situ*. The genetic correlation between *in situ* and invasive was indistinguishable from 1 ($r_g=0.98$, 95% CI = 0.67 to 1.3). The *in situ* vs invasive GWAS revealed one genome-wide significant SNP (rs4566922, $P= 3.263e-10$, near gene *SLC35B3*) on chromosome 6. We will report on PRS results.

There was considerable overlap in genes detected for *in situ* vs invasive, as expected. We found two loci unique to the *in situ* vs. control GWAS that have previously been reported as melanoma GWAS loci, and one potentially novel loci in the case-case GWAS. While it is difficult to characterize a mechanism for *SLC35B3*'s role in differential melanoma diagnoses, identifying differing genetic loci between *in situ* and invasive melanoma could indicate differing underlying genetics influencing individual risk of an invasive or *in situ* melanoma. Building a PRS around the heritable difference in risk between invasive vs *in situ* melanoma has the potential to help reduce overdiagnosis by shifting screening away from lower risk and towards higher risk individuals. More work is needed to identify, and replicate, genetic loci associated with melanoma status.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3395. Exploring the genetic, socio-economic and environmental determinants of drug adherence

Authors:

M. Cordioli¹, A. Corbetta¹, H. Kariis², S. Jukarainen¹, T. Kiiskinen¹, FinnGen Study, Estonian Biobank research team, S. Ripatti^{1,3,4}, L. Milani², A. Ganna^{1,5,3}; ¹Inst. for Molecular Med. Finland, Univ. of Helsinki, Helsinki, Finland, ²Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, ³Broad Inst. of MIT and Harvard, Cambridge, MA, ⁴Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland, ⁵Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

One of the major factors behind the efficacy of pharmacological treatments is patients adherence to their prescribed regimen. A comprehensive investigation combining both genetic and socioeconomic factors is lacking. By leveraging genetic data from FinnGen (N = 356,077, >68 million drug purchases) and nationwide socio-economic and health data from FinRegistry (N=5.5 million, > 790 million purchases), we provide a systematic investigation of adherence determinants across 5 classes of medications (statins, blood pressure (BP) medications, breast cancer medications, antiplatelets, anticoagulants).

For each of these, we defined two phenotypes: adherence (ratio between total purchased quantity and length of treatment) and persistence (purchasing for at least one year vs discontinuation after one purchase). We then looked at the associations between the two phenotypes and genetic risk factors (using polygenic scores, PGS), as well as demographic and socio-economic factors (e.g. region of residence, years of education).

The most notable association, shared across all medications, was between adherence and a higher genetic predisposition to participate in follow-up health questionnaires in UK biobank (β : [0.1%:0.3%], P: $[2.5 \times 10^{-12}; 7.6 \times 10^{-3}]$ - β as percentage increase in adherence per PGS standard deviation). Other notable genetic risk factors were higher BMI and diabetes, which both positively correlated with higher adherence to statins (β : 0.2%, P: 1.6×10^{-6} ; β : 0.3%, P: 9.5×10^{-9}) and BP medications (β : 0.4%, P: 2.5×10^{-27} ; β : 0.4%, P: 1.4×10^{-20}).

Associations between PGSs and persistence replicate those observed for adherence, and a PGS for adherence to statins built in Estonian Biobank significantly associated (P: 0.001) with the adherence measured in our data. We further run a GWAS of adherence and persistence to each drug, which identified no genetic variation associated with the two phenotypes.

We then explored the effect of socio-economic factors and found positive associations between adherence to statins and to BP medications and being male (β : [0.8%, 1.5%], P < 2×10^{-16}) and living in Southern Finland (β : [1.5%, 0.7%], P: $[1.5 \times 10^{-5}, 0.004]$), while educational level intriguingly showed a discordant effect between statins (β : 0.3%, P < 2×10^{-16}) and BP medications (β : -0.1%, P: 9.9×10^{-14}).

Overall, results suggest that adherence is related to behavioral traits (e.g. participation in health questionnaires and perception of risk factors) rather than having direct biologic determinants. Further investigation of additional socio-economic determinants will allow for a better identification of patients at risk of being poorly adherent.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3396*. Expression and Splice QTLs found in COVID-19 patients and controls show differential colocalization and heterogeneity between infectious and non-infectious states.

Authors:

Y. Farjoun¹, T. Nakanishi², J. Willett², T. Lu², S. Zhou³, B. Richards²; ¹Lady Davis Inst., Montreal, QC, Canada, ²McGill Univ., Montreal, QC, Canada, ³McGill, Montreal, QC, Canada

Abstract Body:

There are several different explanations as to why expression and splice QTLs do not often colocalize with GWAS results. One of these is that the difference in expression or RNA splicing is localized to a cell state which is not often measured in RNA experiments. Here we present eQTL and sQTL results calculated on RNA samples from release 7 of The BQC19 (*La Biobanque Québécoise de la COVID-19*). Data were processed by following the GTEx eQTL and sQTL pipelines. The differences between infectious and non-infectious states (defined by comparing the sample date to the onset of symptom date), were highlighted by calculating QTLs on the different sets of samples and following up by a test of heterogeneity. To avoid the results being confounded by population, the results were stratified into the two main populations in the study: NFE (Non-Finnish European) and AFR (African/African-American). The resulting 4 subsets range between 50 and 200 samples in size. The QTL calculation was followed up by testing for colocalization with the HGI loci. The “classical” COVID-19 loci, are present and colocalize as eQTL (e.g., *ABO*) or sQTLs (e.g., *OAS1*) in both infectious and non-infectious samples, however, several HGI loci colocalize differentially in infectious-state and noninfectious state. For example, *NAPSA* and *TYK2* only colocalize with the HGI GWAS for the non-infectious samples, while *IL10RB* only colocalizes for infectious samples. While the overlapping *IFNAR2* gene is the most reliable causal gene for the chr21 locus, this indicates that *IL10RB* may be involved in pathogenesis. In addition to the colocalization results, we compared the QTLs and found several examples where the effect differs significantly between infectious and non-infectious states. These results indicate that expression data that are taken at the appropriate time can colocalize well with GWAS data and highlights several new loci as relevant for infection response to SARS-CoV-2. Our results lay the foundation for further analysis to explore the causal role of gene expression and splicing in COVID-19 outcomes. All data and pipelines are publicly available.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3397. Extending Genome-Wide Association Studies to admixed cohorts with high degrees of relatedness.

Authors:

T. Tan¹, N. N. Shah¹, G. E. Tietz¹, P. Turley², W. Zhou³, E. G. Atkinson¹; ¹Baylor Coll. of Med., Houston, TX, ²Univ. of Southern California, Los Angeles, CA, ³Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

Admixed populations comprise more than one-third of the US population but are extremely underrepresented in Genome-Wide Association Studies (GWAS). This lack of inclusion of admixed populations in GWAS contributes to disparities of the applicability and benefits of findings from health research. Recently, more admixed cohorts are being included in large biobanks and cohorts, such as All of Us, reducing gap of resources but also introducing new statistical and computational challenges. Our group has previously developed a generalized linear model (GLM) based method, *Tractor*, for GWAS in admixed samples, which allows for obtaining accurate ancestry-specific effect sizes and boosting discovery power in the presence of effect size heterogeneity across ancestries. However, most GLM-based software makes a strong, but often inappropriate, assumption of sample independence, whereas in actual large-scale collections, samples are generally related at either ancestral or familial level. New, scalable approaches are therefore necessary that can appropriately account for sample relatedness in all levels for admixed populations.

In this project, we propose a new generalized linear mixed model (GLMM) method for association studies in large biobanks or cohorts with admixed populations, while accounting for confounding factors, such as sample relatedness and population substructure. Similar to *Tractor*, our method conducts genetic association tests by leveraging the power of observing granular ancestral patterns with local ancestry inference to produce more accurate effect sizes, localize association signal, and boost power in the presence of effect size heterogeneity across ancestries.

We constructed ancestry-specific genetic relationship matrices and formulated an ancestry-aware mixed-effect model with multiple variance components, with each one capturing ancestry-level variability. We evaluated the performance of our method through extensive simulation studies and real data analysis and showed its various merits compared to the GLM-based GWAS method, such as effective type 1 error control. Our work will particularly increase the inclusiveness of underrepresented admixed cohorts in GWAS efforts and help reduce health disparity in the long term.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3398. Extremely sparse models of linkage disequilibrium in diverse GWAS

Authors:

A. Wohns¹, P. S. Nowbandegani¹, J. L. Ballard¹, B. M. Neale^{1,2}, A. Bloemendal¹, L. J. O'Connor¹; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

Statistical methods to analyze genetic association data must account for linkage disequilibrium (LD). LD correlation matrices are large (sometimes terabytes) and methods that rely on them are slow (sometimes taking days). The challenge is exacerbated by ancestral diversity, as LD patterns vary with ancestry. Here, we show that LD can be modeled using extremely sparse matrices, hundreds of times smaller than a correlation matrix, enabling efficient algorithms. Our approach leverages genome-wide genealogies: statistical relationships among linked alleles correspond to genealogical relationships among ancestral haplotypes, and we extract these relationships to produce an LD graphical model (LDGM). We derive LDGMs from genome-wide genealogies, validate their accuracy across continental populations, and demonstrate an LDGM-based method for heritability partitioning.

Genome-wide genealogies can be inferred at scale across ancestrally diverse datasets (Wohns et al., 2022 Science), producing efficient representations of genotype data (Kelleher et al., 2019 Nat Genet). Based on the genealogy we identify conditional dependencies between variants, which are the edges of the LDGM. We prove that the LDGM has desired properties (correctness and minimality). We also infer the LDGM precision matrix, which is a sparse regularized inverse of the population-specific LD correlation matrix. We generated LDGMs across 11M common SNPs (MAF<0.01) in 1000 Genomes (N=2504) and inferred LDGM precision matrices for five continental populations (AFR, AMR, EAS, EUR, SAS). These matrices were extremely sparse, with 20 nonzero entries per row and column. They accurately reproduced the LD correlation matrix, with a mean squared error of 0.0012 for four out of five populations (0.0022 in the admixed AMR population). They were equally accurate on average in UK Biobank European samples (MSE=0.0012).

We developed a heritability partitioning method that operates on LDGMs and GWAS summary statistics. This method combines an accurate maximum-likelihood estimation approach (similar to GCTA) with a flexible model that can handle any number of overlapping annotations (similar to S-LDSC), and it can be applied to GWAS in non-European populations. This method produces more accurate enrichment estimates than S-LDSC in simulations.

In other statistical applications as well, slow operations involving correlation matrices can be replaced with fast ones involving our sparse precision matrices, producing order-of-magnitude improvements in runtime. We will publish our LDGMs with open-source software tools enabling their integration into new and existing methods.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3399. Fine-mapping one chromosome at a time: Software and an initial application to UK Biobank

Authors:

K. Yuan¹, M. Lam², C-Y. Chen³, Y. Chen⁴, M. Yu⁵, T. Ge¹, H. Huang⁶; ¹Massachusetts Gen. Hosp., Boston, MA, ²The Broad Inst., Cambridge, MA, ³Biogen, Cambridge, MA, ⁴broad Inst., Cambridge, MA, ⁵Broad Inst., Cambridge, MA, ⁶MGH, Boston, MA

Abstract Body:

Fine-mapping seeks to find the genetic variants underlying the complex traits or diseases. Current statistical fine-mapping methods jointly model the effects of multiple significant variants. The computation cost dramatically increases as more variants are involved, making it almost impossible to apply to whole chromosome data. Researchers solved this by pre-selecting associated loci and simplifying the problem to fine-mapping a hundreds-of-variants locus. However, the loci definition methods are mainly LD-based and problematic when considering multiple populations. Due to the population-specific LD, the loci definitions might be inconsistent across populations. Even after simplification, most fine-mapping software still cannot handle long-range LD regions, such as the Major Histocompatibility Complex (MHC).

SuSiEx is a cross-population fine-mapping method developed by our group. We extended the single-population fine-mapping method, SuSiE, by considering the causal variants shared across populations. Here, we refactored the code to reduce the memory by designing a lite matrix data structure. We boosted the software speed by enabling multi-thread computing and buffered I/O. The software's speed is ~10x faster, and memory consumption is reduced by more than 90%, making it possible to fine-map a full chromosome.

We applied our software to the alanine aminotransferase of Pan-UKBB European chromosome 22 association results. The memory consumption is ~74GB, and the running time is ~2.36 hours with 40 CPUs (clock speed 2.10GHz). A proper server could be easily found on the Google Cloud Platform (GCP), and the computing cost is less than \$10. The whole chromosome fine-mapping found two credible sets less than the loci-based fine-mapping. The further conditional analysis confirmed that one of the missing credible sets was weakly ($r=0.0419$) tagged to a credible set of a different locus. After conditional on a closer credible set, the p-value of another missing credible set's leading variant increased to $2.5e-8$, much closer to the genome-wide significance threshold.

Our software is the first to fine-map a full chromosome at a time with computational resources currently available from commercial cloud providers. With our software, researchers will no longer be bothered by defining loci before fine-mapping analysis. The loci definition inconsistencies will not be a problem in the cross-population fine-mapping. Besides the convenience, applying our software to whole chromosome data could reduce the false-positive discoveries caused by relatively weak but long-range LD.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3401*. Functional Annotations-Informed Whole Genome Sequence Analysis Identifies Novel Rare Variants for AD in the Alzheimer's Disease Sequencing Project

Authors:

S. Lee¹, B. Shi¹, G. Peloso², Y. Wang², N. Heard-Costa³, H. Lin⁴, A. Pitsillides², C. Sarnowski⁵, E. Boerwinkle^{5,6}, P. De Jager⁷, J. Dupuis², S. Seshadri⁸, E. Wijsman⁹, A. DeStefano², M. Fornage^{1,5}; ¹Univ. of Texas Hlth.Sci. Ctr. at Houston McGovern Med. Sch., Houston, TX, ²Boston Univ. Sch. of Publ. Hlth., Boston, MA, ³Boston Univ Sch Med., Boston, MA, ⁴UMass Med. Sch., Worcester, MA, ⁵Univ. of Texas Hlth.Sci. Ctr. at Houston, Sch. of Publ. Hlth., Houston, TX, ⁶Baylor Coll. of Med., Houston, TX, ⁷Columbia Univ Med Ctr, New York, NY, ⁸Glenn Biggs Inst. for Alzheimer's and Neurodegenerative Diseases Univ. of Texas Hlth.Sci. Ctr. at San Antonio, San Antonio, TX, ⁹Univ Washington Sch Med, Seattle, WA

Abstract Body:

Background: Incorporation of functional annotations improves power to identify rare variants (RVs) associated with disease with whole-genome sequencing (WGS) data. We applied an omnibus test in the variant-Set Test for Association using Annotation infoRmation (STAAR-O) framework, incorporating Alzheimer's Disease (AD) specific annotations based on partitioning heritability to identify RVs associated with AD in the multi-ancestry sample of the Alzheimer's Disease Sequencing Project (ADSP). **Method:** We performed association analyses in a sample of 4,074 individuals (1,668 AD cases; 2,406 controls) with WGS as part of the ADSP Discovery/Extension Phase. For each numerical functional annotation from the WGS Annotator, we partitioned all single nucleotide variants (SNVs) into q discrete categories of the continuous functional score ($q=10$). We estimated AD heritability from SNVs in those categories using GCTA. For each functional annotation, we then derived a global weight estimated from the linear regression slope as the average per-SNV AD heritability on the q categories. Among weighted functional scores, we selected the top 12 based on the rank of global weights to leverage the most informative functional annotations for incorporation into the STAAR-O test. We performed an agnostic region-based association analysis using sliding windows, defined as 2kb in length with a skip length of 1kb. Association tests were performed on RVs with minor allele frequency (MAF) $<1\%$, adjusting for sex, sequencing center, platform, study, 4 principal components, and a genetic relatedness matrix. Replication of significant associations was carried out in an independent sample of 10,083 individuals (5,717 AD cases, 4,366 controls) as part of the ADSP Follow-Up Study, using the same analytical models. **Result:** Out of 2.65 million 2-kb overlapping windows with a total minor allele count >10 , two non-consecutive windows on chromosome 17 were associated with AD ($P < 5 \times 10^{-8}$). Both window associations were replicated ($P=0.003$ and 0.016). The top variant of one significant window was rs534148850 (MAF= 0.0005, $P=5.6 \times 10^{-9}$) located downstream of *PLEKHM1P1*, a pseudogene, and *MIR4315-2*, a microRNA with predicted targets enriched in apoptosis pathways. The top variant of the second window was rs532055552 (MAF=0.005, $P=6.2 \times 10^{-9}$) located downstream of *CEP112*, a coiled domain-containing protein involved in the regulation of gamma-aminobutyric acid A receptor surface expression. **Conclusion:** By incorporating AD-relevant functional annotations to a powerful RV association framework, we discovered and replicated two novel genetic regions harboring RVs associated with AD.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3402. Genealogy-wide association improves detection of rare and low-frequency variants in under-sequenced populations.

Authors:

A. Gunnarsson¹, B. Zhang², P. Palamara²; ¹Wellcome Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, ²Dept. of Statistics, Univ. of Oxford, Oxford, United Kingdom

Abstract Body:

The ancestral recombination graph (ARG) compactly represents the genealogical history of a set of individuals at every site along the genome, providing information on both observed and unobserved genomic variation that can facilitate association and other complex trait analyses. We developed an algorithm to efficiently infer large-scale ARGs from SNP array and sequencing data, improving coalescence time inference by extending the sequentially Markovian coalescent model (McVean et al., 2005) to account for allele ages and ARG topology.

We compared our approach to existing ARG reconstruction methods in extensive simulation of SNP array data and observed increased accuracy, including a 36% improvement in Kendall-Colijn distance compared to ARG-Needle (Zhang et al., bioRxiv). We used simulated phenotypes to evaluate linear mixed model association testing of putative unobserved variants in ARGs inferred from array data. While not requiring a sequenced panel, this genealogy-wide association (GeWAS) strategy matched the statistical power of standard GWAS analysis using imputation from thousands of sequenced samples, remaining robust to population structure and SNP ascertainment bias.

We inferred the ARG for UK Biobank individuals from five non-European ancestries, including African (N=6,747), American (N=993), Central/South Asian (N=9,009), East Asian (N=2,763) and Middle Eastern (N=1,616), only using array data. We performed GeWAS testing using the inferred ARGs and compared to a standard GWAS on genotype data imputed using the HRC+UK10K reference panel (N=65k predominantly European haplotypes) in height and six blood-related traits. We used permutation testing to establish genealogy-wide significance thresholds ranging between 1.5×10^{-9} and 6.8×10^{-10} . After stringent LD-based filtering to extract approximately independent associations we found 192 imputed markers and 128 GeWAS markers, 19 of which remained significant after conditioning on any imputed marker. To validate each GeWAS marker, we extracted the most highly correlated variant found in a subset of 9,276 whole-exome sequenced samples, finding 88 unique exome variants. GeWAS markers strongly tagged (average $r=0.74$) these underlying sequencing variants, 8 of which were better tagged than by any imputed variant, and included ARG-specific low-frequency (MAF < 1%) associations with loci harboring common variant signals in European samples.

These results support the use of genealogy-based approaches to complement genotype imputation in the analysis of unobserved genetic variation, particularly for groups that are underrepresented in large-scale sequencing panels.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3403. Gene-by-gene and gene-by-sex interactions identified in the UK Biobank using Explainable Boosting Machines.

Authors:

Y. Nagafuchi¹, H. Hidayatullah², D. Mas Montserrat³, A. G. Ioannidis⁴; ¹Stanford Univ. Inst. for Computational and Mathematical Engineering, Stanford, CA, ²Stanford Univ. Dept. of Statistics, Stanford, CA, ³Stanford Univ. Dept. of BioMed. Data Sci., Stanford, CA, ⁴Stanford Univ. Dept. of BioMed. Data Sci., Inst. for Computational and Mathematical Engineering, Stanford, CA

Abstract Body:

We apply Explainable Boosting Machines (EBMs), which are tree-based, cyclic gradient boosting Generalized Additive Models, to the phenotype prediction problem by using single nucleotide polymorphisms identified via genome-wide association studies in the UK Biobank White British individuals.

This approach allows us to maintain interpretability while identifying interacting pairs of features important to explaining the phenotype in question. To detect pairwise interactions, the model first fits the main effects and then efficiently searches all pairwise feature interactions to fit the remaining residuals. Investigating the epistatic interactions identified by this EBM approach, we found both novel and previously-investigated interactions. In addition, we identified strong non-linear interactions between sex and a number of genomewide significant SNPs. Finally, the non-linear functions modeled by our EBMs allowed us to delineate dominant from recessive acting SNP associations and to model the different effect sizes in homozygous and heterozygous individuals.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3404. Gene-sodium intake interactions reveal new blood pressure loci in a multi-ancestry genome wide association study in UK Biobank and CHARGE

Authors:

M. Kho¹, J. Smith^{2,3}, L. Bielak², P. van der Harst⁴, K. Schwander⁵, T. Kelly⁶, C-T. Liu⁷, A. Correa⁸, C. Miller⁹, P. Peyser², S. Kardia²; ¹Graduate Sch. of Data Sci., Seoul Natl. Univ., Seoul, Korea, Republic of, ²Univ. of Michigan SPH, Ann Arbor, MI, ³Inst. for Social Res., Univ. of Michigan, Ann Arbor, MI, ⁴Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, ⁵Washington Univ. Sch. of Med., St Louis, MO, ⁶Tulane Univ. SPH and Tropical Med., New Orleans, LA, ⁷Boston Univ. SPH, Boston, MA, ⁸Univ. of Mississippi Med. Ctr., Jackson, MS, ⁹Univ. of Virginia Sch. of Med., Charlottesville, VA

Abstract Body:

High sodium intake is an important risk factor for hypertension. Genome-wide association studies (GWAS) have identified thousands of genetic loci associated with blood pressure (BP) traits, but it remains not fully understood whether sodium intake modifies genetic effects. We conducted genome-wide gene by sodium intake (high intake defined by ≥ 220 mmol/day) interaction analyses on BP traits (systolic BP, diastolic BP, mean arterial pressure, and pulse pressure) using the 2 degrees of freedom (df) joint test followed by 1 df test of interaction effects. The study groups included UK Biobank (UKBB) European ancestry participants with daily sodium intake measured from spot urine samples (N=437,725) and participants from the Gene-Lifestyle Interactions Working Group within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium across three ancestry groups (European, African, and Asian ancestries) with daily sodium intake measured from 24-hour or half-day urine samples (N=9,669). We identified 19 and nine previously unreported BP loci in UKBB and CHARGE, respectively, based on the 2df joint test. Through a sensitivity analysis using participants without diuretic use, we found ten additional previously unreported BP loci in UKBB. The most significant locus was rs2299871 in *PPARD*, which has been implicated in several cardiometabolic diseases and plays a role in the stimulation of epithelial sodium channel (ENaC)-mediated renal salt absorption impacting blood volume and pressure. We also identified one and seven significant interactions with sodium intake in UKBB and CHARGE, respectively, using the 1df interaction test. Among them, only one of the interaction loci (*TENM2* in CHARGE European participants) had not been previously reported. Our bioinformatics analyses and functional annotations highlighted several biological pathways and genes implicated in BP regulation through ENaC, atherosclerotic progression, dietary lifestyle behaviors, and education. The druggability analysis suggested that several of our newly identified BP genes are druggable targets or targets of approved drugs and/or compounds under investigation including nutrient supplementation. In conclusion, we identified 39 previously unreported loci that were significantly associated with BP measures through GWAS accounting for interaction with sodium intake, which requires further replication. This study highlights the role of sodium intake interactions in the genetic contribution to BP traits as well as potential to identify druggable targets, suggesting novel insights into sodium intake-related BP regulation and potential interventions.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3405. Genetic association models are robust to common population kinship estimation biases.

Authors:

Z. Hou, A. Ochoa; Duke Univ., Durham, NC

Abstract Body:

Common genetic association studies for structured populations, including Principal Component Analysis (PCA) and Linear Mixed-effects Models (LMM), model the correlation structure between individuals using population kinship matrices, also known as Genetic Relatedness Matrices or “GRMs”. However, the most common kinship estimators can have severe biases that were only recently characterized. Here we characterize the effect of these kinship biases on genetic association. We employ a large simulated admixed family and genotypes from the 1000 Genomes Project, both with simulated traits, to evaluate a variety of kinship matrices (every bias type has two locus weight types, and their theoretical limits for the simulation). Remarkably, we find nearly equal association statistics and performance for kinship matrices of different bias types (when all other features are matched). These empirical observations lead us to hypothesize that these association tests are invariant to these kinship biases, which using linear algebra we prove holds exactly for LMM and approximately for PCA. Our constructive proof shows that the intercept and relatedness (PCs in PCA, random effect in LMM) coefficients compensate for the kinship bias, so the result extends to generalized linear models as long as those coefficients are present and are nuisance parameters. Overall, we find that existing association studies are robust to kinship estimation bias, and our theoretical results may help improve association methods by taking advantage of this unexpected robustness, as well as help determine the effects of kinship bias in other settings.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3406. Genetic determinants of a metabolic model for age

Authors:

S. Xiao¹, M. Argentieri¹, A. Nevado-Holgado¹, K. Nho², M. Arnold³, G. Kastenmüller³, R. Kaddurah-Daouk⁴, N. Amin¹, C. M. Van Duijn¹, Alzheimer Disease Metabolomics Consortium, Accelerating Medicines Partnership Program for Alzheimer's Disease (AMP-AD); ¹Univ. of Oxford, Oxford, United Kingdom, ²Indiana Univ Sch. of Med., Indianapolis, IN, ³Helmholtz Zentrum München, Neuherberg, Germany, ⁴Univ. of Duke, Durham, NC

Abstract Body:

Age is a major driver of common age-related diseases. However, the molecular pathways underlying the effects of age on morbidity and mortality are far from understood. We aimed to construct a blood-based metabolic model for chronologic age and determine its genetic determinants. The study is embedded in UK Biobank and includes 118,021 randomly selected participants who are characterized for 249 metabolites using proton nuclear magnetic resonance (¹H-NMR). We used a gradient boosting ensemble decision tree machine learning method (Light GBM) to construct a sex adjusted model for metabolic age allowing non-linear relationships. Recursive feature elimination was used to reduce the number of metabolites and Shapley values were used to evaluate the impact of metabolites. We conducted a genome-wide association analysis (GWAS) to find genetic determinants of metabolic age. Our final metabolomic age model included 57 metabolites ($R^2 = 0.32$ in test dataset). Albumin, omega 6/omega 3 ratio, citrate, leucine, free cholesterol levels in extra small VLDL, tyrosine and triglyceride levels in large LDL are most important features. To evaluate bias due to existing chronic disease, we repeated the analysis in participants without cancer, cardiologic, lung, liver, metabolic, CNS or locomotor disease at baseline. The correlation between the model based on all participants and those healthy at baseline was 0.980 ($SE=7.24*10^{-4}$). In multivariate Cox models, higher metabolic age at baseline significantly increased the risk of mortality ($p=2.20*10^{-16}$) and incidence of liver ($p=7.87*10^{-7}$), pancreatic ($p=2.11*10^{-2}$) and esophageal ($p=3.65*10^{-2}$) cancer, chronic kidney disease ($p=2.20*10^{-16}$), diabetes ($p=2.20*10^{-16}$), hypertension ($p=4.44*10^{-16}$), chronic liver disease ($p=1.16*10^{-12}$), cardiovascular disease ($p=2.68*10^{-11}$), COPD ($p=5.78*10^{-6}$), vascular ($p=1.09*10^{-3}$) and all cause dementia ($p=9.70*10^{-3}$), osteoarthritis ($p=1.39*10^{-3}$) and obesity ($p=6.61*10^{-3}$). The metabolic age is further associated with blood telomere length ($p=1.06*10^{-8}$) and C-reactive protein ($p=2.20*10^{-16}$) at baseline. Our GWAS identified 27 loci ($p<5*10^{-8}$) associated to the metabolic age score. Regions significant in the GWAS are associated with 47 out of 57 metabolites used to estimate metabolic age but 3 novel regions are uniquely associated to metabolic age: *REV3L* (involved in DNA repair), *C12orf43* (WNT signaling), and *FBXO39* (ubiquitination) but not to any of the metabolites used to construct the score. In conclusion, we have constructed a metabolomics model for age that associates with markers of aging, age related disease and mortality and elucidated its genetic architecture.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3407. Genetic Determinants of Circulating Glycine and Cardiovascular Disease: Causal Relationship or Red Herring?

Authors:

S. Biswas¹, J. R. Hilser¹, N. Woodward², P. Huang², J. Gukasyan², Z. Fouladian², S. Charugundla³, A. J. Lusis³, Z. Wang⁴, W. Tang⁴, S. L. Hazen⁴, J. Hartiala⁵, H. Allayee²; ¹Univ. of Southern California, Los Angeles, CA, ²USC Keck Sch. of Med., Los Angeles, CA, ³UCLA Dept of Med., Los Angeles, CA, ⁴Cleveland Clinic, Cleveland, OH, ⁵Univ of Southern California, Los Angeles, CA

Abstract Body:

Introduction: Circulating glycine levels have been linked to decreased cardiovascular risk in multiple studies. However, given the role of glycine in a wide range of metabolic pathways, evidence for a causal association between glycine and coronary artery disease (CAD) has not been conclusive. We sought to untangle the genetic underpinnings of glycine and its relationship with CAD using Mendelian Randomization (MR) and glycine supplementation in a mouse model of atherosclerosis. **Methods and Results:** We conducted a multi-ethnic weighted Z-score meta-analysis of genome-wide association study (GWAS) results in up to 222,413 subjects of European, Asian, and Hispanic origin from 9 cohorts, including the UK Biobank, Biobank Japan, and the Hispanic Community Health Study. Using Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA), we identified 59 independent loci for glycine, 33 of which were novel. A phenome-wide association scan indicated that most of these variants were associated with other metabolites or CAD risk factors such as BMI, blood pressure, or diabetes (T2D). However, 8 glycine-associated SNPs, including 4 components of the glycine cleavage system (*AMT*, *DLD*, *GLDC*, and *GCSH*), did not yield evidence of pleiotropy. Inverse-variance weighted two-sample MR was performed with the 8 non-pleiotropic SNPs but there was no evidence for a causal association between genetically determined glycine levels and CAD (OR=0.93, 95% CI: 0.66-1.30, p-value=0.66). However, there was modest evidence for causal associations between glycine levels and with systolic blood pressure (Beta= -1.41, SE=0.68, p-value=0.04) and T2D (OR=0.69, 95% CI: 0.53-0.93, p-value=0.03). As a complementary approach, we carried out a dietary glycine study with hyperlipidemic Apoe^{-/-} mice. Feeding of a 2% glycine-supplemented diet led to significant elevations of fasting and non-fasting glycine levels compared to the control 0.3% glycine diet. However, atherosclerotic lesion area was not significantly different between glycine-fed and control groups. **Conclusions:** Our results comprise the largest meta-analysis of glycine to date and yield a more complete picture of the genetic architecture of glycine metabolism. While MR analyses suggested that glycine could be causally related to blood pressure and diabetes, there was no evidence for the same causal relationship with risk of CAD or atherosclerosis in humans or mice.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3408. Genetic evidence of assortative mating in the Japanese population.

Authors:

K. Yamamoto^{1,2}, K. Sonehara¹, S. Namba¹, T. Konuma³, H. Masuko⁴, S. Miyawaki⁵, The Biobank Japan Project, Y. Kamatani⁶, N. Hizawa⁴, K. Ozono¹, L. Yengo⁷, Y. Okada^{1,2,8,5}; ¹Osaka Univ. Graduate Sch. of Med., Suita, Japan, ²Immunology Frontier Res. Ctr. (WPI-IFReC), Osaka Univ., Suita, Japan, ³Japan Tobacco Inc., Takatuki, Japan, ⁴Faculty of Med., Univ. of Tsukuba, Tsukuba, Japan, ⁵Graduate Sch. of Med., the Univ. of Tokyo, Tokyo, Japan, ⁶Graduate Sch. of Frontier Sci., The Univ. of Tokyo, Tokyo, Japan, ⁷The Univ. of Queensland, Brisbane, Australia, ⁸RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan

Abstract Body:

Positive assortative mating (AM) is a commonly observed mating pattern, where individuals with similar phenotypes are prone to get married more often than by chance. Detecting the genetic impact of AM on a polygenic trait has been challenging due to a lack of large-scale spousal data and confounding of population stratification, especially in non-European populations. The recently developed method by Yango et al focused on quantifying the correlation between physically distant trait-associated alleles (i.e., gametic phase disequilibrium [GPD]) using polygenic scores (PGS) derived from partitioned chromosomes (e.g., odd and even numbered chromosomes) without spousal data. Using GPD estimate (θ), they confirmed the genetic evidence of AM on adult height and educational attainment in Europeans. Here, we report the PGS-based analysis of AM in the Japanese population using Biobank Japan Project data and other Japanese independent cohorts ($n = 172,270$). Adopting the leave-one-group-out method, we estimated the GPD estimates across 81 complex traits by calculating the correlation between PGS_{odd} and PGS_{even} with the robust adjustment for population stratification. Our study identified that cardiometabolic diseases (type 2 diabetes [T2D] and coronary artery diseases [CAD]), and behavioral and dietary habits were significantly affected by parental AM ($\theta_{T2D} = 0.018$, standard error [SE] = 0.0025, $P = 5.2 \times 10^{-14}$, $\theta_{CAD} = 0.015$, standard error [SE] = 0.0025, $P = 2.2 \times 10^{-9}$), and this result was population specific. We also identified the shared but heterogeneous impacts of AM on an adult height between the Japanese populations and Europeans from cross-population analysis using the UK Biobank resources ($n = 337,139$, $\theta_{\text{Height in EAS}} = 0.0073$ vs $\theta_{\text{Height in EUR}} = 0.030$). Finally, the consideration of the geographical factor could provide more robust results for the Japanese population. Our study demonstrated genetic evidence of AM in the Japanese population.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3409. Genetic variants associated with serum IGF-1 levels in the Long Life Family Study

Authors:

R. Minster¹, R. A. Sherlala¹, R. K. Cvejkus¹, H. Campbell², A. L. Kuipers¹, M. K. Wojczynski³, B. Thyagarajan⁴, S. V. Ukraintseva⁵, M. F. Feitosa³, J. Mengel-From⁶, J. M. Zmuda¹; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²Washington Univ. Sch. of Med., St. Louis, MO, ³Washington Univ. in St Louis, St Louis, MO, ⁴Univ. of Minnesota, Minneapolis, MN, ⁵Duke Univ., Durham, NC, ⁶Univ. of Southern Denmark, Odense, Denmark

Abstract Body:

Insulin-like growth factor 1 (IGF-1) is the product of a single gene (IGF1) located on the long arm of chromosome 12 and a main ligand in the IGF pathway. The IGF pathway may provide insight into the underlying physiology involving lifespan and healthy aging. Because the IGF pathway is associated with longevity both in human and animal models, performing genome-wide analysis on longevous families may provide additional insights about the biology behind healthy aging. Using data of more than 4,000 participants (aged 24-110 years) from 539 longevous two-generation families, we estimated the heritability of serum IGF-1 to be 0.41 ± 0.04 ($p < 0.0001$), and the proportion of variance attributed to the covariates to be 0.19. Linkage analysis revealed a novel locus associated with IGF-1 levels on 11p14 (peak LOD = 3.48). We also identified a genome-wide significant sequence variant in a discovery association analysis (rs72696993 on 14q21, $p = 4.16 \times 10^{-8}$). Replication analysis of the 212 sequence variants from the discovery analysis with $p < 1 \times 10^{-6}$ in 2,833 participants of the Framingham Heart Study did not replicate rs72696993. However, we confirmed four sequence variants at suggested significance in the discovery located within a previously known IGF-1 locus on 7p12.3. These sequence variants—rs700750, rs700752, rs700753, and rs856582 (all $p \leq 0.00023$) are located in introns 2 and 3 of lincRNA *AC011294.1* and are approximately 800 kb upstream of *IGFBP3*. This locus has been associated with other aging-related traits, such as thyroid-stimulating hormone levels, obesity, and kidney function. Additional genetic studies are required to elucidate the role that IGF-1 plays in age-related morbidity and mortality.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3410. Genetically predicted counts of white blood cells are associated with lower risk of obesity: a Mendelian randomization study

Authors:

Y. Sun, K. Ye; Univ. of Georgia, Athens, GA

Abstract Body:

The prevalence of obesity has increased remarkably worldwide and has been linked to the worsening epidemics of other metabolic diseases. Although white blood cells (WBC) have been shown to be associated with obesity, the causality remains elusive. This study aimed to investigate the causal association between WBC and obesity using bidirectional Mendelian randomization (MR) analyses. Genetic instruments of 20 WBC traits were selected from three genome-wide association studies (sample sizes ranging from 169,219 to 562,243 European individuals). We also leveraged genetic association data on 26 body mass index (BMI)-related traits, including obesity ($n = 50,364$ to $98,697$), overweight ($n = 158,855$), type 2 diabetes ($n = 655,666$), fat mass ($n = 330,762$ to $454,846$), fat-free mass ($n = 331,030$ to $454,850$), BMI ($n = 681,275$), and other related traits ($n = 10,255$ to $232,101$). In the forward MR analysis, one standard deviation (SD) genetic increase in WBC was associated with lower risks of class I obesity (odds ratio (OR): 0.85; 95% confidence interval (CI): 0.78, 0.93; $P = 5.04 \times 10^{-4}$), class II obesity (OR: 0.71; 95% CI: 0.63, 0.81; $P = 2.36 \times 10^{-7}$), class III obesity (OR: 0.63; 95% CI: 0.51, 0.78; $P = 3.44 \times 10^{-5}$), overweight (OR: 0.91; 95% CI: 0.85, 0.97; $P = 3.34 \times 10^{-3}$), and other BMI-related traits. The reverse MR analysis provided strong evidence in support of BMI-related traits causally decreasing the levels of WBC traits. Our bidirectional MR analyses suggested that genetically predicted high WBC was associated with lower risks of BMI-related traits, and BMI-related traits were also negatively associated with the level of WBC. Future studies are warranted to clarify the shared genetic etiology between WBC and obesity and to identify pathways that may underlie both traits.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3411. Genetically-guided phenotype imputation of partial or missing traits in biobank data increases power for genetic discovery

Authors:

C. Eijbouts¹, G. McVean², L. Jostins¹; ¹Univ. of Oxford, Oxford, United Kingdom, ²Genomics plc, Oxford, United Kingdom

Abstract Body:

Biobanks typically provide genetic data on large numbers of individuals, yet diagnoses and other phenotypes of interest are often only available for a subset of participants, or are inconsistent across diagnostic modalities (e.g. between self-reported data and electronic healthcare records), limiting the power and generality of genome-wide association studies (GWAS) in these datasets. We have developed a method to impute traits for which well-powered GWAS cannot be conducted because they are only available in a small subset of samples or because they are completely absent. We use previously-established genetic associations alongside the broad range of phenotypic data available in biobanks to impute and increase GWAS power for such traits. Concretely, we show how genetic associations learned from a small reference sample (e.g. a subset of a biobank, or an external dataset) can inform how relevant phenotypic features available in a larger sample (e.g. comorbidities, medication use) may be used to construct an imputed phenotype which yields increased GWAS power. After demonstrating this concept in simulations, we provide proof-of-concept by using genetic associations for systolic blood pressure learned from 15,433 UK Biobank participants, of which only one was genome-wide significant, to impute the trait in a broader sample of 179,962 participants, recovering 130 independent genome-wide significant loci, of which 55 are supported by previous literature. We then impute transferrin saturation as an example of a trait that is absent in UK Biobank, and find 11 novel associations that have not previously been observed at genome-wide significance, but which are supported by evidence of association at nominal significance in the smaller reference sample. These loci implicate genes that have previously been associated with erythroid traits (*HBSIL*), alcohol consumption (*ALDH1A1*), and risk of liver cirrhosis (*MARCI*). We propose that phenotype imputation guided by genetic associations will enable well-powered analyses of phenotypes that are partially or completely missing from biobanks, increasing the yield of future GWAS and enabling phenotype harmonisation between biobanks.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3412. Genome wide association study of asthma in 1587 French Canadian subjects.

Authors:

A. Eslami^{1,2}, Z. Li², N. Gaudreault², S. Thériault^{1,2}, Y. Bossé^{1,2}; ¹Laval Univ., Quebec, QC, Canada, ²Inst. universitaire de cardiologie et de pneumologie de Québec - Université Laval, Quebec, QC, Canada

Abstract Body:

Context: Asthma is a common respiratory disease with both genetic and environmental risk factors. Heritability estimates for asthma range from 0.55 to 0.90. Numerous genome-wide association studies (GWASs) have been completed and have identified several single nucleotide polymorphisms (SNPs) associated with asthma. **Aims:** Our research aims to perform a GWAS in the Quebec City Case-Control Asthma Cohort (QCCAC) which consists of 1,587 French-Canadian subjects (1,056 asthmatics and 531 healthy controls). We also create a genome-wide Polygenic Risk Score (PRS) to estimate an individual's genetic predisposition for asthma in our cohort. **Methodology:** The genetic association analysis was performed using SAIGE (Scalable and Accurate Implementation of GEneralized mixed model) adjusting for age, sex, and the first 20 ancestry-based principal components. We used summary statistics reported by the Trans-National Asthma Genetic Consortium meta-analysis (23,948 cases and 118,538 controls, 1,991,789 SNPs) to calculate the PRS in our cohort by applying the LDpred2 method, a Bayesian shrinkage approach. The prediction accuracy of models was evaluated by the adjusted area under the receiver operating characteristics curve (AUC). **Results:** Seven distinct genetic loci were identified (a suggestive association threshold of P value $< 1 \times 10^{-6}$) in the GWAS analysis, including asthma-associated SNPs located in *KCNQ5* and *MYL10*. The model with only sex, age, and the first 20 principal components showed an AUC of 0.588 [95% CI: (0.533-0.643)]. After adding PRS to the model the AUC increased to 0.610 [95% CI: (0.556-0.665)]. **Conclusion:** These results suggest that individuals' genetic background may improve asthma risk prediction.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3413. Genomes, exomes, and imputation: Comparing technologies for variant association and discovery in large-scale genetic studies

Authors:

T. Joseph¹, S. Gaynor¹, J. Backman², J. Emberson³, R. Collins³, J. Torres⁴, P. Kuri-Morales⁵, R. Tapia-Conyer⁵, J. Alegre⁵, J. Berumen⁶, Regeneron Genetics Center, J. Marchini¹, T. Thornton¹, G. Abecasis⁷; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Regeneron, Tarrytown, NY, ³Univ. of Oxford, Oxford, United Kingdom, ⁴The Univ. of Oxford, Oxford, Oxon, United Kingdom, United Kingdom, ⁵Natl. Autonomous Univ. of Mexico, Mexico City, Mexico, ⁶Univ. Natl. Autónoma de México, México, Mexico, ⁷Regeneron Pharmaceuticals, Tarrytown, NY

Abstract Body:

Multiple strategies are used to capture genetic variation towards the discovery of genetic factors influencing traits, including imputation from common variants, whole exome sequencing (WES), and whole genome sequencing (WGS). These approaches differ significantly in the technology used, genomic regions targeted, accuracy in detecting rare variants, and per sample cost. We perform the largest systematic comparison of the relative benefits of each approach to date, with a focus on ability to capture variants and detect association signals in coding regions. Our analysis compares each approach in two cohorts: 141,422 individuals of European ancestry in the UK Biobank (UKB), and 9,950 individuals from the Mexico City Prospective Study (MCPS). We apply standardized imputation and filtering approaches, survey genetic variants captured by each technology as characterized by consequence, frequency, and exclusivity to approach, and perform association analyses using both single-variant and gene-based tests. We survey genetic variants discovered by WGS and WES in coding regions. We find differences in variant detection to be highly sensitive to target definition and caller. When applying uniform variant calling and filtering, we find that WGS identifies 2.3% more coding variation using variants annotated with the canonical transcript --- smaller than previous comparisons. Of all coding variants captured by both technologies, 93.2% are observed in both WGS and WES, with 4.5% and 2.3% of variants observed only in WGS or WES, respectively. When variants are annotated by the most deleterious consequence, WGS identifies 4.6% more coding variation, including over 5% specific to WGS. We impute genome-wide variants in 141,422 samples of European ancestry in the UKB using the TOPMed reference panel, and compare association signals to those identified using WGS and WES datasets constructed from the same samples. We identify over 5% more GWAS peaks using the WGS dataset compared to the imputed array and WES datasets. However, when we restrict to independent association signals, defined as GWAS peaks that are greater than 1MB apart, we find only a 2.6% increase for the WGS dataset. Taken altogether, our results suggest there is a large overlap between insights gained from WGS, WES, and imputed array datasets.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3414. Genome-wide analysis of dental caries variability identifies genotype-by-environment interactions

Authors:

T. Zou¹, B. Foxman², D. McNeil³, S. M. Weinberg^{1,4}, M. L. Marazita^{1,4}, J. R. Shaffer^{1,4}; ¹Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, ²Dept. of Epidemiology, Univ. of Michigan Med. Sch., Ann Arbor, MI, ³Dept. of Community Dentistry and Behavioral Sci., Sch. of Dentistry, Univ. of Florida, Gainesville, FL, ⁴Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, PA

Abstract Body:

Background: Genotype-by-environment interactions (GEI) are hypothesized to play a role in the etiology of dental caries (i.e., tooth decay), although their effects are difficult to detect. Genetic variants associated with trait variance are prime candidates for GEI, because trait heteroskedasticity across genotype groups can serve as an indicator of a potential underlying interaction effect. The aim of this study was to investigate GEI effects on dental caries by focusing on a set of variants prioritized using genome-wide variance quantitative trait locus (vQTL) analysis.

Methods: We used Levene's test to perform separate genome-wide vQTL scans of ~5 million variants for a quantitative dental caries experience phenotype (i.e., decayed and filled tooth surfaces) in three cohorts (IFS [n=396], COHRA1 [n=328], and COHRA2 [n=773]) and combined results via meta-analysis. All participants were unrelated children of European ancestry at approximately age 5. We prioritized vQTLs with $P < 1E-6$ for GEI testing. Linear regressions were used to test GEI effects on dental caries with prioritized variants and self-reported environmental factors (e.g., demographic, socioeconomic, behavioral and dietary factors) in the three cohorts separately.

Results: A total of 39 independent vQTLs with $P < 1E-6$ were identified across the three cohorts and the meta-analysis. Some of these vQTLs were located in or near genes with plausible biological roles in dental caries (e.g., *IGFBP7*, *SLC5A8*, and *SHH* involved in tooth development and enamel mineralization). In the GEI analysis, we found that children with certain genotypes of prioritized variants exhibited higher caries experience if they had lower parental educational attainment, had lower household/parental income, brushed their teeth less frequently, consumed sugar-sweetened beverages more frequently, were not breastfed, and were female.

Conclusions: We report the first genome-wide vQTL analysis of dental caries and nominate several novel genes and GEI involved in dental caries for further investigations. Our findings expand the current understanding of the genetic architecture of dental caries.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3415. Genome-wide association studies of brain imaging endophenotypes derived from unsupervised learning identifies genes relevant to brain structure

Authors:

Z. Xie¹, K. Patel¹, H. Yuan², S. Islam¹, W. Zhang¹, M. Fornage³, S. Ji², D. Zhi⁴; ¹Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ²Texas A&M, College Station, TX, ³Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, ⁴U Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract Body:

Understanding the genetic architecture of brain structure is difficult, partly due to lack of descriptors of brain morphology. Existing methods using expert-defined image-derived phenotypes (IDPs) can be incomplete and biased. Here we present an approach of deriving brain imaging phenotypes using unsupervised deep learning. A 3-D convolutional autoencoder consisting of an encoder and decoder is trained on 6,130 UK Biobank participants' T1 and T2 FLAIR brain MRI by foreground mean square error reconstruction loss. Encoder reduces the whole brain MRI that was linearly registered to the MNI152 (Montreal Neurological Institute) space (182x218x182) into 128 dimensional latent space (ENDOs) for both T1 and T2. ENDOs capture all relevant structural variations necessary for the decoder to reconstruct the whole brain MRI. These ENDOs were studied extensively using a wide number of exploratory and statistical approaches and showed strong association with standard image derived phenotypes such as cortical and subcortical structure volumes. We also used decoder perturbation and statistical tests to map the ENDOs back to the brain regions and found meaningful interpretations. Genome-wide association studies of these ENDOs in the held-out UK Biobank white British participants with brain imaging (n=22,962 discovery and n=12,848/11,717 replication cohort for T1/T2) identified 1,132 significant ($P < 5 * 10^{-8} / 256$) SNP-ENDO pairs, out of which 658 are replicated ($P < 0.05 / 1132$). These replicated associations involve 154 SNPs organized into 43 independent loci. Almost all genes identified such as *ALDH1A2*, *EFNA1*, *2&3*, *AQP9*, *C16orf95*, *CPED1*, *MOV10*, *WNT2B*, *CAPZA1*, *NUAK1*, *RHOC*, *FAM3C*, *DAAMI*, *ST7L*, *UQCC1* were previously associated with brain structure, development and defects, and neuropsychiatric disorders. Interestingly, 40 of these loci were previously associated with imaging-derived phenotype brain morphology GWAS (including GWAS-Catalog and Oxford Brain Imaging Genetics Server where GWAS results of 3,144 imaging-derived phenotypes are hosted), 3 are new loci (rs1149935, rs11021216 and rs1791316) that are not reported in the previous IDP-based GWAS. These results established our unsupervised learning methodology as an effective approach for deriving interpretable endophenotypes for brain imaging data.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3416*. Genome-wide association studies of metabolic traits in Samoans

Authors:

J. Wehr¹, J. C. Carlson¹, E. M. Russell¹, M. Krishnan², S. Liu¹, H. Cheng³, T. Naseri⁴, M. S. Reupena⁵, S. Vialia⁶, J. Tuitele⁷, E. Kershaw¹, R. Deka³, N. L. Hawley⁸, S. T. McGarvey⁹, D. E. Weeks¹, R. L. Minster¹; ¹Univ. of Pittsburgh, Pittsburgh, PA, ²Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ³Univ. of Cincinnati, Cincinnati, OH, ⁴Ministry of Hlth., Apia, Samoa, ⁵Lutia i Puava 'ae Mapu i Fagalele, Apia, Samoa, ⁶Natl. Univ. of Samoa, Apia, Samoa, ⁷Dept. of Publ. Hlth.LBJ Tropical Med. Ctr., Faga'alu, American Samoa, ⁸Yale Univ., New Haven, CT, ⁹Brown Univ., Providence, RI

Abstract Body:

Background: Type 2 diabetes (T2D) is a major public health concern for the nation of Samoa in the South Pacific. There have been no genome-wide studies of T2D or associated phenotypes in individuals of Polynesian ancestry, whose population history might result in allele frequencies allowing for the observation of associations undetectable in other populations. Here, we performed genome-wide association studies (GWAS) of six metabolic traits: T2D, fasting glucose, fasting insulin, and HOMA-IR, as well as T2D among individuals with obesity and T2D among individuals without obesity.

Methods: Genotypes were measured from two sources. First 659,492 variants were genotyped on an Affymetrix 6.0 array in 2,890 Samoan individuals recruited in 2010. We then imputed an additional 9 million variants using a Samoan-specific haplotype reference panel derived from 1,285 Samoan individuals whole-genome sequenced by the TOPMed Program. We performed association testing using linear or logistic mixed models adjusting for fixed effects of age, sex, and principal components of ancestry derived through PC-AiR and for random effects of kinship derived from PC-Relate.

Results: There were seven unique loci associated with metabolic phenotypes at $p < 5 \times 10^{-8}$, and an additional seventeen unique loci were associated at $p < 1 \times 10^{-6}$. The most significant signal is in intron 2 of *PARD3B*: an association between rs76755625 and T2D among individuals without obesity ($p = 9.42 \times 10^{-11}$). This locus has been associated with T2D in an earlier multiancestry GWAS and has been associated body mass index in several GWASs. Notable among the other associating loci, variants in *PPARGCIA* were associated with T2D (peak $p = 4.22 \times 10^{-9}$). This locus has not been associated with T2D in genome-wide association studies to date. It encodes PPAR γ coactivator 1 α , which is a transcriptional coactivator for energy metabolism genes and regulates liver gluconeogenesis. *Ppargcla*-knockout mice exhibit disrupted adipose morphology and abnormal glucose homeostasis.

Conclusion: We observed several known and novel genetic loci associated with metabolic phenotypes in Samoans, suggesting that while some of the genetic architecture of metabolic phenotypes is shared across ancestries, there may be unique associations among Polynesians. We are currently investigating potential associations between these loci and metabolic phenotypes in an independent cohort of Samoan adults. Additional studies will be necessary to validate these associations and the determine the biological underpinnings of these associations, which may point to previously unknown biological mechanisms or social determinants of metabolic health.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3417. Genome-wide association study of obstructive sleep apnea in the Million Veteran Program uncovers heterogeneity by sex and genetic ancestry

Authors:

T. Sofer¹, N. Kurniansyah², M. Murray³, Y-L. Ho³, J. Huffman⁴, K. Cho⁵, P. Wilson⁶, D. Gottlieb⁷; ¹Harvard Med. Sch., Boston, MA, ²Brigham and Women's Hosp., Boston, MA, ³VA Boston Hlth.care System, Boston, MA, ⁴VA Boston Hlth.care System, Jamaica Plain, MA, ⁵VA Boston Hlth.care System, Boston, MA, ⁶Atlanta VA Hlth.care System, Decatur, GA, ⁷Harvard Med. Sch./Brigham and Women's Hosp., Boston, MA

Abstract Body:

Introduction: Obstructive sleep apnea (OSA) is a common condition characterized by repeated collapse of the upper airway during sleep, resulting in decreased ventilation, brief arousals, and intermittent hypoxemia. There are known differences in OSA prevalence between men and women and among race/ethnic groups. Previous studies of OSA genetics have been limited by small sample size and low prevalence of diagnosed OSA, leading to misclassification of OSA cases. **Methods:** We performed a GWAS of OSA in over 550,000 unrelated individuals from the VA Million Veteran Program. Analysis was stratified by sex and race/ethnicity-based groups, followed by meta-analysis. OSA was defined based on electronic health records via an unsupervised clustering algorithm utilizing counts of ICD codes and natural language processing-derived concepts, which estimates a probability of disease and a threshold for binary classification. Decision rules for OSA classification were validated by manual chart review of 100 individuals with at least one OSA ICD code. Association analyses were adjusted for age and 10 first principal components of genetic data computed over all MVP individuals, without and with adjustment for linear and squared terms of BMI. We estimated the heritability of OSA using LD Score regression, with LD scores computed from the analytic dataset. We tested identified OSA associations in GWAS of OSA in the FinnGen and the Mass General Brigham (MGB) biobanks. **Results:** There were 4 HARE groups in the analysis, ordered by sample size: White, Black, Hispanic, and Asian. The sample was 90% male. The prevalence of OSA in men ranged from 25.6% (Hispanic) to 21.3% (White). Among women, it ranged from 16.6% (Black) to 11.9% (Asian). The estimated heritability of OSA was higher by 5-10% in BMI-adjusted compared to BMI-unadjusted analysis, and was substantially higher in race/ethnic group-specific analysis (9-21%) compared to multi-ethnic analysis (~7%). We identified 18 independent common SNPs associated with OSA in multi-ethnic BMI-unadjusted analysis, of which 3 remained associated with OSA in BMI-adjusted analysis. Three additional loci were associated with OSA in BMI-adjusted analysis only. Twelve of the OSA SNPs in BMI-unadjusted analysis and four of the SNPs from BMI-adjusted analysis showed evidence of replication in FinnGen and MGB Biobank (one-sided p-value guided by the direction of association in MVP <0.05). Loci associated with OSA in BMI-adjusted analysis were previously reported as associated with waist-to-hip ratio (adjusted for BMI) and height, supporting the importance of anthropometric measures in OSA beyond obesity.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3418. Genome-wide association study of the human metabolome in diverse ancestries identifies an association of the *OPLAH* locus with 5-oxoproline (pyroglutamic acid) in individuals with African ancestry.

Authors:

M. Krishnan¹, A. Howard¹, H. M. Highland¹, D. Lloyd-Jones², B. Rushing¹, S. Sumner¹, K. E. North¹, P. Gordon-Larsen¹, C. L. Avery¹, M. Graff¹; ¹Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ²Northwestern Univ., Chicago, IL

Abstract Body:

Metabolite genome-wide association studies (GWAS) have identified hundreds of variants, several of which also associate with cardiometabolic diseases. However, the majority of metabolite GWAS's have been published in older adult European ancestral population, limiting inference to a fraction of the world's population. Here we report the results of a GWAS of 7,522 metabolite levels in 714 individuals of African and 1,469 individuals of European ancestry from the Coronary Artery Risk Development in Young Adults (CARDIA) study (total n=2,183; mean age = 43.56, 56.1% female). Metabolomic profiling was performed in CARDIA (n=2,183) using an Ultra performance liquid chromatography-mass spectrometer (UPLC-MS) Orbitrap method. Ancestry-specific GWAS and ancestry-combined (via inverse-variance meta-analysis) GWAS of 8,534,915 (African-ancestry) and 5,886,255 (European-ancestry) TOPMed imputed genetic variants (minor allele frequency (MAF) >0.05) were conducted in CARDIA for 7,522 metabolic peaks adjusting for age, sex and the first five ancestral principal components (PC). Significant loci were identified based on a Bonferroni adjusted *P*-value of 6.6×10^{-13} . Metaboanalyst 5.0 was used to predict functional pathway activity of significant metabolomic peaks based on mass to charge ratios (m/z) using the *mummichog* algorithm. A total of 76, 315 and 486 metabolomic peaks contained at least one significant locus in the African, European, and combined ancestry populations, respectively. In stratified analyses, 15 of the 76 metabolomic identified peaks were unique to African ancestry participants. Eight of these metabolomic peaks contained the 5-Oxoprolinase, ATP-Hydrolysing (*OPLAH*) locus, a gene that catalyzes the cleavage of 5-oxo-L-proline to form L-glutamate. *OPLAH* lead variant rs3935209 (p.Ser284Arg) was common in African-ancestry populations (minor allele frequency [MAF] = 45%), but less frequent among other ancestral groups in the Genome Aggregation Database (MAF<10%). Mummichog pathway analysis suggested that the identified metabolomic peaks participate in amino-acid-related pathways, specifically D-glutamine and D-glutamate metabolism. Two of these metabolomic peaks were annotated as 5-oxoproline (pyroglutamic acid) supporting the association of *OPLAH* with key cardiometabolic pathways. This study highlights the use of multiethnic and integrative omics approaches to broaden understanding of the genetic architecture underlying metabolites. Further work is needed to refine understanding of mapped loci and characterize identified loci in relation to cardiometabolic diseases.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3419. Genome-wide classification of epigenetic signal reveals regions of enriched heritability in complex immune traits

Authors:

M. Stricker¹, W. Zhang², W-Y. Cheng³, S. Gazal^{4,5}, C. Dendrou², S. Nahkuri⁶, P. Palamara^{1,2}; ¹Statistics Dept., Univ. of Oxford, Oxford, United Kingdom, ²Wellcome Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, ³Roche Pharma Res. & Early Dev. Informatics, New York, NY, ⁴Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ⁵Ctr. for Genetic Epidemiology, Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ⁶Roche Pharma Res. & Early Dev. Informatics, Zurich, Switzerland

Abstract Body:

Epigenetics plays a key role in regulating the expression of immune system (IS) relevant genes that dictate the therapeutic options for many diseases, but the link between high-dimensional epigenetic data and IS traits has not been fully explored.

We developed a supervised classification algorithm, called EpiNN, that leverages genome-wide epigenetic data to detect IS-relevant regions. We used a low-dimensional encoding of data produced by the ChromHMM model (Ernst & Kellis 2012) together with a handcrafted list of 888 IS loci to train a convolutional neural network, which obtained a testing AUPRC of 0.93 (SE 0.03). We used the trained model to scan the human genome for new IS regions and built a genome-wide annotation of predicted IS relevance, detecting 2765 loci with high (>0.5) EpiNN score. Of these, 1571 did not have an associated Gene Ontology term, while the remaining 1964 were enriched for IS-related function ($p = 7e-21$). To further validate these loci we performed PCR and Western Blot analyses for a constitutively expressed gene, obtaining evidence for the presence of a novel IS-specific transcription start site.

Next, we used LD score regression (S-LDSC, Finucane et al. 2015) coupled with association summary statistics for 176 traits (average $N = 262k$) to evaluate EpiNN's efficacy in detecting genomic regions enriched for heritability in IS traits. EpiNN's genome-wide IS annotation resulted in a significant enrichment and heritability effect size ($|\tau^*| p < 0.05/176$) for 20/26 IS traits (including e.g. Crohn's disease), remaining significant after conditioning on 97 other functional and evolutionary annotations. We performed a meta-analysis of 63 independent traits to confirm the specificity of the detected enrichments for IS phenotypes. IS-related traits were 4.45x (SE 0.09) more enriched for heritability ($p = 9e-92$) than non-IS traits. EpiNN effect size remained significant after further conditioning on annotations for IS-specific histone marks, ChromHMM, and transcription-based data used during training. Finally, we used S-LDXR (Shi et al. 2021) to test for depletion of squared trans-ancestry genetic correlation in European and East Asian association summary statistics for 15 IS traits (average $N_{EA} = 83k$, $N_{EU} = 235k$). The EpiNN annotation was the second most significantly depleted among 63 tested annotations, indicating ancestry-specific effects that may be driven by gene-environment interactions at IS loci impacted by recent adaptation.

These results underscore the promise of leveraging supervised learning algorithms and large epigenetic datasets to detect genomic regions implicated in specific classes of heritable traits and diseases.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3420. Genome-wide imputed differential expression enrichment (GIDEE) analysis identifies trait-relevant tissues

Authors:

D. Nyholt, A. Ghaffar, The International Headache Genetics Consortium; Queensland Univ. of Technology, Brisbane, Australia

Abstract Body:

The identification of pathogenically-relevant tissues for complex traits can be a difficult task. We have developed a novel pipeline named genome-wide imputed differential expression enrichment (GIDEE), to prioritise trait-relevant tissues by combining genome-wide association study (GWAS) summary statistic data with tissue-specific expression quantitative trait loci (eQTL) data from 49 GTEx v8 tissues. This method can be viewed as an extension of transcriptome-wide association studies (TWAS).

The GIDEE pipeline was applied to 29 GWAS datasets comprising nine ‘training’ datasets and 20 ‘discovery’ datasets. The causal tissues or cell types were uncertain or unknown for the discovery dataset traits, whereas the involvement of specific tissues in the pathogenicity of the training dataset traits had been established and reported in the literature. Therefore, the performance of four enrichment tests and their combinations was benchmarked utilising the GWAS training datasets by comparing the ranks for the known pathogenic tissue.

The best performing enrichment test produced an average rank of 1.55 out of 49 for the known pathogenic tissue across the nine GWAS training datasets. We subsequently applied the best performing GIDEE enrichment test to the 20 GWAS discovery datasets to prioritise their likely pathogenic tissues. GIDEE prioritisation may also help identify suitable proxy tissue/cell models (e.g., using enriched tissues/cells that are more easily accessible). The application of our GIDEE approach to GWAS datasets will facilitate follow-up *in silico* and *in vitro* research to determine the functional consequence of their risk loci.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3421*. Genome-wide study on 72,298 Korean individuals in Korean biobank data for 76 traits identifies hundreds of novel loci.

Authors:

K. Nam, J. Kim, S. Lee; Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract Body:

Population-based biobanks facilitate large-scale genome-wide association studies (GWAS) across numerous phenotypes. These extensive resources helped elucidate genetic components of complex traits and identify individuals with a high risk of disease. However, GWAS on diverse ancestry groups are lacking, resulting in deficits in genetic discoveries and polygenic scores. We report a GWAS of 76 phenotypes on 72,298 Korean individuals from the Korean Genome and Epidemiology Study (KoGES), a large biobank conducted by the National Biobank of Korea. Our analysis discovered 2,237 associated loci, including 117 novel associations, many of which were replicated in Biobank Japan (BBJ) at a nominal p-value. Many of the novel loci had very low minor allele frequency among European, demonstrating power increment by utilizing samples from diverse ancestry groups. To fully use the information in the KoGES data, we applied several up-to-date methods for genetic association tests such as survival GWAS and models to incorporate family disease history, discovering additional associations that are not identified in simple case-control GWAS. We also investigated pleiotropy and found that two neighboring genes in chromosome 12, *ERP29* and *NAA25* were the most pleiotropic genes with 28 associated phenotypes. To find East Asian-specific genetic associations, we conducted meta-analyses for 32 phenotypes using KoGES and BBJ GWAS results. We identified 379 novel associations and demonstrated the improved predictive performance of polygenic risk scores (PRS) by using the meta-analysis results. We believe that our results will contribute to elucidating the genetic architecture of complex traits by providing East Asian GWAS on many phenotypes. All the analysis results are publicly available at <https://koges.leelabsg.org>.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3423. GrafPop: A tool to quickly infer subject ancestry from multiple large genotype data sets.

Authors:

Y. Jin, M. Feolo, L. Phan, B. L. Kattman, J. R. Brister; Natl. Ctr. for Biotechnology Information, Natl. Library of Med., NIH, Bethesda, MD

Abstract Body:

GRAF-pop is a unique model-free algorithm to infer subject ancestry from genotypes (Y. Jin et al., 2019). Instead of using dimension reduction methods such as principal component analysis (PCA), it compares sample genotypes across a preselected set of SNPs to the allele frequencies of several reference populations with known ancestry backgrounds, then transforms the statistical scores and eventually project them onto a 2D plane to cluster the subjects. The algorithm was first implemented in the GRAF-pop feature of the GRAF (Genetic Relationship And Fingerprinting) software package, using the 10,000 fingerprint SNPs for ancestry inference. The GRAF-pop software tool can infer subject ancestry very quickly. The results obtained from genotype data collected using different methods and covering different SNPs are comparable and can be plotted in the same graph. GRAF-pop initial version, using only 10,000 SNPs unrelated to ancestry, can clearly cluster subjects into continental level populations (European, African, etc.). In addition, it can distinguish South Asians from East Asians and separate the geographic complex “Hispanics” population into two clusters. GRAF-pop is currently used in NCBI dbGaP curation pipeline and the Allele Frequency Aggregator (ALFA) project to provide population allele frequency from millions of dbGaP subjects (<https://www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/>).

In this presentation I will present an improved implementation of the GRAF-pop algorithm in a new tool GrafPop that uses 10x more markers (100K) to provide more precise clustering. GrafPop can cluster subjects with much finer granularities, for example, Central, Southern and Northern Europeans can now be separated into different clusters. GrafPop tolerates some errors and missing values in the data and hence preprocessing (e.g., filtering, imputing, deduplicating) is unnecessary. The tool quickly estimates the ancestry components for every subject in the data sets and returns results in tables and displays them in scatter plots for convenience of visual inspection.

Addition improvements include 1) support for common genotype input formats including PLINK binary data sets and large, zipped VCF files generated by whole genome sequencing; 2) SNPs are not required to be labeled with dbSNP rs IDs.

GrafPop is released as a stand-alone software, which is independent of the GRAF software package, and is freely downloadable from the following webpage: <https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/Software.cgi>.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3424*. GWAS meta-analysis of DXA-derived bone mineral density identifies 42 novel loci and uncovers biological pathways not previously identified in larger studies of ultrasound-derived bone mineral density.

Authors:

M. Frysz^{1,2}, T. Lu³, H. Lu⁴, B. H. Mullin⁵, F. Rivadeneira⁶, J. Tobias^{1,2}, B. B. Richards³, C. Ohlsson⁷, D. Kiel⁸, C. Medina-Gomez⁴, J. P. Kemp^{9,2}, on behalf of the GEFOS/GENOMOS/CHARGE; ¹Univ. of Bristol, Bristol, United Kingdom, ²MRC IEU at the Univ. of Bristol, Bristol, United Kingdom, ³McGill Univ., Montreal, QC, Canada, ⁴Erasmus MC, Rotterdam, Netherlands, ⁵The Univ. of Western Australia, Perth, Australia, ⁶Erasmus MC Rotterdam, Rotterdam, Netherlands, ⁷Gothenburg Univ., Gothenburg, Sweden, ⁸Hebrew SeniorLife, Boston, MA, ⁹Inst. for Molecular BioSci., The Univ. of Queensland, St Lucia, Australia

Abstract Body:

Background: Osteoporosis (OP) is a skeletal disorder characterized by low bone mineral density (BMD), and an increased risk of fracture (FX). BMD measured by dual-energy X-ray absorptiometry (DXA) at the femoral neck (FN) and lumbar spine (LS), is used to diagnose OP. GWAS has identified >500 BMD-associated loci. However, >90% were detected in a GWAS of BMD estimated by ultrasound (eBMD), which correlates moderately with DXA BMD ($r=0.3 - 0.6$), but is not used to diagnose OP. Aim: To conduct the largest GWAS meta-analysis of FN- and LS-BMD and identify novel biological pathways that regulate DXA BMD and were not detected in larger studies of ultrasound eBMD. **Methods:** GWAS of FN and LS-BMD (adjusted for age & sex) was conducted in ~17 cohorts, and meta-analyzed using a fixed-effects method (NTotal~148,000). Conditional and joint analysis was used to identify novel DXA BMD-associated SNPs that were statistically independent of published eBMD-associated SNPs. SNP-based heritability was estimated by LD-score regression. Candidate genes involved in BMD regulation were identified using downstream analyses including (but not limited to) expression QTL analysis of bone resorbing cells, and serum protein QTL analysis. **Results:** Independent SNPs in 165 and 190 loci were associated ($P<5\times 10^{-8}$) with FN- and LS-BMD respectively. They explained 12% and 14% of FN- and LS-BMD variance and accounted for ~50% of the total estimated SNP heritability of FN- ($h^2_{SNP}=23\%$) and LS-BMD (27%). Novel BMD associated SNPs mapping to the calcium-sensing receptor gene (*CASR*) and 41 other loci were detected >1mb from published eBMD associated lead SNPs. SNPs at several novel loci, including *TTL7* and *BCAR1* were robustly associated with FN- and LS-BMD ($P<5\times 10^{-8}$), but not with eBMD ($P>0.02$). Lead SNPs at the remaining 145 FN- and 168 LS-BMD loci mapped to within 1mb of published eBMD SNPs. Unexpectedly, several were robustly associated with increased DXA BMD, and decreased eBMD (e.g., *DLEU1* & *BCAS3*). In contrast, SNPs in ~40 loci were independent of published eBMD SNPs, despite being in close proximity (e.g., *DAAMI* & *JAG2*). Protein QTL analysis implicated TNFRSF11A, an established drug target for OP, among several other proteins that were predicted to regulate BMD but have not previously been linked to skeletal function (e.g., *CFHR1*). **Conclusions:** GWAS of DXA BMD in a sample that was 3-fold smaller than ultrasound eBMD, identified 42 novel loci and implicated biological pathways that may contribute to the regulation of bone density. These findings highlight the importance of analyzing BMD measures used clinically for diagnosing OP and warrant further investigation as they may uncover new drug targets for OP.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3425. GWAS of breast cancer family history reveals evidence for indirect genetic effects on maternal cancer risk.

Authors:

Z. Sun, Y. Wu, Q. Lu; Univ. of Wisconsin-Madison, Madison, WI

Abstract Body:

Breast cancer is the most common cancer among women worldwide and is the most frequent cause of cancer-related mortality in women. To date, genome-wide association studies (GWAS) have found over 200 significant loci associated with the risk of breast cancer. Additionally, data on breast cancer family history in large biobanks, which has been shown to be an effective proxy of disease risk in GWAS, provide new opportunities to further accelerate GWAS findings. However, existing analytic approaches for GWAS of disease family history lack consideration of indirect genetic effects in families: one's genotype can have an effect on other family members' phenotypes. For example, there has been extensive evidence of maternal genetic effects (also known as "genetic nurture") on child's birth weight and cognition. Here, we expand and employ the DONUTS approach developed by our group and leverage well-powered GWAS on breast cancer and breast cancer family history, to investigate an understudied type of indirect genetic effect: how children's genotypes influence their mother's breast cancer risk. We first performed a GWAS of maternal breast cancer history using samples of European descent in UK Biobank (N = 383,034) while adjusting for age and sex. We found 13 genome-wide significant loci with an expected and substantial overlap with known GWAS loci identified in case-control GWAS. However, one locus on chromosome 22 (rs56312415) showed highly suggestive association with breast cancer family history ($p = 9.1 \times 10^{-8}$) but not in case-control analysis ($p = 5.1 \times 10^{-1}$). This variant is in the *TXNRD2* gene, which encodes a protein belonging to the pyridine nucleotide-disulfide oxidoreductase family and is known to play a role in cellular redox environment and breast cancer pathogenesis. We then applied an extension of DONUTS approach to partition direct and indirect genetic effects through jointly analyzing breast cancer family history GWAS and case-control breast cancer GWAS (meta-analysis of Breast Cancer Association Consortium GWAS and UK Biobank; N = 467,890). The *TXNRD2* locus showed a highly significant indirect effect (on maternal cancer risk; $p = 3.3 \times 10^{-8}$) while the direct effect (on self cancer risk) was not statistically significant ($p = 8.3 \times 10^{-5}$). We further performed genetic correlation analysis and identified differential genetic correlations of self/maternal breast cancer risk with depression and educational attainment. These results provide compelling evidence for indirect genetic effects on maternal breast cancer risk and suggest extra caution for integrating family history data into standard GWAS design.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3426. GWAS of pericarditis derived from a natural language processing model on self-reported free text data identifies a genome-wide significant association on chromosome *2q14.1*.

Authors:

C. German, 23andMe Research Team, N. Eriksson, S. S. Shringarpure; 23andMe, Inc., Sunnyvale, CA

Abstract Body:

Pericarditis is a rare condition involving inflammation of the pericardium. It has been reported as a rare outcome following COVID-19 infections and COVID-19 vaccinations. A main barrier to studying the genetics of rare conditions like pericarditis is gathering enough cases to have adequate power. Empowered by the large advances in deep learning, natural language processing (NLP) is an emerging field of artificial intelligence concerned with processing and understanding natural language (speech and text data). Named entity recognition models, a class of NLP models, aim to identify named entities (such as names, locations, etc.) in text data. scispaCy is an open-source python package developed by the Allen Institute for Artificial Intelligence that has various types of pre-trained named entity recognition models fine-tuned to biomedical, clinical, and scientific text. Here we illustrate the utility of this class of NLP models for genetic discovery when used in conjunction with large-scale self reported free text data and genetic data collected through the 23andMe, Inc. direct-to-consumer platform. We gathered self-reported free text data from a cohort of over 1 million genotyped research-consented individuals, where participants were asked to write out health conditions they had that had not been covered by a survey mainly focused on common diseases (prevalence >1%). We then applied a pre-trained NLP model from the scispaCy library on the free text data to identify cases of pericarditis. We used these individuals as cases to run a GWAS and found a genome-wide significant association in the 2q14.1 region (p-value = 2.0e-14) with an odds ratio of 0.733 (95% CI: [0.675, 0.795]). Our results demonstrate that using NLP models on self-reported free text survey data is a viable method for identifying genetic associations in pericarditis. This approach may be useful for other rare conditions when there is free text data available, where case counts are often the limiting factor in discovery.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3427. GWAS of serum ALT and AST using whole genome sequencing data for 119,009 UK Biobank participants yields novel associations with rare variants.

Authors:

A. Holleman, R. A. Hoffing, A. M. Deaton, L. Krohn, P. LoGerfo, P. Nioi, M. E. Plekan, C. Willis, L. D. Ward; Alnylam Pharmaceuticals, Cambridge, MA

Abstract Body:

Liver disease is a leading cause of mortality globally and has increased in incidence over the past two decades. Understanding the genetic contributions to liver disease holds great promise for guiding the development of more effective and safer therapeutics. Research groups including ours previously leveraged biobank-scale genome-wide imputation and whole exome sequencing (WES) datasets to identify numerous associations of genetic variants with both serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST), enzymes that are biomarkers of liver injury and that in combination are predictors of liver disease. In the current study, we furthered our examination of genetic contributors to ALT and AST by analyzing whole genome sequencing (WGS) data from 119,009 UK Biobank participants with European ancestry (a subset of the samples included in our previous analyses using imputed data). We used REGENIE to perform a genome-wide association study of approximately 71 million single nucleotide variants (SNVs) and indels with a minor allele count of at least 10. After applying a strict Bonferroni correction for the number variants tested, we identified 42 candidate loci associated with ALT and 58 candidate loci associated with AST. We found no novel loci that were associated with both ALT and AST (defining novel as not being reported in GWAS Catalog). However, we did uncover novel loci associated with one or the other enzyme. Of the 42 ALT-associated loci, one was novel: a rare intergenic SNV situated between *CYP11B2* and *LY6E* on chromosome 8 (Beta = -1.49; P = 4.33e-14), with the latter gene broadly expressed including in the liver. Of the 58 AST-associated loci, 3 were novel, all located on chromosome 10: a rare 3-base deletion within an intron of *RSU1* (Beta = -0.59; P = 7.83e-12); a locus spanning 170kb and including two rare SNVs, one within an intron of *CYP2C19* (Beta = 1.54; P = 5.09e-11) and the other approximately 1,100 bases upstream of *CYP2C9* (Beta = 1.54; P = 5.11e-11), both of these genes being primarily expressed in the liver; and a rare intergenic SNV located between *CCNJ* and *ZNF518A* (Beta = 1.49; P = 2.02e-10). The association of these variants specifically with ALT or AST, and not with both enzymes, suggests that they may not be informative for liver disease drug target identification. Yet insofar as these loci contribute to normal variation in ALT or AST levels, it may be useful to account for them when examining potential toxic effects of a drug on the liver. Future research plans include analyzing associations of ALT and AST with copy number variants called based on the WGS data for our sample set.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3428. Hand grip strength and its underlying genetics is associated with widespread protein changes revealed by population-based plasma proteomics.

Authors:

Y. Huang, D. Bodnar, B. Sun, C. Whelan, Biogen Biobank Team, K. Meilleur, J. Shi, H. Runz, E. Tsai; Biogen, Cambridge, MA

Abstract Body:

Hand grip strength (HGS) is a proxy measure of general muscle strength and a predictive measure of multiple adverse health outcomes including disability and mortality. Previous studies have shed light on protein changes in human muscle associated with aging and exercise, but most were limited by small sample sizes due to the rarity of muscle tissue samples. Plasma proteins such as creatine kinase have also been used as biomarkers for certain muscular atrophy and injuries, but systematic changes in blood protein concentrations have not been evaluated for general muscle strength. The UK Biobank Pharma Proteomics Project (UKBB-PPP) measured circulating concentrations of 1,463 plasma proteins in 54,306 UK Biobank participants, for whom HGS was also measured at recruitment using hydraulic hand dynamometer. We conducted a proteome-wide association analysis of HGS adjusted for age, sex, height, body weight, and BMI among 48,699 participants that passed data quality control. Out of 1,463 protein species measured in plasma, 75 positive and 928 negative associations were significant after Bonferroni correction ($p < 3.4e-5$). Strongest positive associations were observed for CA14, ENPP5 and EGFR, while top negative associations included LEP, GDF15 and CLMP. Gene-set enrichment analysis (GSEA) using Enrichr identified myoblast proliferation, TGF-beta activation, and myoblast migration as the top enriched pathways for positively associated proteins, while inflammatory pathways including cytokine production and monocyte/lymphocyte chemotaxis were most enriched for negatively associated proteins. To explore proteomic associations with HGS driven by underlying genetic susceptibility, we conducted a GWAS of maximum HGS among 327,290 UKBB participants of European ancestry that were not included in UKBB-PPP and derived a polygenic score using PRS-CS from 1.17 million variants. HGS-PRS was then calculated for 36,367 European UKBB-PPP participants to assess its proteomic associations. 51 plasma proteins showed significant associations with HGS-PRS, of which 23 were positively associated, and 43 were previously associated with HGS. Consistently, GSEA identified muscle-relevant pathways such as myoblast proliferation and calcium transport for HGS-PRS positively associated proteins and inflammatory pathways for HGS-PRS negatively associated proteins. Our analysis reveals systematic associations of HGS with the plasma proteome and the potential impacts of genetic susceptibility to general muscle strength on muscular and inflammatory pathways.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3429. Haplotype-based fine-mapping of variant associations for complex traits

Authors:

A. Massarat, M. Gymrek; Univ. of California San Diego, La Jolla, CA

Abstract Body:

Over the past decade, genome-wide association studies (GWAS) have identified thousands of genomic loci associated with a range of traits in humans. However, identifying the underlying causal variant(s) driving an association signal is challenging. A typical GWAS signal may harbor hundreds of variants in strong linkage disequilibrium (LD) with each other, most of which are not causal for the trait. Further, even when the causal variant is directly genotyped, it may not have the best p-value in the locus. A variety of methods have been developed to perform statistical fine-mapping, with the goal of inferring the candidate causal variants at a particular GWAS signal. However, existing methods face several important limitations: they typically focus on independent contributions of individual bi-allelic variants such as single nucleotide polymorphisms (SNPs) and assume the true causal variant exists within the set of candidate variants being considered at a locus. Yet increasing evidence suggests that many of the strongest GWAS signals are driven by complex variants such as repeats or structural variants that are not typically directly genotyped, are often multi-allelic, and may be in only modest LD with surrounding SNPs. Existing fine-mapping solutions will not produce reliable results for these cases. Here, we present happler, a new fine-mapping method that models the effects of SNP haplotypes (sets of alleles at one or more SNPs inherited on a single chromosome) rather than of individual SNPs. SNP haplotypes are better able to capture effects of variants that were not directly included in a GWAS dataset, such as repeats, structural variants, and rare variants. To identify trait-associated haplotypes, happler constructs binary trees of regression models composed of correlated variants as nodes, where a path from the root to a leaf identifies a haplotype that can be used as input to an existing fine-mapping method to obtain a probability of causality. We demonstrate the utility of our tool using phenotypes simulated based on causal variants not directly genotyped, and show that we can identify haplotypes harboring the true causal variant. We are currently applying our method to real GWAS data for blood and serum biomarker traits profiled by the UK Biobank. Overall, we envision that a haplotype-based approach will improve the ability to fine-map loci driven by complex variants not typically considered in GWAS. Further, we predict that improved identification of causal variants will ultimately better our understanding of the biological mechanisms driving complex traits and improve our ability to predict disease risk for individual patients based on their genotypes.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3430. Hierarchical scanning strategy reveals signal regions in 260 genes implicated in regulation of gene expression on chromosome 19q

Authors:

X. Zhang, W. Bush; Case Western Reserve Univ, Cleveland, OH

Abstract Body:

Hi-C experiments and FISH imaging have provided evidence of chromatin's dynamic 3D folding and looping structures, including hierarchical Topologically Associating Domains (TADs), within the nucleus. TAD boundaries are traditionally called directly from Hi-C data. Regions between TAD boundaries are believed to be isolated environments that constrain enhancer-promoter interactions. However, it's hard to associate the unstable TAD with gene regulation. Besides, we still lack an understanding of the function of the hierarchical TADs and the sub-loops. This study assumed that the sub-loops have a similar enrichment pattern of chromatin interaction as TADs. We further restrict the analysis to focus on the regulation effect of rare genetic variants located within transcription factor binding sites (TFBSs). Instead of calling hierarchical TADs directly from Hi-C data, we implemented scan tests on RNA-seq and WGS data to identify clusters of TFBSs associated with gene expression levels, ignoring the physical distance between the TFBS and gene along the genome. We can then use the identified clusters as a proxy to describe the hierarchical chromatin structures. We first concatenated overlapping TFBSs into single initial scanning units. Then we utilized a flexibly shaped spatial scan framework (PSCAN, Tang et al., 2020) to define scan windows while accommodating the positions of the concatenated TFBSs. For each gene expression trait, p-values from set-based rare-variant variance-effect tests were combined to evaluate the global association; non-overlapping signal regions were also detected. With such a scanning and testing strategy, we can identify tissue- and gene-specific TFBS clusters along the genome and infer the hierarchical insulated loop structures like hierarchical TADs based on these clusters. We applied this scanning strategy to chr19q using GTEx (v7) Whole Blood data (N=369). We defined 4,653 clusters of concatenated TFBSs within the scanning region and applied the afore-described scan test to 302 gene expression traits on chr19q. We identified signal regions for 260 genes, suggesting that the TFBSs residing in the signal regions have a joint effect on gene expression regulation. Further investigation into these signal regions will inform both the cis- and trans-regulation effects of transcription factor binding events. We can infer potential hierarchical structures of chromatin using this approach.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3431. HLA Analysis Tool Kit v2.0

Authors:

W. Choi¹, B. Han²; ¹Seoul Natl. Univ. Graduate Sch., 103 Daehakro, Jongro-gu, Korea, Republic of, ²Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Abstract Body:

Fine-mapping association signals arising in Human Leukocyte Antigen(HLA) region can offer important information in investigating immune-related diseases. We previously introduced HLA Analysis Tool-Kit(HATK; Bioinformatics 2021) to facilitate the fine-mapping analysis for researchers, providing (1) preprocessing of DNA and amino acid sequences from the IPD-IMGT/HLA database, (2) preparation of virtual marker data for association test with given HLA type data, and (3) visualization of association signals. Here we present HATK v2.0, which improved in several aspects. First, HATK v2.0 now can handle more non-classical HLA genes such as *HLA-MICA*, *-MICB*, *-V*, or *-E*, while the HATK only worked restricted to the 8 major HLA genes, i.e. *HLA-A*, *-B*, *-C*, *-DPA1*, *-DPB1*, *-DQA1*, *-DQB1*, and *-DRB1*. Second, HATK v2.0 provides automated phasing to its output marker panel, which can make it easier for researchers to conduct downstream haplotype analysis. Although the HATK provided the omnibus test that can perform both the association test of an amino acid locus and conditional haplotype analysis of phased amino acid data, the HATK didn't provide a phasing module. Therefore, researchers were required to phase their data by themselves. Third, researchers can now use HATK v2.0 to generate a custom reference panel that can be used in HLA imputation by CookHLA. Currently, the MakeReference module included in the SNP2HLA software is being widely used for this function given the typed HLA and SNP data. However, this module only works on the outdated version of the IMGT database and is fixed to the outdated hg18 genomic coordinate. HATK v2.0 provides the same functionality but with high flexibility where users can choose a specific IMGT database version and their preferred genomic coordinate. HATK v2.0 is freely available at <https://github.com/WansonChoi/HATK>.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3432. Hybrid autoencoder for robust ancestry inference in the presence of data artifacts and relatedness.

Authors:

M. Yuan¹, H. Hoskens¹, S. Goovaerts¹, N. Herrick², M. Shriver³, S. Walsh², P. Claes^{1,4}; ¹KU Leuven, Leuven, Belgium, ²Indiana Univ. - Purdue Univ. Indianapolis, Indianapolis, IN, ³Penn State Univ, University Park, PA, ⁴Murdoch Children's Res. Inst., Melbourne, Australia

Abstract Body:

Analysis of genetic ancestry remains an important topic in human genetics and bioinformatics. Conventional methods are typically affected by laboratory artifacts and related individuals in the dataset. Thus, complex quality control procedures are often required to ensure accurate inference of genetic ancestry. In this work, we propose a novel hybrid approach, referred to as SAE-IBS, which consists of a Singular Autoencoder (SAE) generalized by an Identity-by-State (IBS) similarity matrix, for robust ancestry inference in the presence of missing data, genotyping errors, and relatedness. SAE-IBS yields an orthogonal latent space, like matrix decomposition-based methods, and can extract more comprehensive latent features by exploiting the non-linear nature of neural network-based methods. We demonstrate that the proposed approach achieves comparable performance with principal component analysis (PCA), the most widely used method for genetic clustering and ancestry inference through classification of global population by continent. Moreover, the ancestry space obtained by SAE-IBS exhibits interesting patterns: it has a PCA-like variance-covariance structure, yet different population groups can be separated with fewer dimensions compared to PCA. This layout remained consistent under different hyperparameter settings and was independent of the dimensionality used to train the model, leading to a strong and desired stability of the population structure inference and an enhanced dimensionality selection after training. Furthermore, with the flexibility in training neural networks, we show that a robust ancestry space in the presence of relatedness can be obtained by simply changing the loss function from L2 norm to L1 norm, whereas the least-square matrix decomposition encountered in PCA leads to sensitivity to relatedness. Finally, we integrate the principle of denoising training into our model and propose an extension by imposing robust projections through the incorporation of an additional loss, i.e., Denoising SAE-IBS-L. The proposed extension reaches higher accuracy than existing methods for projecting poor quality target samples (genotyping errors and missing data) onto a reference ancestry space.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3433. HybridGWAIS-Web: A fast and secure epistasis detection web service

Authors:

L. Wienbrandt, C. Priess, D. Ellinghaus; Inst. of Clinical Molecular Biology, Kiel Univ., Kiel, Germany

Abstract Body:

Genome-wide association interaction studies (GWAIS) have become increasingly important as it is believed that genetic interactions play a significant role in genetic variation causing complex diseases. For assessing gene-gene (GxG) interactions, logistic regression as implemented in the PLINK tool is a powerful and commonly used framework. However, fitting regression models for each pair of markers in a genome-wide data set is an extremely compute intensive task.

We recently demonstrated that we are able to speedup PLINK's epistasis test by a factor of more than 1,600 on a hybrid computing architecture that combines a Field-Programmable Gate Array (FPGA) with a Graphics Processing Unit (GPU). Analysis of our exemplary case-control data set with 16,000 samples typed at 1.2 million markers takes 3:45 hours on our system, while computation with PLINK takes more than 260 days, even when run with 32 threads on a 16-core system with hyper threading.

We now introduce the HybridGWAIS-Web web service which makes our tool, along with the computing power of the hybrid combination of FPGAs and GPUs, freely available. The web service allows researchers to perform GWAIS without access to high-performance computing resources to analyze their genome-wide data sets quickly and conveniently.

Besides the multiplicative logistic regression test (as implemented in PLINK) HybridGWAIS-Web provides other statistical tests as well, such as mutual information, interaction gain, or the BOOST epistasis screening method. Third-order interaction tests between three genetic markers are also available, and linkage disequilibrium (r^2 -correlation) values between each pair of markers can optionally be calculated on-the-fly.

HybridGWAIS-Web combines easy access to our hybrid computing system, many convenient algorithmic features (such as chromosome region selection, result filtering of best results, etc.), user data protection (usage is compliant with the European General Data Protection Regulation (GDPR)), and security (through exclusively encrypted connections, password-protected data access and optional 2-factor authentication).

For now, users can upload data sets for which the predicted runtime analysis would take no longer than one day. In that time, users can run a logistic regression test on a case-control data set with e.g. up to 100,000 samples typed at 1.25 million markers. On a usual CPU-based server-grade compute node these tests would take more than 4 years using PLINK's original epistasis test.

The HybridGWAIS-Web service is freely available at <https://hybridcomputing.ikmb.uni-kiel.de>.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3434*. Identification of age-related genetic and molecular determinants and their link to disease.

Authors:

T-D. Michalettou¹, M-G. Hong², J. Fernandez³, S. Sharma⁴, C. A. Brorsson⁵, R. W. Koivula^{3,6}, J. Adamski⁴, S. Brunak⁷, E. T. Dermizakis⁸, P. W. Franks⁶, M. McCarthy⁹, E. Pearson¹⁰, J. Schwenk¹¹, M. Walker¹, DIRECT consortium, A. Brown¹⁰, A. Vinuela¹; ¹Newcastle Univ., Newcastle upon Tyne, United Kingdom, ²Natl. Bioinformatics Infrastructure Sweden, Stockholm, Sweden, ³Univ. of Oxford, Oxford, United Kingdom, ⁴HelmholtzZentrum München, Neuherberg, Germany, ⁵Technical Univ. of Denmark, Lyngby, Denmark, ⁶Lund Univ., Lund, Sweden, ⁷Copenhagen Univ., Copenhagen, Denmark, ⁸GSK, Geneva, Switzerland, ⁹Genentech, South San Francisco, CA, ¹⁰Univ. of Dundee, Dundee, United Kingdom, ¹¹KTH Royal Inst. of Technology, Stockholm, Sweden

Abstract Body:

Ageing is a complex process, involving changes in multiple molecular, cellular, tissue and organ systems, which has a great impact on disease risk for many different conditions. To identify the age-related changes in molecular phenotypes that may lead to increased disease risk with age, we evaluated the influence of age, type 2 diabetes (T2D) status, sex and BMI on whole blood and plasma transcriptomic, proteomic and metabolomic data from the DIRECT consortium (N=3,029 participants). We found 10,687 genes (66%) and 287 proteins (77%) were differentially expressed with age (FDR<0.05). We found 148 cases where both the gene and the corresponding protein were associated with age, of which 50% showed an opposite directions of effect. For example, gene *REG4* showed a decrease in expression with age (Pval=2.40e-46, β =-0.058) while the protein abundance increased as participants aged (Pval=2.62e-15, β =0.006). While this result is counter-intuitive, it is in agreement with results of other analyses we have seen, where genetic factors frequently have opposite effects on gene and protein levels. We observe the highest proportion of age associations when analysing metabolite levels. Of the targeted and un-targeted metabolites evaluated (n=349), 78% were associated with age (FDR<0.05), including L-Carnitine (Pval=5.91e-17, β =0.020) and DHEA-S (Pval=4.27e-86, β =-0.056). We additionally evaluated the independent effects of sex, BMI and T2D status on these molecular phenotypes. We found 10191 genes associated with sex, the majority of which overlapped with genes associated with age. T2D status was associated with the largest number of proteins (308 associations) and sex and BMI were associated with similar numbers of metabolites (335 and 305 associations respectively). We found 8 proteins associated only with T2D after controlling for BMI, age and sex, including TSHB, GP6 and SLAMF7. We also find examples of sex-specific effects of age. For example levels of the SOST protein increase with age in men, but decrease in women. Most measured molecular phenotypes are significantly associated with age, however the effect of age frequently differed across related phenotypes. By identifying age-related changes in molecular phenotypes we are one step closer to uncovering the link between ageing and disease, either because these phenotypes are the mediating actors, or because these phenotypes interact with the genetics of the relevant disease.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3435. Identification of key biomarkers associated with obesity using multi-omics data integration

Authors:

A. Liu¹, L. Jiang¹, K-J. Su¹, X. Zhang¹, Y. Gong¹, c. qiu¹, J. Greenbaum¹, Z. Luo¹, Q. Tian¹, Z. Ding², H. Shen¹, H-W. Deng¹; ¹Tulane Univ. Sch. of Med., New Orleans, LA, ²Tulane Univ. Dept. of Computer Sci., New Orleans, LA

Abstract Body:

Obesity has emerged as a major global public health problem in the 21st century, spreading rapidly across both developed and developing countries and disproportionately affecting the most vulnerable and socially disadvantaged populations. Omics-based technologies have been put forward to provide a better understanding of obesity etiology. While each omics technology is capable of depicting a portion of the biological information, integrating multiple types of omics data can provide a more comprehensive picture of the underlying biological processes. Specifically, for obesity, existing research has shown that incorporating data from multiple omics technologies can improve the performance of patient clinical categorization in comparison with only using one single type of omics data. Therefore, to effectively take advantage of the interactions and complementary information in multi-omics data, the application of integrative analysis methods is very necessary. Here, we will apply a multi-omics integrative method using multi-view graph attention networks for binary and multi-class biomedical classification by combining multi-omics data—such as single nucleotide polymorphisms (SNPs), copy number variations, mRNA gene expression, miRNA, DNA methylation and metabolomics. This method jointly explores omics-specific learning with graph attention networks and cross-omics correlation learning with view correlation discovery network (VCDN) for effective multi-omics data classification. We will apply the proposed method on obesity classification tasks in a multiethnic sample (61% Caucasian and 39% African American) from Trans-omics Integration of Multi-omics Studies for Male Osteoporosis Study (n = 919). The top 30 SNPs, copy number variations, mRNA, miRNA, DNA methylation and metabolomics features will be identified by layer-wise relevance propagation as the key important biomarkers associated with obesity. The identification of obesity biomarkers and the understanding of their biological roles would be transformed into consistent benefits for the effective prevention, intervention, and treatment for both obesity and obesity-related diseases.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3436. Identification of rare genetic coding variants associated with Alzheimer's disease.

Authors:

Z. Khurshid^{1,2}, J. Farrell², T. Tong^{1,2}, C. Zhu², E. Martin³, M. Pericak-Vance³, A. Naj⁴, A. Kuzma⁴, L. Cantwell⁴, O. Valladares⁴, L-S. Wang⁴, G. Schellenberg⁴, J. Haines⁵, K. Lunetta⁶, Y. Leung⁴, X. Zhang^{1,2,6}, L. Farrer^{1,2,6}, Alzheimer's Disease Sequencing Project (ADSP); ¹Bioinformatics Program, Boston Univ., Boston, MA, ²Dept. of Med. (BioMed. Genetics), Boston Univ. Sch. of Med., Boston, MA, ³Hussman Inst. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, ⁴Penn Neurodegeneration Genomics Ctr., Dept. of Pathology and Lab. Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, ⁵Dept. of Population and Quantitative Hlth.Sci. Case Western Reserve Univ., Cleveland, OH, ⁶Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA

Abstract Body:

Background: Alzheimer's disease (AD) is the most common cause of dementia resulting in memory loss and an impaired cognitive function. Previous whole exome sequencing (WES) studies of samples including as many as 11,000 subjects have identified genome-wide significant associations with rare variants in several novel genes, but also highlighted the need for investigations in even larger samples. We combined WES and WGS data assembled by the Alzheimer Disease Sequencing Project (ADSP) to increase power for detection of associations with rare variants in gene coding regions.

Methods: We developed an efficient computational pipeline to perform both pre-merging and post-merging genotype, variant and sample-level quality control (QC) on WGS data containing 16,905 individuals (in short, 17K WGS) and WES data for 20,504 individuals. The resultant sample included participants from European (EA, 11,279 AD cases and 8,924 controls), African American (AA, 2,757 AD cases and 4,336 controls), and Dominican (DR 1,438 AD cases and 3,256 controls) ancestries. In the total sample set (joint analysis and ethnicity stratified datasets), we tested for association of AD with 250,465 bi-allelic variants that passed our QC pipeline with minor allele count ≥ 20 , call rate $\geq 80\%$, HWE p -value $\geq 1 \times 10^{-8}$ using a logistic model implemented in GENESIS including covariates for age, sex, exome capture kit, read length, and principal components (PCs) of ancestry. The same model was also implemented without adjusting for age to account for the unique nature of the 20K WES where the selected controls were older than AD cases.

Results: Using a study-wide significant (SWS) threshold of $p = 0.05/250,465 = 2 \times 10^{-7}$ in the joint analysis, we found associations with variants from 7 independent loci. As expected, the top-ranked association was observed with a SNP in APOE (rs429358, $p = 1.12 \times 10^{-209}$) which was also SWS in EAs ($p = 4.77 \times 10^{-74}$), AAs ($p = 2.31 \times 10^{-47}$), and DRs ($p = 1.70 \times 10^{-14}$). Ethnic stratified analysis also showed SWS associations with rs367563342 in ICK ($p = 7.74 \times 10^{-11}$) in EAs and rs200944727 in CPSF1 ($p = 4.02 \times 10^{-8}$) in DRs.

Discussion: We demonstrated how merging WGS and WES datasets can help us increase genetic power and find novel rare coding variants that provide insight into AD pathogenesis. Multiple known and novel AD loci were identified but the known R47H rare variant in TREM2 that reached SWS in prior studies was not found. This could be due to sample set differences, as samples in previous study were enriched for AD cases with familial disease as this wasn't the case for the current WGS and WES samples.

Keywords: Alzheimer's disease, genome wide association study (GWAS), rare genetic coding variants

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3437. Identifying circulating proteins as biomarkers for age at menarche and age at natural menopause: insights from a Mendelian randomization study

Authors:

N. Yazdanpanah¹, M. Yazdanpanah¹, D. Manousaki²; ¹CHU Sainte-Justine, Montréal, QC, Canada, ²Sainte Justine Hosp. Univ. of Montreal, Montreal, QC, Canada

Abstract Body:

Objective: Age at menarche (AAM) and at natural menopause (ANM) in humans are highly heritable traits. In an effort to characterize novel biomarkers associated with the female reproductive lifespan, we aimed to identify circulating proteins, whose genetically predicted levels are associated with age at menarche or natural menopause, by performing a two-sample Mendelian Randomization (MR) study. *Research Design and Methods:* Utilizing *cis* genetic determinants (protein quantitative trait loci or pQTL) of up to 1379 circulating proteins from five large genome-wide association studies (GWAS), we performed a two-sample MR study, to screen for causal associations of these proteins with AAM and ANM in women from the REPROGEN GWAS (n=370,000 for AAM and n=200,000 for ANM). Further, pleiotropy-robust MR methods were used in sensitivity analyses using both *cis* and *trans*-pQTL. As follow-up analyses for the MR-prioritized proteins, we performed colocalization, pathway and enrichment analyses using expression profiles. *Results:* We identified 18 circulating proteins associated with AAM and 14 proteins associated with ANM. Our pathway enrichment analysis suggests involvement of the AAM-related proteins in peptide hormone biosynthesis in pituitary. Among these proteins, HPGDS (*cis*-pQTL: rs1965049) showed colocalization with AAM. Moreover, our pathway enrichment analysis for the 14 proteins associated with ANM demonstrates functional effects of these molecules on the immune system, while we found enriched expression of their *cis*-pQTLs in the immune system and adipose tissue. CPNE1 (*cis*-pQTL: rs12481228) ADAMTS13 (*cis*-pQTL:rs71503194) and LY9 (*cis*-pQTL:rs12128261) proteins strongly colocalized with ANM. *Conclusions:* Our findings highlight a potential role of the genes encoding the prioritized circulating proteins in the physiology of the reproductive lifespan in women, while these proteins can serve as biomarkers predicting age variations in menarche and natural menopause. Further study of the shared and specific proteomic profiles associated with ages at menarche or natural menopause may help identifying pathophysiological pathways and new targets for treatment of extreme variations in timing of puberty or menopause.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3438. Identifying common genetic susceptibility underlying comorbid phenotypes using Binomial Regression.

Authors:

S. Ghosh, P. Panja; Indian Statistical Inst., Kolkata, India

Abstract Body:

The models suggested in MultiPhen (O'Reilly et al., 2012) and BAMP (Majumdar et al, 2015) provide an alternative to study population-based genetic association with multivariate phenotypes by exploring the dependence of genotype on phenotype instead of the naturally arising dependence of phenotype on genotype. However, these tests for association are based on the null hypothesis of no association with any of the constituent traits of the multivariate phenotype vector. Thus, such tests do not provide evidence of pleiotropy. With respect to a pair of comorbid phenotypes (both binary, a combination of binary and quantitative or both quantitative), we aim to modify the proposed BAMP (Binomial regression-based Association of Multivariate Phenotypes) approach to test the null hypothesis of no association with at least one of the phenotypes versus the alternative hypothesis of association with both the phenotypes. Since the likelihood ratio test requires a constrained maximization (over the two coordinate axes) under the null hypothesis, it is analytically difficult to obtain the asymptotic distribution of the log-likelihood test statistic under the null hypothesis and hence, we use permutation procedures to determine the thresholds for rejection. We carry out extensive simulations under different genetic models and correlation structures of the bivariate phenotype to evaluate the type 1 error rates and power as well as compare the performance of the proposed test procedure to separate univariate tests with standard multiple testing corrections. We find that while combining univariate tests results in inflated probabilities of type 1 error, the proposed test maintains the appropriate size. We also show that an extension of the method to more than two comorbid phenotypes is theoretically straight forward with minimal increase in the computational burden.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3439. Identifying composite biomarkers in multi-omics datasets of osteoporosis for drug discovery

Authors:

M. Alam¹, H. Shen², H-W. Deng²; ¹Tulane Univ., New Orleans, LA, ²Tulane Univ, New Orleans, LA

Abstract Body:

Background: Many statistical machine approaches could identify novel features of the etiology of complex diseases by analyzing multi-omics data. However, they are sensitive to some deviations in distribution when the observed samples are potentially contaminated with adversarial corrupted outliers (e.g., a fictional data distribution). Likewise, statistical advances lag behind supporting comprehensive data-driven analyses of complex multi-omics data integration. **Methods:** We propose a novel non-linear M-estimator-based approach, “robust kernel machine regression (RobKMR),” to improve the robustness of statistical regression and the diversity of fictional data to examine the higher-order composite effect of multi-omics datasets. We address a robust kernel-centered Gram matrix to estimate the model parameters accurately. We also propose a robust score test to assess the marginal and joint Hadamard product of features from multi-omics data. We apply our proposed approach to a multi-omics dataset of osteoporosis (OP) from Caucasian females **Results:** Experiments demonstrate that the proposed approach effectively identifies the inter-related risk factors of OP. With solid evidence (at p-values ≤ 0.00001), biological validations, network-based analysis, causal inference, and drug repurposing, the selected three triplets ((DKK1, SMTN, DRGX), (MTND5, FASTKD2, CSMD3), (MTND5, COG3, CSMD3)) are significant biomarkers and directly relate to OP. Overall, the top three selected genes (DKK1, MTND5, FASTKD2) and one gene (SIDT1 at p-values ≤ 0.001) significantly bond with four drugs- Tacrolimus, Ibandronate, Alendronate, and Bazedoxifene out of 30 candidates for drug repurposing in OP. **Conclusion:** We developed a novel robust approach to identify inter-related risk factors in multi-omics data. Benchmarking experiments based on simulation and real datasets analysis demonstrated that our proposed approach provides a competitive performance compared with the existing ones to derive a statistic for testing the inter-related risk factors. Our proposed method can identify significant biomarkers (DKK1, MTND5, FASTKD2, and SIDT1) for OP studies. **Acknowledgments** This work is benefited by the support of U19AG05537301 and R01AR069055.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3440. Identifying functionally-related SNPs and genes from phenome-wide genetic effect correlations using GWAS summary statistics

Authors:

L. Chen, J. Tubbs, P. Sham; The Univ. of Hong Kong, Pokfulam, Hong Kong, China

Abstract Body:

We observe a ubiquitous distribution of non-trivial correlations of summary statistics (e.g. Z-scores, log P-values) across numerous genome-wide association studies (GWAS) on distal SNPs and genes, revealing the importance of investigating phenome-wide genetic-effect correlation. Analogous to genetic correlation, it is an aspect of genetic architecture that quantifies the similarity of causal effect sizes between two SNPs across the phenome, but has been relatively unexplored. Inflation of GWAS-wide correlations of summary statistics can arise due to true genetic-effect correlations or through bias by other confounders, such as linkage disequilibrium (LD) and GWAS non-independence. However, no current methods can be used to distinguish between augmentation from a true genetic-effect correlation signal and confounders. We have developed a statistical framework (LRCP) to quantify the contribution of each by examining the relationship between Z-score cross products and LD cross components. **LRCP**, using both likelihood and regression-based methods on cross components of summary statistics, is the first attempt to infer functional relationships of arbitrary SNP pairs from their genetic-effect correlation estimate. Applying LRCP to large collections of GWAS data on quantities of traits, we find some evidence for widespread genetic-effect correlations in the genome, providing a new approach to identify functionally-related SNPs, gene clusters and novel pathways.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3441. Identifying non-common variant sources of clinical comorbidities through integration of electronic health data and genetics for schizophrenia.

Authors:

T. Vessels¹, N. Strayer², K. Choi³, T. J. Morley², L. Han², H. Lee³, J. Smoller^{4,5}, Y. Xu², D. Ruderfer²; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Massachusetts Gen. Hosp./ Harvard, Boston, MA, ⁴Massachusetts Gen Hosp, Boston, MA, ⁵Broad Inst., Cambridge, MA

Abstract Body:

Background: Patients with schizophrenia (SCZ) have a reduced life expectancy of 10-20 years that is often a byproduct of comorbidity, which accounts for 70% of premature mortalities. We hypothesize that comparing comorbidities from electronic health record (EHR) data to genetic associations will provide insights into which comorbidities are effects of the SCZ genetic etiology vs consequences of treatment, behavior, or other environmental factors. **Methods:** We defined clinical comorbidities from a logistic regression of pairwise comparisons of diagnostic billing codes (PheCodes), adjusting for demographic and other covariates from two health care systems (Vanderbilt University Medical Center, VUMC; Mass General Brigham, MGB). Comorbidities were calculated on 250,000 randomly selected patients from each site individually and then averaged together. SCZ polygenic risk scores (PRS) in 64,190 genotyped patients of European ancestries were generated. A TOST test of equivalence was used to further evaluate the clinically significant SCZ comorbidities on whether they showed a well-powered absence of association with the SCZ PRS. **Results:** Clinical comorbidities were highly consistent across both institutions overall ($r=0.79$) and for SCZ specifically ($r=0.85$). We show significant replication of comorbidities previously identified in the literature including psychiatric conditions, epilepsy, diabetes, and hyperlipidemia (25/39 with $p<0.05$). There is correlation between the clinical and genetic associations both within ($r = 0.53, p = 6.2 \times 10^{-118}$) and across institution ($r = 0.49, p = 8.4 \times 10^{-81}$). After multiple test correction, there were 59 significant comorbid associations with SCZ. Of these, 11 had significant equivalence of SCZ PRS at both sites, indicating an absence of contribution to phenotypic comorbidity from common variant SCZ genetics. We further limited to those with non-significant SCZ PRS association, resulting in 5 phenotypes: alteration of consciousness, neurological disorders, epilepsy recurrent seizures convulsions, adverse drug events and drug allergies, and convulsions. **Discussion:** These 5 comorbid phenotypes may represent examples caused by other risk factors including rare variation, disease or treatment consequences or medical practice. Adverse drug events and drug allergies could be due to pharmacogenomics or interactions of multiple medications. Investigation into potential causal relationships among the remaining four using more detailed EHR data (e.g. EEGs, medications) and time is ongoing. This work presents an opportunity to identify and interpret non-genetic relationships among disease phenotypes.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3442. Identifying the conserved versus divergent structural sequences of the human pangenome

Authors:

H. Lee, S. Greer, D. Pavlichin, H. Ji; Stanford Univ., Stanford, CA

Abstract Body:

The Human Pangenome haploid assemblies provide an opportunity to determine which sequences are among the most conserved versus divergent in the human population. To conduct this study, we developed and applied a k-mer indexing strategy that enables one to make rapid comparisons among multiple different haploid genomes as well as parental haploid variation. This method was used to characterize 94 haploid assemblies and two reference genomes GRCh38 and CHM13. Highly conserved k-mer sequences had the following properties: (1) occurring only once per a given haploid genome and (2) having the same unique representation across all haploid assemblies. To identify segments indicating structural divergence and variation, we calculated the distance between tandem pairs of these highly conserved sequences. Then, we determined which pairwise segments varied in their distances across the entire pangenome data set. Based on the pairwise distance among 21 million highly conserved sequences, we identified 2,312 loci segments in which the pairwise distance diverged from both the human genome reference and other haploid assemblies among the pangenome set. These divergent pairs were indicators of loci with a structural variation. One-third of these divergent loci were private, occurring in only a single haploid assembly. Interestingly, multiple hotspots were identified across the human genome. In summary, we developed a rapid and scalable framework to characterize highly conserved versus divergent features across large sets of haploid assemblies. We identified a set of highly conserved sequences across the entire pangenome set. In addition, we discovered variable divergent loci that occur in a single or sets of haploid genomes. This k-mer-based analysis can be efficiently scaled to cover more haploid assemblies as they are generated. The outputs include simple-to-understand metrics that are useful in making comparisons from individual patient genomes to the Pangenome reference set.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3443. Identity-by-descent strategies uncover cryptic relatedness, deliver novel pedigrees, and facilitate gene discovery in amyotrophic lateral sclerosis

Authors:

K. Williams¹, L. Henden¹, I. Li¹, E. P. McCann¹, S. Topp², E. L. Scotter³, N. Grima¹, C. Dobson-Stone⁴, M. Mrkela³, J. B. Kwok⁴, G. Halliday⁴, R. Pamphlett⁴, R. H. Roxburgh³, T. Lin⁵, N. Laing⁶, M. Needham⁷, D. Schultz⁸, S. Mathers⁹, S. Vucic⁴, M. Kiernan⁴, R. Henderson⁵, P. McCombe⁵, A. Henders⁵, G. A. Nicholson¹⁰, D. B. Rowe¹, B. Smith², I. P. Blair¹, SALSA-SGC Consortium; ¹Macquarie Univ., Sydney, Australia, ²Kings Coll. London, London, United Kingdom, ³Univ. of Auckland, Auckland, New Zealand, ⁴Univ. of Sydney, Camperdown, Australia, ⁵Univ. of Queensland, Brisbane, Australia, ⁶Harry Perkins Inst. of Med. Res., Nedlands, Australia, ⁷Fiona Stanley Hosp., Perth, Australia, ⁸Flinders Med. Ctr., Adelaide, Australia, ⁹Calvary Hlth.Care Bethlehem, Parkdale, Australia, ¹⁰Concord Repatriation Gen. Hosp., Sydney, Australia

Abstract Body:

Do we really know our extended family? We may know our 3rd degree relatives (first cousins, great grandparents) or 4th degree relatives (e.g. first cousins once removed). But who knows their 5th, 6th, or even 9th degree relatives? We can leverage genomic data and our powerful new identity-by-descent algorithms [1] to uncover these relationships (cryptic relatedness) in large disease cohorts and use these genetic relationships for disease gene discovery.

Next-generation sequencing methods have fast-tracked gene discovery in many genetic diseases, but consequently have depleted the research space of genetically informative families (i.e. more than one affected individual with a DNA sample). This has constrained genetic studies in amyotrophic lateral sclerosis (ALS), where genetic factors remain to be identified in ~90% of patients, and reduced disease penetrance is common. Only 10% of ALS patients have a family history of disease, yet heritability estimates suggest all forms of ALS (familial and 'sporadic') have a significant genetic component. To facilitate gene discovery in ALS, we propose 'creating' new informative families by detecting cryptic relatedness using identity-by-descent (IBD) methodologies on SNP microarray data from an integrated cohort of 1,939 sporadic and familial ALS and FTD (frontotemporal dementia) patients from Australia and New Zealand. XIBD software [1] performed ~2 million pairwise IBD comparisons on the ALS and FTD patients. For each pair of samples, degree of relationship was estimated using total lengths of inferred IBD segments.

For individuals with a known pathogenic repeat expansion in *C9orf72*, we uncovered cryptic relatedness among 60% of sporadic and familial ALS and FTD patients from Australia and New Zealand. Among sporadic and familial ALS and FTD individuals without a known causal mutation, we identified 55 novel pairs, 1 triad, and a sextet of individuals ranging from 2nd (avuncular) to 6th degree relatives (2nd cousins once removed, the accuracy limit of our methodology). In one novel pair, retrospective genealogy analysis successfully linked the two pedigrees from different neurology clinics in Sydney, Australia. IBD analysis in this pair implicated 6 genetic linkage regions comprising 2.4% of the genome and encompassing 714 genes, defining critical regions for gene discovery.

In the absence of large historical pedigrees, genetic data can be successfully used to accurately uncover cryptic relatedness and confirm founder effects. We have employed identity-by-descent methodologies to genetically 'create' 57 new ALS/FTD families for gene discovery efforts.

[1] Henden, et al. (2016) *Bioinformatics*. 32:2389-91

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3444. Imaging derived phenotypes of the knee reveal novel genetic and clinical risk factors associated with knee osteoarthritis severity.

Authors:

B. Flynn¹, E. M. Javan¹, E. Kun¹, Z. Trutner², P. Jayakumar², V. Narasimhan¹; ¹The Univ. of Texas at Austin, Austin, TX, ²Dell Med. Sch., Austin, TX

Abstract Body:

Several large-scale studies of the genetic basis of musculoskeletal (MSK) disease including knee osteoarthritis (OA) have been carried out, though almost all rely on pre-existing disease assessment based on patient electronic health records (EHRs). These EHRs for knee OA are often unable to capture information about disease severity, only report individuals who sought care, and provide no additional information about clinically relevant biomarkers of disease progression. Though radiography is a mainstay in the clinical assessment of knee OA, it is challenging to perform manual annotation of disease severity at Biobank scale. To address this limitation, we trained a deep learning model to automatically grade knee OA severity based on the widely used Kellgren-Lawrence (KL) classification system, using training data from 200 dual energy X-ray absorptiometry (DXA) derived X-rays annotated by a team of orthopedic surgeons (AUROC = 0.98). We applied this model to 35,347 DXA images of the UK Biobank, and showed that 1,093 (33%) of the most severe cases (KL>2) not coded in the ICD-10 had self-reported knee pain. To obtain additional endophenotypes associated with knee OA, we performed deep learning based segmentation (98.9% accuracy) to measure the minimum joint space width (mJSW) between the femur and tibia and the tibiofemoral angle (TFA). After controlling for height, sex, age and body fat percentage, KL, mJSW and TFA remain significantly associated with leg pain while walking ($p=1.43e-84$, $p=7.65e-25$, $p=3.48e-9$), with fractures and falls ($p=9.12e-4$ and $p=0.0381$) in the past few years also associated with mJSW. We performed genome wide association studies (GWAS) for mJSW (N=29,257), TFA (N=24,456), and binary KL severity (N=28,435). We report 14 genome wide significant loci associated with mJSW (heritability 40.2%) and 2 loci associated with TFA (heritability 14.9%) not previously described for knee OA by the largest OA GWAS conducted to date, and 1 locus in common. Polygenic risk scores estimated from mJSW were also significantly predictive of OA status. We found no significant genetic associations with KL (heritability 3.5%), though mJSW and KL had a genetic correlation of -0.83 ($p=6.3e-5$). Our results suggest that for complex diseases where radiography is a primary diagnosis tool, image derived phenotyping is a cost-effective and scalable solution for quantifying disease severity, and that systematic measurement of quantitative endophenotypes can enable major increases in statistical power for genetic analyses. Our results also suggest that a major proportion of the risk of knee OA can be localized to common genetic determinants of knee anatomy.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3445*. Impact of heterogeneity-by-ancestry on GWAS in admixed populations.

Authors:

R. Mester, K. Hou, A. Bhattacharya, Y. Ding, K. Burch, B. Pasaniuc; Univ. of California, Los Angeles, Los Angeles, CA

Abstract Body:

Admixed individuals present both a challenge and an opportunity in disease mapping studies due to the unique genetic architecture of admixed populations. As methods for attributing variants to particular ancestry backgrounds become more reliable, disease mapping methods must balance the potential for capturing the additional signal with the anticipated loss of power due to correction for genetic structure within analyses. These tradeoffs become even more complex when causal SNP effect sizes differ by local ancestry. We investigate several statistical and biological factors which may result in this type of causal effect size heterogeneity, including differences of minor allele frequency and linkage disequilibrium across local ancestries.

We assess disease mapping methods in admixed populations to determine which methods are best suited for traits with heterogeneity in allelic effect sizes and frequencies of causal variants. We simulate admixed genotypes and phenotypes with different underlying causal architectures using information from the 1000 Genomes Project. We assess several scores in admixed populations ranging from 1 degree of freedom tests such as ATT, ADM, SNP1, and MIX that model the local ancestry signal to 2 degree of freedom tests such as SUM and TRACTOR that allow for heterogeneity in allelic effects across ancestries. We investigate metrics of both power and effective sample size, such that conclusions can be extended regardless of the number of individuals included in any given study.

We show that 1 degree of freedom tests such as MIX (Pasaniuc et al, 2011) are more powerful when allelic effects are similar across ancestries (Hou et al, 2021), and specifically that MIX has a larger effective sample size when the ratio of causal effects by local ancestry is ≥ 0.75 . However, 2 degree of freedom tests such as TRACTOR (Atkinson et al 2021) perform optimally when causal effects are in opposite directions or a large enough level of heterogeneity in same-direction causal effects is present. Using data from admixed individuals in the UK Biobank, we also investigate the extent to which differential linkage disequilibrium across local ancestries is expected to induce this amount of heterogeneity by local ancestry in causal effect sizes. From both scientific and social perspectives, it is important that admixed populations are incorporated more effectively in genetic studies. By providing direct examples of how these methods work and the benefits that can be derived from them, we enable studies to maximize their power and effective sample size in admixed populations.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3446. Implementing Mendelian Randomization to Assess Foetal Risk from Intrauterine Prescriptive Drugs for the Treatment of Diabetes, Hypertension and Thyroidism in Pregnancy

Authors:

C-J. Barry¹, N. M. D. Davies¹, V. Walker¹, A. Havdahl², G. Davey Smith¹, C. Burden¹; ¹Univ. of Bristol, Bristol, United Kingdom, ²Univ. of Oslo, Oslo, Norway

Abstract Body:

Background: Pregnant women are excluded from clinical trials for ethical and practical reasons; thus, little is known about the causal impact of intrauterine medication exposure on developing foetuses, creating high uncertainty within pharmacological research. Yet many chronic conditions necessitate maternal medication use during pregnancy, such as diabetes, hypertension, and thyroid disorders. Left untreated, these conditions have been found to be associated with adverse neonatal outcomes such as preterm birth, intrauterine growth restriction and foetal loss. The objective of this study is to discern whether genetic proxies for maternal drug exposures to medication for the conditions of interest can indicate causal adverse neonatal drug reactions to intrauterine exposure, through the application of Mendelian Randomization (MR).

Methods: We conducted a one-sample within-family MR analysis of maternal genetic drug targets on neonatal outcomes using parent-offspring trio data from The Norwegian Mother, Father and Child Cohort Study (MoBa). We identified genome-wide significant genetic proxies for drug targets related to treatments for the conditions of interest using DrugBank. We investigated the association of these maternal genetic variants with neonatal outcomes such as gestational age, birth weight, mode of delivery and APGAR score. We used within family models including parental genotype to assess potential intrauterine drug exposure effects, whilst controlling for environmental confounding.

Results: The cohort of interest to this study contains complete genetic data on up to 29,964 parent-offspring trios. For the relevant treatments, as indicated by the British National Formulary, we have identified and extracted a total of 1,064 maternal SNPs as potential instruments for maternal-neonatal intrauterine drug exposure. Preliminary summary statistics indicate analysis should be sufficiently powered, using the available linked phenotypic data. Adverse neonatal outcomes of interest to this study include those such as gestational age, birthweight, mode of delivery and Apgar score.

Conclusions: Genetic epidemiological data can provide evidence about the risks to neonates and benefits to mothers of intrauterine medication exposure. A combination of pharmacoepidemiological and genetic data is perhaps the best way to evaluate the potential risks currently associated with maternal medications, when clinical trials are less feasible. Thus, evidence from this study may be used with existing literature, clinical trials, and alternative study types to guide physicians and mothers during pregnancy.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3447. Improved breast cancer risk stratification by integration of a cross-ancestry polygenic model with clinical risk factors.

Authors:

P. Tshiaba¹, J. Sun¹, T. Tunstall¹, D. Ratman¹, B. Levy¹, P. Shah¹, J. Weitzel², M. Rabinowitz¹, A. Kumar¹, K. Im¹; ¹MyOme, Inc, Menlo Park, CA, ²Natera, Inc, Austin, TX

Abstract Body:

Breastcancer (BC) is the most common cancer diagnosed among women in the U.S exceptfor skin cancers. However, germline pathogenic mutations in high penetrancesusceptibility genes are rare and are associated with only 5-10% of BCs.Polygenic risk scores (PRS) have been shown to provide genetic informationabout BC risk beyond family history (FHx), but most have been developed forwomen of European ancestry. We developed a cross-ancestry PRS (caPRS) andcombined it with the Tyrer-Cuzick (T-C) clinical model to generate anintegrated risk score (caIRS). Thestudy included a training cohort and two independent validation cohorts consistingof women of diverse ancestries. We used a variety of algorithms to build PRSmodels including PolyPred, LD-Pred2, PRS-CS and PRS-CSx. We defined a caPRS asa linear combination of the best performing PRS model for each ancestry,weighted by fractional ancestry and PRS effect size. Individual ancestryinference and PRS centering were performed using principal components withindividuals in the 5 superpopulations of the 1000 Genomes Project used as thereference. Correlations between the caPRS and clinical variables included inthe T-C model were tested and accounted for during model training. The caIRSwas trained and calibrated in a cohort of 125,317 women. The caIRS wasvalidated in a cohort of 24,878 women from the Women’s Health Initiative (WHI)and a second cohort of 119,187 women from the UK Biobank (UKB). The validationfocused on remaining lifetime risk (RLR); with a RLR of $\geq 20\%$ classified as highrisk. Improvement in risk stratification was assessed using odds ratio (OR) perstandard deviation (SD) of PRS and area under the receiver-operator curve(AUC). caPRSquantile was highly correlated with odds of BC in all population groups tested.The caIRS had a stronger association with BC than the T-C score alone in bothvalidation cohorts, with an overall OR per SD of 1.78 (95% CI 1.69-1.88, $p=2.6 \times 10^{-105}$) compared with 1.35 (95% CI 1.27-1.42, $p=1.4 \times 10^{-23}$) in the WHI and 1.76 (95% CI 1.72-1.80, $p < 10^{-324}$) compared with 1.28 (95% CI 1.25-1.32, $p=5.8 \times 10^{-75}$) in the UKB. The caIRS improved risk prediction in allpopulations tested including self-reported Caucasian, Hispanic, African/AfricanAmerican and Asian women. The largest gain in performance was observed inHispanic women with a 10% increase in AUC. We demonstrated thataddition of a cross-ancestry polygenic risk score to the well-used clinicalrisk predictor Tyrer-Cuzick results in significant improvement in prediction ofbreast cancer with higher gains in multi-ethnic populations like Hispanic womenwhere T-C is less well calibrated.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3448. Improved genotyping of rare variants from AppliedBiosystems™ Axiom™ microarrays, using Support Vector Machine (SVM) prediction models.

Authors:

T. Webster, O. Mizrahi-Man, A. Mittal, J. Schmidt; Thermo Fisher Scientific, Santa Clara, CA

Abstract Body:

Databases such as ClinVar, contain rare genetic variants of importance, and researchers would like to design and use microarrays that detect the occurrence of such rare variants. However, genotyping rare variants is more challenging than genotyping common variants, and false heterozygous (Het) genotypes for rare variants called on microarray platforms can cause problems in the analysis of data by researchers. In 2021 we designed an improved algorithm specifically aimed at genotyping rare variants and developed a number of predictors. The new algorithm (RHA) significantly improved the positive predictive value (PPV) virtually without losing any correct predictions. Here we report on further highly significant improvements by using the RHA predictors and other predictors in an SVM prediction model. Significantly, by changing the probability cutoff values above which the genotype is called Heterozygous in the prediction model, one can tune the genotyping calls towards maximum sensitivity or maximum PPV.

The SVM prediction models were trained on genotypes from the UK Biobank under UK Biobank Resource Application Number 55681. The UK Biobank is a large-scale biomedical database and research resource, containing in-depth genetic and health information from half a million UK participants, all genotyped on Thermo Fisher Scientific's Applied Biosystems™ Axiom™ microarrays. Whole exome sequencing data for an approximately 200,000 subset of participants was also made available at the time of this study and were used as the truth data.

Microarray Het genotypes were collected from ultra-rare variants, those with computed minor allele frequency less than 0.001%, and were divided into training and test set data. In the test set data the RHA algorithm achieved 73.0% PPV while retaining 99.3% of the true Hets produced by the original AxiomGT1 genotyping algorithm. Using the SVM probability as a filter, the tradeoff between %PPV vs %True Hets retained can be tuned within a range of increasing PPV values. One probability cut-off achieves PPV values in the test set data that reach 95% with 93.8% of the true Hets retained, while an alternative cutoff achieves essentially no loss of true Hets (99% retained), while increasing the original PPV to 86.6%. This evaluation of SVM prediction models has been successfully extended to rare genotypes from additional Axiom™ microarrays.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3449. Improved Prediction Performance of Polygenic Risk Score with Gene Expression Data

Authors:

H. Xu; Augusta Univ., Augusta, GA

Abstract Body:

Polygenic risk score combines the genetic contribution from multiple genetic variants across the genome and has the potential clinical utilities in predicting disease risk for common human diseases. Current polygenic risk scores are based on the results from genome-wide association studies, where the information are from common genetic variants. With the availability of gene expression data, many gene expressions have been shown to be associated with the disease status. In this study, we build a polygenic risk score based on both common and gene expression data. We incorporate the effects of differentially expressed genes with $FDR < 0.05$ by including them into prediction model with the corresponding effect size estimates. Results from our simulation study show that our polygenic risk score has improved predictability measured by the C-statistic. Our polygenic risk score also has lower prediction error from cross-validation than the risk score without rare genetic variants. We applied our polygenic risk score method to the breast cancer data from the Cancer Genome Atlas to predict the risk of developing breast cancer.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3450. Improved risk prediction using functionally calibrated polygenic risk scores in admixed populations.

Authors:

X. Wang¹, S. Gonzales², C. Doyle¹; ¹Univ. of North Texas, Denton, TX, ²Univ. OF NORTH TEXAS, Denton, TX

Abstract Body:

Polygenic Risk Scores (PRS) are summaries of genetic information, typically using summation of effect size weighted allele counts for disease associated genetic variants. PRS can predict the risk of a disease for an individual, guide the selection of treatments, and benefit the development of medicines. Genome-wide association study (GWAS) summary data enables PRS estimation. However, the majority (>78%) of GWAS rely on individuals of European descent. Prediction accuracy with PRS remain moderate for most diseases, especially in admixed populations. This is largely due to challenges in 1) accurately estimating the effect sizes of genetic variants 2) identifying whether these variants are functionally relevant and 3) accounting for ancestry-specific linkage disequilibrium. In this study, we propose a Bayesian method that more accurately estimates the ancestry specific effect sizes of alleles for functionally relevant genetic variants by incorporating ancestry information, multiple GWAS and functional annotations into a prior distribution. PRS employing this method improves prediction accuracy for disease status. Simulation studies demonstrate that the proposed method outperforms alternative methods regarding the area under receiver operating characteristic curve for binary traits and the correlation coefficient between the predicted and observed trait values for quantitative traits. Application to admixed population data from the UK Biobank, and the 41 traits in the electronic Medical Records and Genomics (eMERGE) Network reveals that the proposed method outperforms comparable methods in risk prediction for most of the diseases/traits.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3451. Improvements in performance of polygenic risk scores for Alzheimer's disease in the Midwestern Amish using founder population-specific weights

Authors:

M. D. Osterman¹, Y. E. Song¹, L. D. Adams², R. A. Laux¹, L. J. Caywood², M. B. Prough², J. E. Clouse², S. D. Herington², S. H. Slifer², A. Lynn¹, M. Fuzzell¹, S. L. Fuzzell¹, S. D. Miller¹, K. Miskimen¹, L. R. Main¹, D. A. Dorfsman², A. F. Zaman², P. Ogrocki³, A. J. Lerner³, J. M. Vance², M. L. Cuccaro², W. K. Scott², M. A. Pericak-Vance², J. L. Haines¹; ¹Case Western Reserve Univ., Cleveland, OH, ²Univ. of Miami, Miami, FL, ³Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH

Abstract Body:

Alzheimer's disease (AD) is the most common type of dementia and is estimated to affect nearly six million Americans. Risk for AD is multifactorial, including both genetic and environmental risk factors. AD genomic research has generally focused on identification of risk variants. Using this information, polygenic risk scores (PRSs) can be calculated to quantify individual degree of disease risk due to genetic factors. The Amish are a founder population descended from German and Swiss Anabaptist immigrants. They experienced a genetic bottleneck after arrival in the United States, making it likely that their genetic architecture of AD risk is different from the broader European ancestry population. Here, we compare the performance of PRSs derived from genome-wide association studies (GWASs) of Amish individuals to those derived from a large European ancestry GWAS.

Our study population consisted of 1,981 Amish adults recruited from communities in Indiana (IN) and Ohio (OH). Participants were screened for cognitive impairment with further evaluation for AD or non-AD dementia. Genotype data were imputed using the Haplotype Reference Consortium Panel. The Amish individuals were split into two groups based on the primary site of recruitment (IN or OH) to minimize relatedness and allow for a training and validation set. For each group, genome-wide association analysis using a mixed model approach was conducted, accounting for inferred relatedness by use of a genetic relationship matrix. Age, sex, and relationship-adjusted principal component 1 (PC1) and PC2 were considered as covariates. PRSs were then calculated using weights from the other Amish group (i.e. IN Amish weights for OH Amish PRS), with and without a 500kb region surrounding either primary *APOE* variant. Area under receiver operating characteristic curve (AUC) was calculated for predictive models with and without covariate adjustments. The AUCs were then compared to predictive models using PRSs derived from the Jansen et al. (2019) summary statistics.

We found that the AUCs from the non-*APOE* PRS models improved from 0.55 to 0.63 in the OH group by using the IN Amish-specific weights instead of the European ancestry weights. They similarly improved from 0.54 to 0.63 in the IN group. The findings were consistent after inclusion of *APOE* genotype with an improvement in overall model AUC of 0.64 to 0.67 in the OH group and 0.56 to 0.65 in the IN group by using the Amish-specific weights.

The results show improvement in AD prediction through use of Amish-derived PRS weights. This work highlights the value of using founder population-specific weights when performing PRS analysis for founder and other populations.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3452. Improving Cross-Population Polygenic Risk Scores with a Tree-Guided Deep Learning Method

Authors:

E. Layne, Y. Li, M. Blanchette; McGill Univ., Montreal, QC, Canada

Abstract Body:

A common issue when training polygenic risk score (PRS) predictors is that training data will often have many individuals of a particular ancestry group (typically European), and relatively few from other backgrounds. Despite ongoing research efforts, there remains a gap in the ability of most PRS methods to generalize across ethnic backgrounds in this context of unbalanced training data.

Here, we present the results from adapting DendroNet, a machine learning that accounts for evolution in the function mapping genotypes to phenotypes, for this purpose. We modify the methodology for a multi-population PRS setting. Our method first constructs a binary tree, representing the relatedness of all population groups in the training set. The model includes a global set of per-SNP effect sizes, along with a set of SNP-specific effect size changes corresponding to each edge in the population tree. Effect sizes used for each population are a sum of the global effect sizes and their changes along the edges leading to it. L1 regularization is used to enforce sparsity amongst the effect sizes. This method combines the advantages of being able to leverage the large sample sizes of European datasets with the ability to learn effect sizes specific to other groups for whom only smaller datasets may be available.

We report results on the task of predicting expression levels of 15,000 genes, using a combined dataset of European and African individuals from the 1000 Genomes Project. We find that in approximately 1% of genes, the use of our framework's population-specific predictions result in a significantly improved predictive power, compared to a population agnostic lasso regression, while losses in predictive power are never observed. Ongoing work is investigating the power of this approach on various UK-BioBank polygenic traits.

Furthermore, we present two different implementations of our algorithm: a memory-efficient coordinate ascent implementation, capable of operating on very large-scale datasets in memory constrained environments, and a gradient-descent implementation, capable of leveraging GPU acceleration for high training speeds.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3453. Improving fine-mapping by modeling infinitesimal effects.

Authors:

R. Cui^{1,2}, R. Elzur^{1,2}, M. Kanai¹, J. Ulirsch¹, O. Weissbrod³, M. Daly⁴, B. Neale², Z. Fan⁵, H. Finucane²; ¹Broad Inst. of MIT and Harvard, Cambridge, MA, ²Massachusetts Gen. Hosp., Boston, MA, ³Eleven Therapeutics, Tel Aviv, Israel, ⁴Harvard Med. Sch., Cambridge, MA, ⁵Yale Univ., New Haven, CT

Abstract Body:

Fine-mapping aims to identify genetic variants that causally impact a given phenotype. State-of-the-art Bayesian fine-mapping algorithms (for example: SuSiE and FINEMAP) are widely applied in practice, but it remains challenging to assess their calibration (i.e., whether or not the posterior probability of causality reflects the true proportion of causal variants) in real data, where model mis-specification almost certainly exists and true causal variants are unknown. We first propose a real-data benchmarking strategy, Replication Failure Rate (RFR), that assesses consistency of fine-mapping between different sample sizes of the same cohort. Based on evidence from both RFR and functional enrichment, we believe that SuSiE and FINEMAP are likely miscalibrated in real data. Next, we contend that non-sparse genetic architecture, of several possible model misspecifications we tested in large-scale simulations, is likely the main contributor to this suspected miscalibration. We develop new fine-mapping methods, SuSiE-inf and FINEMAP-inf, that extend the computational ideas of SuSiE and FINEMAP to model infinitesimal effects in addition to a small number of sparse causal effects of interest. Our methods exhibit better calibration in simulations and improved RFR and functional enrichment in real data, with minimal loss of power and competitive computational cost. Furthermore, using the sparse fine-mapped variants generated by our methods to perform cross-population genetic risk prediction in the UK Biobank, we observed a substantial increase in predictive accuracy over SuSiE and FINEMAP. Our work improves our ability to pinpoint causal variants for complex traits, a fundamental goal of human genetics.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3454. Improving the accuracy and interpretation of polygenic risk score through modeling the pathways of disease and multiple risk factors

Authors:

Y. Yang, N. Lorincz-Comi, G. Li, X. Zhu; Case Western Reserve Univ., Cleveland, OH

Abstract Body:

Introduction The prediction of complex disease, one of the key components of precision medicine, is of great research and clinical interest. The polygenic risk score (PRS), a composite genetic risk indicator, has become a standard tool for quantifying genetic risk of complex disease. However, traditional PRS typically only explain a limited amount of disease heritability and improves risk prediction for only limited subsets of individuals. Although many improvements to the PRS have been developed, the gap between PRS-explained variance and disease heritability is still substantial. We thus propose a new pathway-guided PRS to improve prediction power and the potential for clinical utility. **Methods** It is well known that complex diseases are associated with multiple risk factors. Many genetic associations with complex disease may be through indirect associations with risk factors. Thus, we can construct a PRS based on direct and indirect genetic effects, which can improve the accuracy and interpretability of a PRS. In the implementation of our new PRS, we first constructed risk factor and disease-specific PRSs, then combined these PRSs into a final disease PRS by using penalized nonconvex regression. The estimated coefficients from risk factor and disease-specific PRSs can be considered the causal effects in multivariable Mendelian Randomization. **Results** We performed a simulation study in which UK Biobank genotype data was used to generate 20 risk factors with heritability $h^2=0.5$ each and the outcome disease was generated by the first 2, 5, or 10 risk factors and direct contributions of selected genetic variants. The results show that as the outcome disease is associated with more risk factors, traditional PRS performs worse while the pathway-guided PRS becomes better. Using 10 risk factors, the pathway-guided PRS explained 64% more of the variance in a continuous trait compared to the traditional PRS. For a binary trait, the pathway-guided PRS boosted the traditional PRS by ~17% in terms of Area Under the ROC Curve. The pathway-guided PRS also reliably selected true risk factors for the outcome disease. **Conclusion** We propose a pathway-guided PRS that is epidemiologically interpretable and can improve disease prediction accuracy. The significant improvement in prediction accuracy over a traditional PRS can be attributed to the leveraging of information on multiple risk factors for the disease of interest. The pathway-guided PRS also has a clear interpretation of risk factors affecting an outcome. We will apply the new pathway-guided PRS method to the prediction of cardiovascular disease outcomes in UK Biobank data and compare our method to existing alternatives.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3455. Impseer: Machine learning prediction of imputation quality

Authors:

M. Saad¹, K. Kunji²; ¹Qatar Computing Res. Inst., HBKU Research Complex, QCRI, Qatar, ²Qatar Res. Computing Ctr. (HBKU), Doha, Qatar

Abstract Body:

Background: Imputation is an important step in modern Genome-Wide Association Studies (GWASs) to cost-effectively increase genomic coverage and allow meta-analysis. Even with the decreasing costs of whole genome sequencing, the need for imputation will remain. Determining imputation quality is crucial for downstream association analysis. Here, we propose a machine learning framework to improve imputation quality estimation and therefore increase association statistical power.

Methods: We selected and generated many pre- and post-imputation features that can impact imputation quality, such as allele frequencies, reference and GWAS panel sample size, coverage, and linkage disequilibrium. A C++ program, Impseer, was written to generate these features. XGBoost was used to estimate the Pearson correlation (R^2) between the imputed dosages and the ground truth. Imputation was performed on the 1000 genomes data using Minimac and Impute, and their imputation quality measures, RSQR and INFO, were among the generated features.

Results: XGBoost models using all features predicted R^2 better than Minimac3's RSQR or Impute's INFO alone, with 11.5% and 21% improvement for rare (MAF: 0-0.01) variants and 6.4% and 3.7% improvement for common (MAF: 0.01-0.5) variants when compared to RSQR and INFO respectively. Gains occurred across all chromosomes and MAF intervals tested.

Conclusions: XGBoost predictions should be used instead of RSQR/INFO to determine imputation quality and therefore improve inclusion/exclusion filtering of SNPs for association analysis.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3456. Imputation of heritable gene expression reveals transcriptomic associations with neuroimaging phenotypes.

Authors:

N. Hoang¹, N. Sardaripour¹, Y. Chen¹, J. Park¹, M. Benton², J. Capra³, M. Rubinov¹; ¹Vanderbilt Univ., Nashville, TN, ²Baylor Univ., Waco, TX, ³Univ. of California San Francisco, San Francisco, CA

Abstract Body:

Motivation: Transcriptome-level association studies on neuroimaging phenotypes have revealed strong mappings between brain-wide gene expression profiles and measures of healthy brain states, such as intraregional homogeneity of neural activity. However, recent work is limited by the availability of complementary neuroimaging and transcriptomic data for large human populations. Instead, group-average data, such as the Allen Human Brain Atlas (AHBA) of six postmortem brains, are analyzed. While this approach is informative, it does not capture the high degree of individual variability present in human brain organization. This has led to a gap in associating neuroimaging phenotypes and gene expression with respect to individual variability in large populations.

Methods: In this work, we bridge this gap by imputing the heritable expression of several hundred genes in 10 brain regions for 890 individuals from the Human Connectome Project (HCP; 28.68 ± 3.73 mean age, 485 females). We applied PrediXcan to impute the genetically regulated component of gene expression using *cis*-regulatory genetic variants. This allowed us to study the otherwise inaccessible expression profiles for brain regions in living individuals. We found a strong association between the interregional correlations of the imputed gene expression patterns and assayed expression from the AHBA ($r = 0.69$, $p < 0.01$), demonstrating the convergence between measured and imputed expression.

Results: Our transcriptome-wide association studies on neuroimaging phenotypes in the HCP cohort identified 40 genes whose expression across the brain regions of interest correlated with intraregional homogeneity of neural activity. Further, we leveraged the PrediXVU catalog of gene-to-medical-phenome mappings derived from the Vanderbilt University biobank to identify 16 neurological and psychiatric phenotypes associated with these genes ($p \leq 0.01$, FDR).

Conclusions: Collectively, our approach bridges an existing gap in human neuroimaging studies by leveraging gene expression modeling to impute heritable expression in brain regions with existing neuroimaging data for many individuals. This modeling approach, and its rigorous statistical validation, opens a new direction for studying transcriptomic correlates of neural phenotypes. Ultimately, we propose that this approach can help reveal genetic mechanisms underpinning variation in human brain organization, and in this way support future discovery of genes linked to healthy and diseased brain states.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3457. In silico GWAS: the rapid and accurate (pre)computation of genetic disease associations using only population-level data.

Authors:

C. Foley^{1,2}, H. Runz³, B. Sun^{2,3}; ¹Data Sci. and Engineering, Res. and Development, Optima Partners, Edinburgh, United Kingdom, ²Univ. of Cambridge, Cambridge, United Kingdom, ³Biogen Inc., Cambridge, MA

Abstract Body:

Large-scale cohorts and biobanks with genetic measurements have significantly enhanced discovery of genetic links to disease and multi-omic phenotypes. For many diseases however, the extent of their genetic links are either incomplete or unknown. Broadly, resolution of risk variants and risk loci associated with disease can be improved by increasing the number of cohort and biobank samples, under a similar design, thereby increasing power to detect novel associations. There are two important limitations with this approach: (a) additional samples can be very costly to acquire; and (b) a method to pre-compute the expected number of additional risk variants and loci likely to be detected as a function of increased sample-size, for each disease, is currently unavailable. In combination, unsupervised enrolment of additional participants into cohorts and biobanks can be financially costly with few guarantees of enhancing biological insight.

Here we present In silico GWAS (isGWAS), a new algorithm, tool and online portal. Using only cohort and population-level information, isGWAS accurately approximates genetic variant and disease association summary data that is routinely computed in traditional genome-wide association study (GWAS) regression analyses, i.e., beta, standard error and p-values. isGWAS is lightweight and highly computationally efficient relative to traditional GWAS, as it avoids storing and processing individual-level data on each of potentially hundreds of thousands of individuals. For a given disease, users can specify a sample size for a cohort or biobank and isGWAS returns predictions of genetic links to disease. isGWAS therefore overcomes the outstanding challenge of estimating the expected number of additional disease associations identified as a function of sample-size. isGWAS can predict risk variants and loci from large-scale GWAS in currently understudied ancestries and cohorts as well as estimating likely gains in GWAS association signals from cross-cohort and cross-biobank analyses.

We demonstrate the accuracy and impact of isGWAS using Schizophrenia, Parkinson's, ALS and genetic data from several large-scale cohorts and biobanks to: (i) validate the tool by matching predicted results to known disease associated loci; (ii) predict future genetic disease discoveries from biobanks of increasing size; and (iii) combine population biobanks across ancestry to increase GWAS discoveries.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3458. Incorporating functional annotation with bilevel continuous shrinkage for polygenic risk prediction

Authors:

Y. Zhuang¹, N. Y. Kim², L. G. Fritsche¹, B. Mukherjee¹, S. Lee²; ¹Univ. of Michigan, Ann Arbor, MI, ²Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract Body:

Genetic variants in different functional categories have different shares of contribution to the heritability of complex traits, and recent studies have shown that the incorporation of functional annotation can improve the predictive performance of polygenic risk scores (PRSs). In addition, for certain phenotypes where only a small proportion of variants can be considered causal, PRS methods that employ a Bayesian framework with continuous shrinkage can account for such sparsity. It is possible that the annotation group level effect is sparse. However, the number of PRS methods that incorporate both annotation information and apply continuous shrinkage on effect sizes is limited. Here, we propose a PRS method which utilizes the functional annotation with a bilevel continuous shrinkage prior for each variant and each annotation group. This prior accommodates the varying genetic architectures both on the variant-specific level and on the functional annotation level. Computationally efficient Gibbs sampling is used for the posterior update. The proposed method uses GWAS summary statistics instead of individual-level data, and accounts for local linkage disequilibrium (LD) patterns through an external LD reference panel. We conducted simulation studies and investigated the predictive performance in settings with different genetic architectures. Results indicated that when there was a relatively large variability of group-wise heritability contribution, the gain in prediction performance from the proposed method was on average 8.0% higher AUC compared to PRS-CS. The proposed method also yielded higher predictive performance compared to PRS-CS in settings with different overlapping patterns of annotation groups and obtained on average 8.4% higher AUC. We applied the proposed method to both binary traits (e.g., type 2 diabetes) and quantitative traits (e.g., BMI) in three real world data sources (the UK Biobank, the Michigan Genomics Initiative (MGI), and the Korean Genome and Epidemiology Study (KoGES)), and two sources of annotations: ANNOVAR, and pathway information from the Kyoto Encyclopedia of Genes and Genomes (KEGG). While the degree of improvement is not consistent across different traits and sample population, we demonstrated that the proposed method generally improved PRS predictive performance.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3459. Incorporating functional annotations of rare variants in interpretable neural networks identify novel Alzheimer's disease risk loci

Authors:

A. Das^{1,2}, C. Lakhani², T. Raj³, D. Knowles^{1,2}; ¹Columbia Univ., New York, NY, ²New York Genome Ctr., New York, NY, ³Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Rare variants play a significant role in Alzheimer's disease (AD), but identifying their pathological effects remains challenging. However, the increase in whole-genome sequencing data availability now allows us to better investigate functions of non-coding as well as coding rare variants. Studies have used rare variants in burden and overdispersion tests, but limited research exists on building predictive models from rare variants. Previous strategies to identify novel relationships and increase predictive ability of complex diseases include applying machine learning methods and grouping variant-level information on the gene-level. In this analysis, we trained an interpretable neural network to predict AD status from functional annotations of rare variants. We input variant-level annotations into per-gene multilayer perceptrons to create latent gene scores and fed these scores, along with APOE-e4 status, age, sex, and ancestry traits, into a logistic network. Variants were separated by coding and non-coding groups, and annotations were divided by pre-transcriptional and post-transcriptional effects.

To train and test our model, we used data from the Alzheimer's Disease Sequencing Project (ADSP) consisting of whole-genome sequencing data for 5,122 AD cases and 6,599 controls over the age of 65. The majority of individuals were of European (57.2%), African (22.6%), or Hispanic (18.9%) ancestries. Functional annotations were selected from the Whole Genome Sequencing Annotator (WGSA) database, which contains over 800 annotations per variant. Our model predicts AD with 62.5% accuracy and an AUROC of 0.641, which is comparable to accuracies from other AD studies, and outperforms a baseline logistic regression on age, sex, APOE-e4 status, and ancestry (AUROC = 0.630). This indicates that the inclusion of rare variants alongside functional annotations via a carefully designed model structure improved performance. Further, given the interpretable structure of the model, we are able to see which features contributed most significantly to disease prediction. As expected, the well-established AD risk allele, APOE-e4, had a notably larger coefficient than all other features. Additionally, we observed BAG3 as one of the top candidates, which has been associated with Tau degradation, a common AD pathology. Collectively, we present a new approach to predict AD status and identify candidate genes for future studies.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3460. Incorporating related individuals in genome-wide association studies to reduce biases from familial effects and population stratification in Mendelian randomization.

Authors:

W. Jiang, H. Zhao; Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Abstract Body:

Mendelian randomization (MR) is a statistical approach to exploring the potential causal relationships between exposures and outcomes from genome-wide association studies (GWAS). Traditional MR methods assume all collected individuals are independent, therefore only unrelated individuals in GWAS are included in the MR analyses. Nevertheless, these analyses can be biased due to uncontrolled confounding factors between genetic variants and outcomes, such as dynastic effects, assortative mating, and population stratification. Recently, some novel MR methods have been proposed to reduce the biases by considering individuals with specific relationships such as siblings or parent-offspring trios. Since recruiting related individuals usually consumes more resources than recruiting unrelated individuals, the sample sizes in family-based GWAS and corresponding MR analyses are often limited, leading to the lack of statistical power to identify causal relationships.

In this study, to alleviate the biases from familial effects and population stratification and to address the power issue in family-based MR analyses, we proposed a novel MR method that can jointly analyze the data collected from both unrelated and pedigree-based individuals. The proposed method is robust to horizontal pleiotropy by allowing $\leq 50\%$ of selected genetic instrumental variables violating the exclusion restriction assumption. Simulation and real data applications on the UK Biobank data demonstrated that the proposed method has comparable performance with family-based MR methods in terms of alleviating biases. Besides, the method achieved higher power than family-based MR methods due to the increased sample size. Sensitivity analyses were conducted to demonstrate the robustness of the proposed method to horizontal pleiotropy. After considering the potential biases from familial effects and population stratification, we confirmed the previous finding that a high body mass index (BMI) increases the risk of hypertension. In contrast, the causal effect of BMI on educational attainment was significantly reduced after bias adjustment.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3461. Inference for set-based effects in genome-wide association studies with multiple interval-censored outcomes

Authors:

J. Choi¹, R. Sun²; ¹Rice Univ., Houston, TX, ²Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Abstract Body:

Massive genetic compendiums, such as the UK Biobank, have become an invaluable resource for identifying genetic variants that are associated with complex diseases. Due to the difficulties of massive data collection, a common practice of these biobanks is to collect interval-censored data via questionnaires and linkage of routine health check ups. Interval-censored data occurs when the time until an event of interest is known to occur during two time periods, but the exact time is unknown. However, there is a current lack of methodology to perform genetic association testing with interval-censored outcomes. Therefore, researchers are often forced to transform the data, such as into binary outcomes, which can lose information that may help in understanding the etiology of complex diseases. In this work, we develop a set-based inference method for jointly testing the association between multiple interval-censored outcomes and a group of genetic mutations, such as those in a gene or pathway. Combining multiple outcomes in genetic association tests can increase statistical power while identifying key biomarkers that are associated with multiple traits. Simulations show that combining multiple interval-censored outcomes can detect causal variants with increased power over using a test that only considers single outcomes. We further validate the value of jointly testing multiple correlated interval-censored outcomes by testing for the genetic effects of bone fractures and falls data from the UK Biobank. Fracture risk and fall susceptibility, which have been shown to be heritable traits, are important to investigate because of their prevalence in healthcare and the high costs associated with these outcomes. Our test was able to identify genes that have been previously associated with both outcomes. Other novel significant genes were identified as well, showing the potential of jointly testing multiple correlated outcomes to detect genes that are significant but are too weak to detect using a single outcome test.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3462. Insights into the comorbidity between type 2 diabetes and osteoarthritis: a genetics view.

Authors:

A. Arruda; Helmholtz Association Munich, Munich, Germany

Abstract Body:

Osteoarthritis (OA) and type 2 diabetes (T2D) are two of the most prevalent chronic health disorders worldwide. Observational studies report a positive epidemiological association between the diseases beyond their common risk factors, such as obesity and increasing age. Taking into consideration that the world's obesity rates, and average age are rising, this comorbidity pair can be considered an increasing global health challenge. Here, we aim to disentangle the genetic correlation between OA and T2D. Using summary statistics of large-scale GWAS from T2D (n=898,130) and OA phenotypes (n=490,345), we investigate the genetic intersection between the traits by performing statistical colocalization analysis of established association signals. For colocalizing regions, we derive a set of high confidence likely effector genes based on biological lines of evidence, including colocalization with molecular QTL from disease relevant tissues. Additionally, for each of these genes, we perform Mendelian randomization analyses between expression QTL and each disease. Seventeen genome regions show robust evidence of colocalization between T2D and at least one OA phenotype. Sixteen genes were defined as high confidence effector genes, including *TCF7L2* and the obesity related *FTO* and *IRX3* genes. *TCF7L2* is among the leading signals for T2D risk but had not been identified as associated with OA at genome-wide significance levels to date. We find statistical evidence that *TCF7L2* expression is causal for T2D and protective against knee OA. The identified shared effector genes support the epidemiologically known link between BMI and the investigated comorbidity. Moreover, we find that high confidence effector genes for hip OA and T2D are enriched for biological pathways associated with skeletal formation. In summary, we present an approach to disentangle the shared genetic aetiology of T2D and OA based on openly available GWAS summary statistics. By incorporating functional genomics data, we then derive a list of sixteen high confidence effector genes for the comorbidity.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3463. Integrating external controls by regression calibration for genome-wide association study

Authors:

L. Zhu; Michigan Technological Univ., Houghton, MI

Abstract Body:

The genome-wide association study (GWAS) has successfully revealed many diseases associated variants. For a case-control study, the adequate power of association test can be achieved with a large sample size, although genotyping large control samples is expensive. A cost-effective strategy is proposed by integrating external control samples with publicly available genotyped data. However, the naïve integration of external controls may inflate the type I error rates if ignoring the systematic differences (batch effect) between studies, such as the differences in sequencing platforms, genotype calling procedures, population stratification, and so forth. To account for batch effect, we propose to integrate External Controls into Association Test by Regression Calibration (iECAT-RC) in case-control association studies. Simulation studies showed that iECAT-RC not only can control type I error rates but also can boost statistical power in all models. We also applied iECAT-RC to the UK Biobank for the hypertension disease by considering genotype calling as the batch effect. The real data result shows that iECAT-RC can detect additional hypertension associated locus.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3464. Integrating rare and common variant association data to improve therapeutic target discovery.

Authors:

P. LoGerfo¹, L. D. Ward², A. M. Deaton¹, M. E. Plekan¹, C. Willis¹, R. A. Hoffing³, A. M. Holleman¹, L. Krohn¹, P. Nioi²; ¹Alnylam Pharmaceuticals, Cambridge, MA, ²Alnylam, Cambridge, MA, ³Alnylam Pharmaceuticals, Allston, MA

Abstract Body:

GWAS of common variants is a powerful method to detect loci associated with biomarkers and disease. However, most of the trait-associated common haplotypes consist of noncoding variants that span multiple genes, hampering causal gene identification. Although colocalization with eQTL and pQTL data have been useful in some cases to identify causal genes and directions of effect at GWAS loci, these data remain limited.

Rare-variant association studies (RVAS) that aggregate rare coding variants from exome sequencing data complement GWAS and have the advantage of providing a putative causal gene and direction of effect to guide the development of therapeutics. Distance from a sentinel GWAS variant to the closest gene body has been demonstrated as the most predictive feature for choosing a causal gene (Mountjoy et al., Nat. Genet. 2021), as the closest gene is over 50-fold enriched for harboring an independent rare variant association for the same phenotype (Backman et al., Nature 2021). Thus, we used UK Biobank array genotyping data and exome sequencing data to perform RVAS with a GWAS prior, testing only the genes closest to the GWAS sentinel SNPs to lower the multiple testing burden.

One notable result was a GWAS and RVAS association between variants in *MYLIP* and levels of LDL cholesterol. The sentinel SNP for association with LDL at this locus, *rs6920309*, is an intronic SNP 4930 bp from the 5' end of *MYLIP*. Rare (MAF < 1%) predicted loss of function (pLOF) and damaging missense (CADD score > 25) variants in *MYLIP* associate with a 0.16 SD decrease in LDL cholesterol (N = 643 carriers, p = 4.1 x 10⁻⁵). This association surpassed a p-value threshold Bonferroni-corrected for the number of genes tested for association with LDL (N = 202). *MYLIP* regulates LDL cholesterol uptake through direct ubiquitination of the LDL receptor.

These results suggest that inhibition of *MYLIP* gene expression may reduce serum LDL cholesterol levels by enhancing LDL uptake. Targeted silencing of *MYLIP* mRNA may have therapeutic application in reducing the risks of cardiovascular disease posed by elevated LDL cholesterol, though additional research is needed.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3465. Integration of gene expression data in Bayesian association analysis of rare variants

Authors:

G. Zhong, Y. Choi, Y. Shen; Columbia Univ., New York, NY

Abstract Body:

The statistical power to identify risk genes by rare *de novo* variants is generally low due to rarity of genotype data. Previous studies have shown that disease risk genes usually have high expression in relevant cell types, although for many diseases the identity of these cell types are largely unknown. Recent efforts in single cell atlas in human and model organisms produced large amount of gene expression data. Here we present two new methods, xTADA and VBASS, that integrate expression data to improve power of rare variants association analysis. Optimized for bulk RNA-seq and single-cell transcriptomics data respectively, xTADA and VBASS model the association of disease risk as a function of expression profiles of relevant tissue or cell types in Bayesian frameworks. VBASS uses both analytical likelihood function and neural network approximations in joint probability calculation, and it learns the importance of cell types jointly from expression and genetics data. On simulated data, both methods show proper error rate control and better power than extTADA, the state-of-the-art Bayesian method. We applied the methods to published datasets and identified more candidate risk genes than extTADA with supports from literature or data from independent cohorts.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3466. Integration of International Mouse Phenotype Consortium (IMPC) and UK Biobank data to identify genes associated with blood molecular phenotypes

Authors:

H. Haseli Mashhadi¹, V. Munoz-Fuentes², K. Babalola¹, IMPC Consortium, R. Wilson³, T. Groza⁴, H. Parkinson⁵; ¹European Bioinformatics Inst. (EBI), Hinxton, United Kingdom, ²EMBL EBI, CAMBRIDGE, United Kingdom, ³European Bioinformatics Inst. (EMBL-EBI), Cambridge, United Kingdom, ⁴EMBL - European Bioinformatics Inst., Hinxton, United Kingdom, ⁵EMBL-EBI, Hinxton, United Kingdom

Abstract Body:

The International Mouse Phenotyping Consortium (IMPC, www.mousephenotype.org) is a global research infrastructure which aims to ascertain function to every protein-coding gene, using the mouse as a model. The IMPC focuses on studying genes that are poorly characterised or the function of which is unknown, to ultimately inform human health and disease. In this study, we focused on blood associated phenotypes from the IMPC and the human data from the UK Biobank (www.ukbiobank.ac.uk) to investigate the genes that are associated with haematology and clinical blood chemistry traits. Comparing the two resources, we found 36 (20 haematological and 16 biochemical) traits that had been collected for both species. We investigate sexual dimorphism using data collected from 20K+ wildtype mice and 500K+ humans using a novel method to score the multicentre statistical results. Despite the strong agreement between the data collected by the network of globally distributed centres in the IMPC, our analyses show the agreement between humans and mice in the overall pattern of (males higher) Eosinophil and Monocyte cell counts. We also integrated murine and human gene-phenotype associations using gene orthology, comprising 7,488 mouse-human orthologues that have been studied by the IMPC. Finally, taking advantage of the broad phenotyping pipeline applied by the IMPC, we look at pleiotropic effects of genes associated with blood phenotypes, to uncover relevant physiological pathways. For this study, we acknowledge the UK Biobank for granting access to 500K+ participants in the UK as well as the support by the NIH common fund mechanism: 10U54HG00637 and EMBL-EBI Core Funds.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3467. Integrative approaches to unravel sex- and age-associated gene signatures and networks.

Authors:

K. Johnson, A. Krishnan; Michigan State Univ., East Lansing, MI

Abstract Body:

Many complex traits and diseases vary in their incidence and presentation in people of different ages and sexes. Yet, historically, age and sex have been largely ignored in basic/preclinical studies and in translating findings to clinical research and medical interventions. As a result, we lack fundamental understanding of how the genetic basis of physiology and disease is influenced by sex and age. Our goal is to develop a computational, data-driven framework that can aid scientists in systematically bridging these gaps in understanding age- and sex-specific mechanisms. Key to creating these comprehensive frameworks is the massive collection of public omics data. However, there are significant challenges to using these data as-is, foremost being that most samples are not annotated with age and sex. Therefore, we first manually curated ~30,000 microarray and RNA-seq samples associated with age and sex labels. Using this curated gold standard, we trained one-vs-all logistic regression models with elastic net regularization to classify transcriptome samples into age groups separately for females and males. Overall, the classifiers are able to discriminate between age-groups in a biologically meaningful way in each sex across technologies, with slightly lower performance in the middle age groups and in the few age groups where we have a low number of positive examples. In addition, the coefficients of the trained models capture 'gene signatures' that are characteristic to each age group. Enrichment analysis of these gene signatures helped us identify both novel and previously known functions, phenotypes, and diseases associated with the age-group in each sex. We are currently extending these models to predict sex and age-group labels for all publicly available human transcriptomes. Next, we will use these predicted labels to group samples into sex/age-groups, build coexpression networks based on samples within individual datasets, and integrate these networks to build sex- and age-specific gene functional interaction networks. Our gene signatures and networks will be valuable tools for hypothesis generation and studying biological processes as influenced by age and sex.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3468. Interrogating causality of epidemiological risk factors on adolescent idiopathic scoliosis: a two-sample Mendelian randomization study.

Authors:

D. Manousaki¹, F. Ghanbari², N. Otomo³, Y. Koike³, A. Khanshour⁴, S. Ikegawa³, C. Wise⁵, C. TERAO⁶; ¹Sainte Justine Hosp. Univ. of Montreal, Montreal, QC, Canada, ²Res. Ctr. of the CHU Sainte-Justine Univ. Hosp., Montreal, QC, Canada, ³RIKEN, Yokohama, Japan, ⁴Scottish Rite Hosp. for Children, Dallas, TX, ⁵Scottish Rite for Children, Dallas, TX, ⁶RIKEN IMS, Tokyo, Japan

Abstract Body:

Background: Adolescent idiopathic scoliosis (AIS) is the most common form of pediatric musculoskeletal disorder. Observational studies have pointed to several risk factors for AIS, but almost no evidence exists to support their causal association with AIS. Here, we applied Mendelian randomization (MR), a method known to limit bias from confounding and reverse causation, to investigate causal associations between epidemiological risk factors and AIS risk in Europeans and Asians. **Methods:** For our two-sample MR studies, we used single nucleotide polymorphisms (SNPs) associated with Body Mass Index (BMI), Waist-Hip ratio, Lean mass, childhood obesity, Bone Mineral Density (BMD), 25-hydroxyvitamin D (25OHD), age at menarche, and pubertal growth in large European genome-wide association studies (GWAS), and with BMD and age of menarche in Biobank Japan. We extracted estimates of the above SNPs on AIS risk from the European or Asian subsets of the largest multi-ancestry AIS GWAS (N= 7,956 cases & 88,459 controls). **Results:** The results of our inverse-variance weighted (IVW) MR estimates suggest no causal association between the above risk factors and risk of AIS. Pleiotropy-sensitive MR methods yielded similar results. However, restricting our analysis to European females with AIS, we observed a causal association between estimated BMD and the risk of AIS (IVW MR OR=0.1, 95% CI 0.01-0.7, P=0.02), but this association was no longer significant after adjusting for BMI, body fat mass, and 25OHD, and remained significant after adjusting for age at menarche in a multivariable MR. Moreover, BMD and age at menarche were not associated with risk of AIS in our Asian MR analysis. **Conclusions:** We demonstrated a protective causal effect of BMD on AIS risk in females of European ancestry, but this effect was modified by BMI, body fat mass, and 25OHD levels. Future MR studies using larger AIS GWAS are needed to investigate small effects of the above exposures on AIS.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3469. Inverted genomic regions between reference genome builds in humans impact imputation accuracy and decrease the power of association testing

Authors:

C. Chiang¹, L. Xia¹, J. L. Cahoon¹, D. Conti¹, C. A. Haiman¹, L. Kachuri², X. Sheng¹; ¹Univ. of Southern California, Los Angeles, CA, ²Stanford Univ., Stanford, CA

Abstract Body:

Over the last two decades, the human reference genome has undergone multiple updates as we complete a linear representation of our genome. Two versions of human references are currently used in the biomedical literature, GRCh37/hg19 and GRCh38. Conversions between these versions are critical for quality control, imputation, and association analysis. In the present study we show that single nucleotide variants (SNVs) in regions inverted between different builds of the reference genome are often mishandled bioinformatically. Depending on the array type, SNVs are found in approximately 2-5 Mb of the genome that are inverted between reference builds. Coordinate conversions of these variants are mishandled by both the TOPMed imputation server as well as routine in-house quality control pipeline, leading to underrecognized downstream analytical consequences. Specifically, we observe that undetected allelic conversion errors for palindromic variants in these inverted regions would destabilize the local haplotype structure, leading to loss of imputation accuracy and power in association analyses. Though only a small proportion of the genome is affected, these regions include important disease susceptibility loci, resulting in biased variant-specific signals due to poor imputation. For example, the *P*-value of a known locus (rs10763546) associated with prostate cancer on 10q11.22 would drop from 2.86×10^{-7} to 0.0011 in a case-control analysis of 20,286 Africans and African Americans (10,643 cases and 9,643 controls). These errors have implication for GWAS discovery and development and application of polygenic risk scores, which often rely on *p*-values for selection of index variants. We devise a straightforward heuristic based on applying the popular tool *liftOver* to basepairs preceding and succeeding the SNV of interest to easily detect these variants in the inverted regions between genome builds. We show that correcting for these alleles will locally improve imputation accuracy. Our software, *triple-liftOver*, can be found at <https://github.com/GraceSheng/triple-liftOver>.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3470. Investigating miRNA role in the neuropathology of Major Depression in a large postmortem brain sample.

Authors:

Z. Taylor¹, J. Drake¹, A. Denham¹, S-A. Bacanau², J. Heon Shin³, J. Kleinman³, T. Hyde³, V. Vladimirov¹; ¹Dept. of Psychiatry and Behavioral Sci., Coll. of Med., Texas A&M Univ., College Station, TX, ²Virginia Inst. for Psychiatric and Behavioral Genetics, VCU, Richmond, VA, ³Lieber Inst. for Brain Dev., Johns Hopkins Univ., Baltimore, MD

Abstract Body:

Major Depression (MD) is a debilitating disorder characterized by low mood and anhedonia that affects roughly one out of every six adults worldwide. Several large GWAS studies have already identified genome-wide significant variants associated with MD. However, little can be said about the functional impact of these variants from the GWAS alone. A complementary approach to understanding the underlying neuropathology of MD involves identifying transcriptome changes associated with major depression in postmortem brain tissues.

MicroRNA (miRNA), a class of small non-coding RNA with important gene regulatory functions and high expression in the brain, has been studied in relation to neuropsychiatric phenotypes. In this study, we used miRNA-Seq to compare the expression of 947 miRNAs between 150 MD patients and 150 matched controls in the subgenual anterior cingulate cortex (sACC). We applied statistical and bioinformatic analyses, e.g., single miRNA differential expression analysis (DEA), weighted gene co-expression analysis (WGCNA), expression quantitative trait loci (eQTL), and miRNA/mRNA correlation-based analyses to detect a set of miRNAs with a converging role in the etiology of MDD. We identified 50 miRNAs whose expression was associated with MD (at FDR of 5%). These included miRNAs that were both previously associated with MD and unique to this study as well as miRNAs associated with neurodevelopment or other psychiatric illness. The network analyses detected two significant miRNA modules associated with MD at the Bonferroni corrected $p \leq 0.05$, the module eigengenes (ME) of which were also negatively correlated with mRNA modules significantly associated with MD. Our eQTL analysis identified 86 significant cis- and 8 trans-eQTLs modulating the expression of these miRNAs. We further showed the eQTLs for these miRNAs are enriched for signals in GWAS of MD. Lastly, we note A to I RNA editing for a number of our top miRNAs highlighting additional biological mechanisms by which miRNA contribute to neuropathology of MD.

In conclusion ours is the largest to date postmortem brain miRNA expression study of major depression, and our ongoing analyses provides solid evidence of the importance of miRNA as a contributing factor to the development of MD.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3471. Investigating the effect of body size between menarche and first birth on breast cancer risk in later life: A lifecourse Mendelian randomization study

Authors:

G. Power¹, A. Hughes¹, T. Richardson¹, J. Heron¹, L. Bhatta², B. Brumpton², B. Asvold², J. Tyrrell³, T. Frayling³, A. Havdahl⁴, G. Davey Smith¹; ¹MRC Integrative Epidemiology Unit, Univ. of Bristol, Bristol, United Kingdom, ²Norwegian Univ. of Sci. and Technology, Trondheim, Norway, ³Univ. of Exeter, Exeter, Devon, United Kingdom, ⁴Univ. of Oslo, Oslo, Norway

Abstract Body:

Nulliparity is associated with increased reproductive malignancies and early first full-term pregnancy has been found to reduce risk of breast cancer. There is also evidence that increased weight in childhood is protective against breast cancer. Our research is focused on body size at different time points across the lifecourse and its effect on breast cancer risk, to understand the time frame in which undifferentiated nulligravid breast is most susceptible to carcinogenic insults. However, separating the effects of risk factors at different stages of the lifecourse is challenging due to confounding in conventional epidemiological settings. This is a key motivation behind using a Mendelian randomization (MR) approach.

Conventionally, MR studies use a single measurement to estimate the effects of an exposure on an outcome. Effects obtained are therefore often interpreted as the lifetime effect of the genetically predicted exposure. Our research exploits the notion that genetic associations may arise from the direct effects of the same inherited variants at different stages throughout life. This research seeks to assess the association between genetic variants and measures of body size taken between menarche and first birth at different intervals across the lifecourse. Using previously validated instruments for childhood and adult body size (n=453169) as well as instruments under development using data from several large prospective cohort studies (including Avon Longitudinal Study of Parents and Children (ALSPAC) Cohort (n=13964), HUNT Study (n=7683) and Norwegian Mother and Child Cohort Study (MoBa) cohort (n=36238), we will run univariable and multivariable MR to simultaneously estimate the effects of age-specific genetic proxies for body size on breast cancer. Thus far, we have identified independent associations with body size between menarche and <20 years and childhood body size (mean age: 10 years) as well as body size between menarche and <20 years and adulthood body size (mean age: 56.5 years) using ALSPAC and MoBa cohort data, with genetic correlation coefficients of $r_G=0.68$ and $r_G=0.46$, respectively. We have also successfully used genetic variation to separate the effects of early and later life body size in earlier research, which found novel evidence that higher childhood body size reduces fractures and higher adult body size is a risk factor for fractures. In addition, higher body size in childhood does not have a direct effect on cardiovascular disease in later life, rather, the effect of genetically predicted childhood body size on the cardiovascular disease outcomes analysed are a result of larger body size persisting into adulthood.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3472. Investigating the impact of the rare pathogenic variants linked to Intellectual Disability genes on the cognitive ability in adults: exome analysis of the UK Biobank cohort

Authors:

A. Kumar¹, T. Cuppens², K. Vachon³, M. Leclercq⁴, A. Droit^{5,6}, I. Dunham^{5,7}, F. Bolduc⁸; ¹EMBL-EBI, Cambridge, United Kingdom, ²Ctr. de recherche du CHU de Québec - Université Laval, Québec, QC, Canada, ³Dept. of Pediatric Neurology, Edmonton, AB, Canada, ⁴Université Laval, Québec, QC, Canada, ⁵Ctr. de recherche du CHU de Québec-Université Laval, Québec, QC, Canada, ⁷European Molecular Biology Lab., European Bioinformatics Inst., Cambridge, United Kingdom, ⁸Univ Alberta, Edmonton, AB, Canada

Abstract Body:

Aim/Background: Cognitive impairment, mental retardation, or specific intellectual disability (ID) affected individuals comprise an estimated 3-5% of world population. Extensive research has been done in identification of genetic variants including single nucleotide variants (SNVs) and copy number variants (CNVs) and elucidating their link to ID. Most of this research has been performed on children or young populations pertaining to addressing the developmental delay and early onset of IDs. However, reduced penetrance of these variants has also resulted in unaffected obligate adult carriers in the population. Hence, in the current work we extensively focus on investigating the UK Biobank (UKBB) cohort for the impact of rare (MAF $\leq 1\%$) protein coding pathogenic variants (SNVs) pertaining to genes already known to be associated with ID. **Methods:** 50,000 UKBB samples previously jointly-genotyped (GATK pipeline) were processed for variant annotation and downstream filtering. The variants were annotated for frequency (gnomAD), ensembl-vep annotations, pathogenicity (CADD, Clinvar etc). Subsequently, variants were filtered (MAF $\leq 1\%$) for protein coding, impact on the transcript (high & moderate), variant consequences, pathogenicity and variant quality metrics. Finally, the filtered variants pertaining to non-ID individuals in UKBB were regressed (using a generalized linear model) upon a set of 7 cognitive tests available in the UKBB portal where each individual was assigned a cognitive score based on their respective performances on these tests. **Results:** In the first stage of our analysis we randomly sampled 5000 non-ID UKBB individuals. After applying stringent filtering criteria we retained 39 pathogenic variants across 757 individual carriers. Upon linear regression with one of the cognitive tests (Pairs-Matching Test), variants pertaining to CRB1 (P<0.00125) and PNKP (P <0.03) were found to be statistically significant. This supports our assertion regarding the impact of pathogenic variants ID genes in unaffected carriers. We plan to extend this analysis for full 50K and 500K UKBB individuals and incorporate other cognitive tests for further investigation.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3473. Investigating the potential impact of PCSK9-inhibitors on mood disorders using eQTL-based Mendelian randomization

Authors:

A. Aman¹, E. A. W. Slob², J. Ward¹, B. Cullen¹, N. Graham¹, D. Lyall¹, N. Sattar¹, R. J. Strawbridge¹; ¹Univ. of Glasgow, Glasgow, United Kingdom, ²Univ. of Cambridge, Cambridge, United Kingdom

Abstract Body:

Background: Prescription of PCSK9-inhibitors has increased in recent years but not much is known about off-target effects. *PCSK9*-expression is evident in non-hepatic tissues, notably the brain, and genetic variation in the *PCSK9* locus has been recently shown to be associated with mood disorder-related traits. We investigated whether a genetic reduction in expression of *PCSK9* mRNA might have a causal adverse effect on mood disorder-related traits.

Methods: We used genetic variants in the *PCSK9* locus associated with reduced *PCSK9* expression (eQTLs) in the European population from GTEx v8 as the exposure, proxying PCSK9-inhibitors. The largest European ancestry genome-wide association studies of PCSK9 protein, and LDL cholesterol levels (positive controls), and three mood disorder-related traits (major depressive disorder, mood instability, and neuroticism) were used as outcomes. We conducted summary-based Mendelian randomization analyses to estimate the causal effect of reduced *PCSK9* expression on mood disorder-related traits. The analysis was replicated using data from eQTLGen, Brain-eMETA, and the CAGE consortium.

Results: We find that genetically reduced *PCSK9* gene-expression levels were significantly associated with reduced PCSK9 protein levels but not with increased risk of mood disorder-related traits. Further investigation of nearby genes demonstrated that reduced *USP24*-expression level was significantly associated with increased risk of mood instability (p-value range = 5.2×10^{-5} - 0.03) and neuroticism score (p-value range = 2.9×10^{-5} - 0.02), but not with PCSK9 protein levels.

Conclusion: Our results suggest that genetic variation in this region acts on mood disorders through a PCSK9-independent pathway, and therefore PCSK9-inhibitors are unlikely to have an adverse impact on mood disorder-related traits.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3474. Investigating the relationship between immune disease genetic risk and Alzheimer's Disease incidence.

Authors:

K. Clark, L-S. Wang; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Alzheimer's Disease (AD) is a neurodegenerative disease characterized by devastating memory loss and cognitive deterioration that affects more than 5 million adults over the age of 65. Physiological changes including amyloid plaque deposition and tau tangles begin 15 to 20 years before symptom onset, making AD difficult to treat beyond symptom management. Recent AD genome-wide association studies (GWAS) have identified more than 20 genetic loci associated with AD, many of which have been shown to play a role in the activation of the innate immune system. However, the biology underlying the association is still unexplained. To address this, we conducted a Polygenic Risk Score (PRS) analysis to identify immune risk factors and comorbidities relevant to AD. Using publicly available GWAS summary statistics, we calculated risk scores for 5 immune traits that have been previously shown to be associated with incidence of AD: Crohn's Disease, Lupus, Rheumatoid Arthritis, Psoriasis, and Multiple Sclerosis. We then applied logistic regression to the scores to identify correlations with AD case/control status in European samples from the National Alzheimer's Coordinating Center (NACC). In the future we plan to repeat this analysis in samples of African and East Asian ancestry. This work will allow us to use the known biological mechanisms behind the chosen immune diseases to better understand AD pathogenesis. This could potentially lead to drug repurposing and the advancement of specialized treatment for people of all ethnic groups.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3475. Investigation of Genetic Variants and Causal Biomarkers Associated with Brain Aging

Authors:

J. Lee, K. Jangho, S. Lee; Seoul Natl. Univ., Seoul, Korea, Republic of

Abstract Body:

Delta age is a biomarker of brain aging that captures differences between the chronological age and the predicted biological brain age. As large-scale datasets such as UK Biobank that contain the neuroimaging data and other biomarkers are becoming readily available, various methods to measure the brain-specific age have been developed. While existing research on delta age has been focused on prediction methods and finding an association with other phenotypes, we identified which groups of biomarkers causally affect the degree of brain aging. Moreover, an appropriate feature attribution method for medical images was used to identify brain regions that drive high delta age. First, we trained a 3D convolutional neural network model for age prediction with the T1-weighted structural brain MRI of 25,656 healthy white British samples in the UK Biobank. The test mean absolute error in the healthy individuals was 2.64 years. A visual saliency map of brain regions from integrated gradients showed that lower volumes in the fornix and the lower part of the thalamus are critical predictors of high delta age. Second, we performed a genome-wide association analysis (GWAS) for 38 million array-genotyped and imputed genetic variants for delta age. The results identified six genome-wide significant loci, including variants in *KLK3*, *AS1* and *STX6*. Additionally, we carried out the same GWAS procedure with the average voxel value of the two regions (lower volume in the fornix and the lower part of the thalamus) to evaluate whether the delta age prediction was truly driven by those regions. We observed nearly identical associated variants and large negative genetic correlations between delta-age and lower volume in the fornix (genetic $\text{corr} = -0.32$) and lower part of the thalamus (genetic $\text{corr} = -0.50$), respectively. Lastly, we examined the causal impacts of 249 metabolomic biomarkers and 61 blood-related phenotypes on brain aging using MR-Egger Mendelian randomization. Eosinophil count ($p\text{-value} = 5.16\text{E-}6$) was the most significant, followed by eosinophil percentage, neutrophil count, and total protein. All of these immune-related biomarkers causally increased the delta age. Our analysis revealed regions in the brain that are susceptible to the aging process and provided evidence of the causal and genetic connections between immune responses and brain aging.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3476. Investigations of the potential causal effects of the immune response to varicella-zoster virus on multiple health conditions: a Mendelian randomization phenome-wide association study

Authors:

X. Yu, H. Guo, A. Lophatananon, K. Mekli, K. Muir; The Univ. of Manchester, Manchester, United Kingdom

Abstract Body:

Background The immune response to infections is partly driven by the individual's genes, especially for genes in the major histocompatibility complex (MHC) region. The identification of herpes zoster virus (VZV) immune response-related genes helped understand the occurrence of disease in those persons infected with the virus. Apart from the infection, the reactivations of VZV can be a potential causal factor for multiple health conditions [1]. **Methods** To explore whether VZV specific immunity has a causal role on multiple health conditions, a Mendelian randomization phenome-wide association study (MR-PheWAS) of anti-VZV IgG levels with 1,730 health conditions was conducted using the UK Biobank cohort. To increase statistical power, we relaxed the p-value threshold of associations from a genome-wide association study. For each trait, we performed five MR approaches on two sets of instrumental variables (IVs), either including single nucleotide polymorphisms (SNPs) in the major histocompatibility complex (MHC) region or not. **Results** We identified 49 SNPs associated with anti-VZV IgG levels as IVs, five of which were located in MHC region. We found 75 traits passed conventional p value thresholds in at least three methods, 60 if we remove IVs in MHC region. Only 28 traits overlapped between the two IV strategies. The overlapping results indicated potential causal effects of higher anti-VZV IgG levels on increased risk of 14 diseases and lower risk of 9 diseases. The anti-VZV IgG levels were causally associated with 5 biomarker-related traits. **Conclusions** The anti-VZV IgG levels were causally associated with multiple health conditions, which provides new insight into these diseases. Attention should be paid to the MHC SNPs when selecting IVs for MR analysis.

[1] M. A. Nagel and D. Gilden, 'Complications of Varicella Zoster Virus Reactivation', *Curr. Treat. Options Neurol.*, vol. 15, no. 4, pp. 439-453, Aug. 2013, doi: 10.1007/s11940-013-0246-5.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3477. Is the portability of polygenic scores a matter of genetic ancestry alone?

Authors:

J. Wang¹, M. Zietz², J. Mares³, V. Narasimhan¹, M. Przeworski³, A. Harpak¹; ¹The Univ. of Texas at Austin, Austin, TX, ²Columbia Univ. Irving Med. Ctr., New York, NY, ³Columbia Univ., New York, NY

Abstract Body:

A major obstacle hindering the broad adoption of polygenic scores (PGS) is their lack of “portability” to people that differ—in genetic ancestry, environmental exposures or other characteristics—from the GWAS samples in which the PGSs were estimated. Currently, it is unclear to what extent the poor portability of PGS is caused by sample differences in genetic ancestry. While previous studies have shown decreased prediction accuracy in other ancestries, these groupings confound changes in allele frequencies and linkage disequilibrium patterns with many other factors, including distinct distributions of environmental effects. We investigate relative prediction accuracy as a function of continuous genetic divergence from the GWAS sample to a target genome, using the fixation index, F_{ST} . Across diseases and physiological traits in the UK Biobank, we find that prediction accuracy does not depend on genetic divergence from the GWAS sample when $F_{ST} < 0.01$, beyond which, prediction accuracy decays linearly with F_{ST} . For each trait, we examine whether covariates such as zip code or reported race explain prediction accuracy above and beyond genetic divergence from the GWAS sample. Finally, we examine how portability depends on genetic divergence from the GWAS sample in the specific genomic loci that contribute to the PGS. Discrepancies between this measure, versus genome-wide genetic divergence, as predictors of portability suggest that the confounding of genome-wide genetic ancestry with other factors partially explains the decrease in portability with genetic distance from the GWAS sample. Together, our analyses help disentangle the factors that impede the portability of PGS to groups underrepresented in a GWAS sample, and guide future decisions on the applicability of PGS to a given target sample.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3478. Isoform level transcriptome wide association study of prostate cancer risk reveals large number of transcript specific associations

Authors:

P. Barman, S. K. McDonnell, Z. Fogarty, D. J. Schaid, S. N. Thibodeau, N. B. Larson; Mayo Clinic, Rochester, MN

Abstract Body:

Prostate cancer (PRCA) is a highly heritable and polygenic disease trait, and large GWAS of PRCA risk have identified hundreds of independent genome-wide significant SNPs. Recent transcriptome-wide association studies (TWAS) leveraging tissue-specific expression quantitative trait loci (eQTL) datasets have linked multiple candidate genes to genetic risk of PRCA. However, these analyses have focused on total expression, and many genes have multiple isoforms with unique protein products. This information is largely lost when modeling total expression, and isoform-dependent associations with risk may go undetected. In this study, we used a large normal prostate tissue eQTL dataset to conduct an isoform-level TWAS (isoTWAS) of PRCA risk based on results from the PRACTICAL PRCA case-control GWAS meta-analysis. Our eQTL dataset consisted of 471 normal prostate tissue samples with available RNA-Seq and imputed Illumina Infinium 2.5M genotype data. Isoform-level expression values were generated using StringTie; after low abundance isoform filtering, expression was observed for 125,403 transcripts from 13,271 genes. We modeled relative isoform abundance using a log-ratio transformation relative to the highest expressed transcripts, and the FUSION TWAS software was used to train expression prediction models for genes with >1 expressed transcript. Further filtering based on minimum cis-heritability ($P < 0.01$) using a 50kb buffer led to successful training of isoform-level models for 11,550 transcripts corresponding to 4850 genes, including 2533 new genes that did meet similar criteria using the same data based on total expression. IsoTWAS was then performed using the PRACTICAL PRCA risk summary statistics, and significant associations were conservatively declared at the gene-level based on a Bonferroni-adjusted result for the top transcript model per gene, with gene-level significance then declared at an $FDR < 0.05$. We identified a total of 427 significant isoTWAS associations with PRCA risk, with the top 3 genes corresponding to *NKX3-1* ($P = 1.7E-40$), *PPP1R14A* ($P = 4.2E-27$), and *SIDT1* ($P = 1.0E-23$). When compared to the previously published PRACTICAL multi-tissue TWAS, 1303/4850 genes (26%) were novel with respect to available analysis results (104 significant at $FDR < 0.05$). For overlapping genes, while many significant isoTWAS genes were also significant in the PRACTICAL multi-tissue TWAS, often the isoTWAS yielded a stronger signal and/or sole evidence of a prostate-specific result. These findings suggest additional focus should be afforded to resolving PRCA genetic risk and the prostate transcriptome at a transcript level.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3479. Joint analysis of GWAS and multi-omics QTL summary statistics reveals a large fraction of GWAS signals shared with molecular phenotypes

Authors:

Y. Wu¹, T. Qi², J. Zeng¹, J. Yang²; ¹The Univ. of Queensland, Brisbane, Australia, ²Westlake Univ., Hangzhou, China

Abstract Body:

Molecular quantitative trait loci (xQTLs) are often harnessed to prioritize genes or functional elements underpinning variant-trait associations identified from genome-wide association studies (GWAS). Here we introduce a method, OPERA, that can jointly analyse summary statistics of GWAS and multiple omics layers of xQTLs to identify molecular phenotypes associated with complex traits because of shared causal variants. Applying OPERA to summary-level GWAS data for 20 complex traits ($n=20,833-766,345$) and xQTL data from five omics layers ($n=100-31,684$) finds ~42% of the GWAS signals shared with at least one xQTL, about half of which are not shared with expression QTLs. The signals shared with multiple molecular phenotypes, such as the *MSMB* locus for prostate cancer, are particularly informative to infer plausible regulatory cascades mediating the GWAS effects. Future studies with more molecular phenotypes, measured considering spatiotemporal effects in larger samples, are required to obtain a more saturated map linking molecular intermediates to GWAS signals.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3480*. Latent factor analysis reveals genetic components across GWAS traits

Authors:

A. Omdahl¹, S. Chhetri¹, Y. He², A. Battle²; ¹Johns Hopkins Sch. of Med., Baltimore, MD, ²Johns Hopkins Univ., Baltimore, MD

Abstract Body:

Today's GWAS measure the genetic landscape of thousands of phenotypes. Recent computational methods integrate findings across these studies to clarify relationships between traits and the pleiotropic activity of variants they share. These methods have uncovered novel genetic associations in rare diseases, identified distinct genetic disease etiologies, and prioritized genes regulating cellular development relevant to specific phenotypes. However, these approaches are limited by the number of traits they can jointly analyze, have been applied in few trait contexts, and can be difficult to robustly interpret. Furthermore, the effect of shared participants between studies may confound interpretation of these results and has not been closely examined.

To address these limitations, we present genomewide association sparse matrix factorization (gwasMF), a sparse factor analysis method for finding latent genetic components across GWAS summary statistics. gwasMF groups phenotypes into sparse clusters and identifies SNPs associated with these groups. The method employs weighted alternating-least-squares L1 regression and is initialized to detect broad pleiotropic effects. gwasMF may be applied across multiple cohorts and in specific or diverse sets of phenotypes. We also provide a framework to interpret gwasMF latent factors by projecting them onto a broader set of variants and testing for annotation and tissue-specific enrichment.

When applied to a diverse set of phenotypes, gwasMF identifies tissue-specific latent components and components enriched for pleiotropic SNPs. These sparse factors better group traits in a tissue-specific manner than non-sparse methods. gwasMF applied to Type 2 Diabetes-related phenotypes identifies factors corresponding to previously described pathways of obesity-mediated insulin resistance and beta-cell function. We also examine female infertility using gwasMF, which has no reported GWAS associations, leveraging summary statistics for related phenotypes with known associations where sharing of related genetic influences is likely. Finally, we evaluate the effects of cohort overlap between GWAS studies in the setting of factor analysis using both real and simulated data and consider matrix whitening as a possible solution. These findings merit follow-up to develop best practices for GWAS factor analysis and application of gwasMF to other understudied phenotypes. Future methods development integrating cohort overlap, LD structure, and variant annotations will further enrich the biological interpretability of our method.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3481. Leveraging cohorts from diverse ancestries to uncover generalizable genotype driven fetal- and adult-specific brain transcriptomic mechanisms in psychiatric disorders.

Authors:

A. Jajoo¹, C. Chatzinakos¹, C. Wen², PsychoENCODE Consortium, T. Bigdeli³, M. Gandal², N. Daskalakis¹; ¹McLean Hosp., Harvard Med. Sch., Belmont, MA, ²UCLA, Los Angeles, CA, ³SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY

Abstract Body:

Brain transcriptomic imputation utilizes genotype-expression reference panels to build predictive models genetically regulated gene expression (GReX). GReX models can be applied to GWAS to conduct transcriptome-wide association studies (TWAS) to prioritize gene-trait associations (GTAs) with functional importance. However, reference panels and GWAS studies have been biased toward European Ancestry (EA) samples, posing limitations for the identification of generalizable GTAs. To overcome this challenge, we propose meta-analyzing TWAS results across the ancestries by applying ancestry-specific models to the respective ancestry GWAS. To this effect, we trained ancestry-specific GReX models, of Admixed African Ancestry (AA) and EA, separately for healthy adult DLPFC tissue ($N_{AA}=165$, $N_{EA}=453$ subjects) and fetal brain tissue ($N_{AA}=164$, $N_{EA}=292$). Fetal models predict fewer genes and have lower R^2 model performance for overlapping genes compared to same ancestry adult models ($R^2=0.17$ for EA_adult vs $R^2=.11$ for EA_fetal, $P<e-16$). Within each tissue type, R^2 were comparable between ancestries, but AA models predicted 45% fewer genes compared to same tissue EA models likely due to less power. EA models compared to AA models performed worse in AA test set, while AA models performed comparably in EA test set. We applied these models on GWAS of bipolar disorder (BD), major depressive disorder (MDD), post-traumatic stress disorder (PTSD), and schizophrenia (SCZ), using at least one GWAS for each ancestry within each disorder. The average correlation of GTA z-scores was $Rho=0.49$ between adult and fetal analyses. Fetal tissue yielded a higher percentage of GTAs for SCZ which was expected because of neurodevelopmental origins, but surprisingly this was true for PTSD which has a later age of onset and lower heritability. The common GTAs from fetal and adult analyses tended to cluster in particular genomic regions (e.g, 17q21.31 in PTSD). Furthermore, the general observation is that bi-ancestry TWAS meta-analysis using just EA GReX models yields the least percentage of Bonferroni significant GTAs, while using ancestry-specific GReX models for respective ancestry GWAS yields the highest percentage (e.g., 2x increase in SCZ, 30x increase in PTSD). Our work shows that fetal and adult brain tissue based GTAs reveal shared and distinct genetic underpinnings of psychiatric disorders that operate in multiple stages, and that bi-ancestral TWAS is more beneficial when ancestry-specific GReX models can be applied on respective ancestry-specific GWAS.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3483. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries

Authors:

J. Zeng, Z. Zheng, P. Visscher; The Univ. of Queensland, Brisbane, Australia

Abstract Body:

Polygenic prediction of complex traits, including diseases, plays an important role in research and medical applications of the fast-growing genomic data in humans, but clinical applications in most diseases are currently impeded by the limited prediction correlation between the genetic predictor and outcome. Functional genomic annotations provide orthogonal information about the locations and effects of causal variants underlying complex traits, which can be used to improve polygenic prediction. In this study, we developed a Bayesian method (SBayesRC-eigen) to integrate GWAS summary statistics with functional annotations for polygenic prediction. We derived a low-rank model based on eigen-decompositions on quasi-independent LD blocks so that our method is resource-efficient and scalable to analysis of whole-genome imputed or sequence variants. We analysed 28 traits (including 8 diseases) in UK Biobank using 7 million imputed common SNPs and 96 per-SNP annotations. Compared to SBayesR using 1 million HapMap3 SNPs without annotations, SBayesRC-eigen improved prediction accuracy by 14% in European ancestry and by up to 33% in trans-ancestry prediction, outperforming state-of-the-art methods that incorporate annotations such as LDpred-funct and PolyPred-S. We identified factors affecting accuracy of prediction exploiting annotations, and found that the integration of GWAS with annotation data is most beneficial for traits with low heritability or small sample sizes, provides additional information to allele frequency and linkage disequilibrium categories, and is best used in a unified analysis where all parameters are estimated in one model. In addition, we observed a significant interaction between SNP density and annotation data, evidenced by a 2-fold higher annotation-derived improvement in prediction accuracy from 7 million SNPs. Furthermore, the per-SNP predictability enrichment in the hold-out sample was in proportion to the per-SNP heritability enrichment in the GWAS sample, suggesting that functional genetic architecture is informative to predicting which variants are important for polygenic prediction. Based on these findings, we predict that analysing whole-genome sequence variants with high-quality trait-relevant functional annotations will maximize the accuracy of polygenic prediction within and between ancestries.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3484. Leveraging Molecular-Diagnosis Workflow to Identify Modifier Genes in Diseases

Authors:

K. Schmitz-Abe, A. Tam, Q. Li, J. Lin, P. Agrawal; Children's Hosp. Boston, Boston, MA

Abstract Body:

Analysis of genetic data has helped advance the study of many rare diseases in part by providing molecular diagnoses to guide investigation. Meanwhile, many other rare conditions are either non-monogenic, have unidentified modifiers, or are poorly understood at the molecular level and need alternative approaches. We approach this problem by leveraging a collection of genetic and clinical data at Boston Children's Hospital (BCH) and adapting our previously described molecular-diagnosis Variant Explorer pipeline (VExP) for gene-based analysis. This dataset contains 7414 samples from 3093 families, 1240 of which are trios. We capitalize on the richness of this dataset to reduce noise by selecting mutations that meet a full-penetrance dominant/*denovo* model or full-penetrance recessive/compound-heterozygous model. Additionally, we use the results from various mutation-effect and spliceogenicity prediction software, constraint scores and MAFs from gnomAD, and quality scores for the aligned reads in the neural network of our molecular-diagnosis pipeline to score candidate variants. Furthermore, our pipeline facilitates downstream analysis by flagging candidates in genes with matches to key terms found in HGMD, OMIM, MGI, HPO, and Monarch as well as pathways found in GO or KEGG and identified as enriched through our newly developed Monte Carlo approach and providing links to relevant entries in these phenotype and pathway databases. Taking variants that were thus filtered, scored, and flagged by our workflow, we collapsed them by gene in four ways: loss-of-function, missense, synonymous and loss-of-function + missense. Then, we perform a Fisher exact test for each candidate gene between families affected by a particular condition ("cases") and families without the condition of interest ("controls"). In addition to exploring published methods for gene-based analysis, we are applying our *ad hoc* approach to identify potential modifier genes that differentiate fast-progressing cystic fibrosis from slow-progressing cystic fibrosis in patients homozygous for the *CFTR* F508del mutation. Our approach has identified potentially significant associations for known causal gene *CFTR* and potential modifier genes, including *SLC26A9*, *ACE*, *SCNN1D*, and *IL8*.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3485. Leveraging multiple reference panels by stacked regression TWAS identifies 6 novel independent risk genes for Parkinson's disease

Authors:

R. Parrish¹, A. Buchman², Y. Wang², P. De Jager³, D. Bennett², M. Epstein¹, J. Yang¹; ¹Emory Univ., Atlanta, GA, ²Rush Univ. Med. Ctr., Chicago, IL, ³Columbia Univ Med Ctr, New York, NY

Abstract Body:

Existing transcriptome-wide association study (TWAS) methods often assume a single reference panel of transcriptomic and genetic data from a relevant tissue to train imputation models of gene expression to estimate eQTL weights to be integrated with GWAS data. However, multiple reference panels of a given tissue often exist; for example, the Religious Orders Study (ROS), Rush Memory and Aging Project (MAP), and Genotype-Tissue Expression project (GTEx) all profile transcriptomic data from prefrontal cortex tissue. Additionally, there often exist publicly available expression imputation models which utilize a variety of reference cohorts and regression methods. We hypothesize that leveraging expression imputation models of the same target gene that are fitted from multiple reference panels and by multiple regression methods can improve TWAS performance, compared to the standard use of a single reference panel and single regression method.

We developed the Stacked Regression based TWAS (SR-TWAS) tool that uses the ensemble Machine Learning technique of Stacked Regression to form optimal linear combinations of expression imputation models (i.e., base models) of the same target gene and tissue type. By combining these base models, SR-TWAS can leverage multiple reference panels with increased effective training sample sizes as well as multiple regression methods which assume different underlying genomic architecture.

This study conducted a TWAS of Parkinson's disease (PD) using the most recent GWAS summary dataset (n~482k). We applied SR-TWAS to base models trained on multiple reference panels from prefrontal cortex tissue by TIGAR and PrediXcan. SR-TWAS identified 50 significant TWAS risk genes of PD, including 35 novel TWAS risk genes. Of these, 14 are known GWAS risk genes, and 33 are within 1MB of a known GWAS risk gene. From these results we curated 11 independent significant TWAS risk genes by selecting the most significant genes among genes that have shared test regions. We identified 6 novel independent TWAS risk genes. Two of these novel risk genes are known GWAS risk genes (CNTN1, LINC02210) and 3 are within 1MB of known GWAS risk genes (CCNT2, CCDC158, PTP4A2P2).

Our results show the advantage of using multiple reference panels and regression methods for SR-TWAS. Our SR-TWAS tool is expected to increase expression imputation accuracy and TWAS power for mapping risk genes of complex diseases and will be publicly available on Github.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3486. Leveraging multiple traits to detect shared non-additive genetic variation in genome-wide association studies

Authors:

A. Bass, S. Bian, A. Wingo, T. Wingo, D. Cutler, M. Epstein; Emory Univ., Atlanta, GA

Abstract Body:

Detecting and characterizing non-additive genetic variation in genome-wide association studies provides valuable insights into the genetic architecture of a trait. Identifying this variation---such as gene-by-gene or gene-by-environment interactions---can be challenging due to low statistical power and difficulty in specifying unknown (or latent) genetic and environmental variables. While there are strategies to detect non-additive variation for a single trait, there is limited research on incorporating multiple traits even though this can substantially increase power. In particular, a shared interaction effect influences not only the variance of a trait but also the covariance between different pairs of traits, and these differential covariance patterns can be harnessed to improve power. We propose a new framework that leverages this information across multiple related traits to detect covariance quantitative trait loci (covQTL), i.e. any latent non-additive genetic variation that impacts the covariance among traits. Our approach implements a flexible Kernel-based testing procedure that compares the similarity of one or many genotypes to appropriately transformed traits without having to specify the latent variables. The method, called Primary and Latent Covariance Analysis (PLCA), accounts for population structure and other known covariates while being computationally efficient for genome-wide analysis. We validate PLCA on simulated data and demonstrate it on the UK Biobank data using five obesity-related traits measured across 330,328 individuals with over 5 million single nucleotide polymorphisms. We find strong evidence of covQTLs which implies genome-wide polygenicity of non-additive effects. Thus PLCA can be applied to multiple related traits to discover shared non-additive genetic variation and help reveal underlying biological mechanisms in genome-wide studies. PLCA is implemented in the publicly available R package `plca`.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3487. Leveraging polygenic risk scores of biomarkers to identify obesity endotypes.

Authors:

M. Bagheri¹, B. De Meulder², M. Shi³, J. Annis¹, J. Peterson¹, E. Brittain¹, J. Ferguson¹, J. Mosley¹; ¹Vanderbilt Univ. Med. Ctr., Nashville, TN, ²Association European Inst. for Systems Biology and Med., Vourles, France, ³VUMC, Nashville, TN

Abstract Body:

Background: Obesity (body mass index [BMI]≥30) is strongly associated with cardiometabolic diseases. However, the risk is not always associated with the severity of obesity, suggesting there are subgroups of obese individuals who do not develop adverse outcomes. We hypothesized that clustering approaches that use multi-omics data can identify biologically-informed subgroups with differing patterns of disease associations and differing underlying risk mechanisms. **Methods:** Genetic data from Framingham/Malmö and KORA/TWINS was used to develop single nucleotide polymorphism-based predictors of expressed proteins and metabolites. The genetically predicted levels of biomarkers were used to impute biomarker levels in 4,625 obese genotyped BioVU participants. We constructed multi-omic clusters of obese individuals, defined as obesity endotypes, using Similarity Network Fusion in conjunction with Consensus clustering. The association between each endotype and 14 cardiometabolic diseases was assessed using multivariable logistic regression models adjusting for sex, age, BMI, lipid profiles, glucose and medications. Pathway and enrichment analyses were performed using Reactome and MetaboAnalyst. **Results:** We identified 5 obesity endotypes that were associated with different diseases: endotype1: chronic kidney disease (OR:1.22, 95% CI:1.03-1.45, p=0.02); endotype2: no association; endotype3: chronic pulmonary embolism (OR:4.58, 95% CI:1.36-13.98, p=0.009); endotype4: atrial fibrillation (OR:0.73, 95% CI:0.57-0.92, p=0.009); and endotype5: hypertension (OR:0.65, 95% CI:0.46-0.95, p=0.02) and peripheral arterial disease (OR:1.28, 95% CI:1.01-1.61, p=0.04). Each endotype was related to distinct pathways; notably endotype2 was related to Interleukin-6 signaling. **Conclusion:** This innovative network modeling provides supporting evidence for obesity endotypes, and novel mechanisms for progression of only some endotypes to cardiometabolic diseases. This may inform future prognostic tools constructed from risk prediction models in precision medicine.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3488. Leveraging the use of meta-analysis summary statistics to improve gene expression prediction models.

Authors:

M. Mews¹, A. Naj², J. Below³, B. Kunkle⁴, W. Bush⁵; ¹Case Western Reserve Univ., Cleveland, OH, ²Univ. of Pennsylvania, Philadelphia, PA, ³Vanderbilt Univ. Med. Ctr., Nashville, TN, ⁴Univ. of Miami, Miami, FL, ⁵Case Western Reserve Univ, Cleveland, OH

Abstract Body:

Genetically regulated gene expression as assessed via Transcriptome-wide association studies (TWAS) can be used to elucidate the mechanistic basis of Alzheimer's Disease (AD). However, a limitation of current TWAS approaches is the requirement for individual-level datasets with both genotypes and gene expression values to produce effective prediction models. We sought to improve the predictive performance of existing gene-based models by leveraging easily accessible summary statistics taken from large-scale expression quantitative trait loci (eQTL) meta-analyses.

We used Brain-Cortex *cis*-eQTL data from European-descent individuals (Average N=2,547) and re-ran an eQTL meta-analysis using effect estimates (N=15 cohorts; no GTEx) and their standard errors. To perform gene-based elastic-net regressions using the produced summary statistics, we used the *lassosum* R package (Mak *et al.* 2017) with 1000 Genomes data as a genomic reference and GTEx v8 genotype, normalized Brain-Cortex gene expression and covariate data (N=205) as a testing panel for model validation and dynamic selection of the optimal mixing/penalty parameters. The maximum R² produced by this pipeline was compared to elastic-net models within the PredictDB resource which were generated using only Brain-Cortex GTEx_v8 data where the mixing parameter was set at 0.5. The SNP-weights generated by our models were subsequently applied to AD meta-analysis summary statistics from Kunkle *et al.* 2019 via S-PrediXcan to assess whether predicted gene expression is associated with AD risk.

Our pipeline produced gene-based models that had an R² predictive performance that improved an average of 11.7% (SD: 6.9%) over PredictDB models in 96.2% of genes tested (N=4,540) when restricting our pipeline to identical SNPs found in the PredictDB model. Without this SNP restriction, our models had an improved predictive performance averaging 8.3% (SD: 7.6%) based on R² over PredictDB models in 86.7% of genes tested (N=4,598). The optimal mixing parameter dramatically differed from PredictDB with a median of 0.08 when SNPs were not restricted highlighting the utility of including variants with small effects. Using models with the restricted SNP set, we identified 12 genome-wide significant gene expression models (Effect sizes: -9.0, 1.9) associated with AD relative to only four using PredictDB models (Effect sizes: -0.9, 0.8).

We demonstrate improvement upon existing TWAS reference panels and increase gene-expression predictive performance by leveraging the power of summary statistics. Our findings show AD association studies based on TWAS approaches may be improved by meta-analysis reference panels.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3489. Leveraging tissue-wide functional regulatory information improves portability of trans-ancestral polygenic risk score applications

Authors:

B. Crone, A. Boyle; Univ. of Michigan, Ann Arbor, MI

Abstract Body:

Polygenic risk scores (PRS) are an effective tool in stratifying disease risk for individuals of European ancestries. The predictive accuracy of these risk models continues to grow as genome-wide association study (GWAS) sample sizes increase. However, large GWAS cohorts are largely European-centric, restricting power in trait-variant association discovery within understudied non-European populations. Trans-ancestral PRS models take advantage of associations in large European cohorts, but predictive accuracy is hindered by differences in genetic linkage between populations. Recent studies have shown that leveraging functional genomic information in prioritizing non-coding variants in PRS calculation can improve prediction accuracy of both within- and trans-ancestral applications. These functional methods limit prioritized mutations to explicitly modeled cell-type-specific transcription factor binding sites, fail to incorporate perturbations across whole tissues or organs and may not perform well with unknown or missing transcription factor data. Here we present a SNP prioritization model for trans-ancestral PRS applications that encompasses functional regulatory evidence from the RegulomeDB, a genome-wide probabilistic model of tissue-specific regulatory activity with SNP-level resolution. We construct both single-tissue and multiple-tissue models to prioritize functionally-enriched genomic mutations for trans-ancestral PRS applications. We show that by highlighting functional regulatory mutations across whole tissues and organs we improve the predictive accuracy of trans-ancestral PRS models in complex disease by identifying shared regulatory mechanisms across ancestries.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3490*. Linking the joint genetic structure of neuroanatomical phenotypes with psychiatric disorders.

Authors:

A. Auvergne, N. Traut, L. Troubat, H. Julienne, R. Toro, H. Aschard; Inst. Pasteur, Paris, France

Abstract Body:

There are increasing evidence that brain magnetic resonance imaging (MRI) phenotypes are linked to mental disorders and that the two harbor substantial genetic correlations. However, identifying the architecture linking genetics, MRI traits and mental disorder has proven challenging, and new approaches are needed for deciphering these complex relationships. Here, we show how a genetically-driven approach can reveal latent components underlying MRI phenotypes that are strongly associated with mental disorder.

We first conducted univariate and multivariate genome-wide association studies (GWAS) for nine MRI-derived brain volume phenotypes in 20K UK Biobank participants, using JASS, a robust and computationally efficient multitrait analysis pipeline we recently developed. We clustered these variants based on their multitrait association with MRI phenotypes using an optimized k-medoids approach along an innovative data-driven algorithm for selecting the number of clusters. We then assess the ability of those clusters to distinguish disease-associated and non-disease-associated variants as compared to the original MRI phenotypes. In practice, we derived genetic risk score (GRS) for both the original MRI phenotypes and each inferred cluster, and tested the association between these GRS and six mental disorders: bipolarity, attention-deficit/hyperactivity disorder (ADHD), autisms, schizophrenia, obsessive-compulsive disorder (OCD) and major depressive disorder (MDD).

We observed very limited association between the GRS of MRI phenotypes and disease, with only two of them showing nominal association with some GRS (SCZ, min $P=0.01$, and ADHD, min $P=4 \times 10^{-4}$). Conversely, we identify nominal association with multiple GRS of clusters for all diseases, some of them remaining significant after Bonferroni correction (SCZ, min $P=3 \times 10^{-6}$, and BIP, min $P=4 \times 10^{-5}$). By construction, those associated clusters correspond to linear combination of the volumetric phenotypes studied, and we argue that they can be interpreted as latent endophenotypes. Altogether, this approach can offer a powerful tool to identify genetically-driven neuroanatomical latent structure associated with mental disorders.

Grants :

This research was supported by the FRM (ECO202106013759).

This research has been conducted using the UK Biobank Resource under Application Number 18584.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3491. Local ancestry at the MHC region is associated with disease heterogeneity in a multi-ethnic lupus cohort.

Authors:

O. Solomon¹, C. Lanata², C. Adams³, J. Nititham², K. E. Taylor⁴, S. Chung⁴, M. Dall'Era⁴, J. Yazdany⁴, B. Pons-Estel⁵, T. Tusié-Luna⁶, B. P. Tsao⁷, E. F. Morand⁸, M. E. Alarcón-Riquelme⁹, L. F. Barcellos³, L. A. Criswell²; ¹Univ. of California, Berkeley, Berkeley, CA, ²NHGRI, Bethesda, MD, ³Univ. of California Berkeley, Berkeley, CA, ⁴Univ. of California San Francisco, San Francisco, CA, ⁵Regional Ctr. for Autoimmune and Rheumatic Diseases (CREAR) of Rosario, Rosario, Argentina, ⁶Inst. Natl. de Ciencias Médicas y Nutrición Salvador Zubiran and Inst. de Investigaciones Biomédicas de la Univ. Natl. Autónoma de Mexico, Mexico City, Mexico, ⁷Dept. of Med., Med. Univ. of South Carolina, Charleston, SC, ⁸Sch. of Clinical Sci., Monash Univ. Faculty of Med., Nursing & Hlth.Sci., Melbourne, Australia, ⁹Pfizer—Univ. of Granada—Andalusian Government Ctr. for Genomics and Oncological Res. (GENYO), Granada, Spain

Abstract Body:

Systemic lupus erythematosus (SLE) is an autoimmune disease that results in debilitating clinical manifestations that vary in severity by race and ethnicity with a disproportionate burden for admixed populations. Differences in global and local genetic ancestry can shed light on the mechanisms that contribute to these disparities, including increased lupus nephritis prevalence, younger age of onset, and autoantibody status in non-White populations. A total of 1,139 SLE patients of European, African American, and Hispanic race and ethnicities were genotyped using the Affymetrix LAT1 World array. Global ancestry proportions were estimated using ADMIXTURE and local ancestry was estimated using RFMIXv2.0. Genetic map coordinates were used to generate windows of consecutive local ancestry across the genomes. We tested for associations between SLE manifestations including lupus nephritis, age of onset, and anti-double stranded DNA (anti-dsDNA) autoantibodies with both global ancestry proportions and local ancestry windows. Analyses were conducted genome-wide, at candidate SNPs, and within the MHC region (chr6:28477797-33448354). In African American patients, increased European local ancestry was associated with anti-dsDNA autoantibodies in the 18 of 50 (36%, FDR- $p < 0.05$) MHC windows indicating increased European ancestry in patients with anti-dsDNA autoantibodies in those MHC windows. In Hispanic patients, 10% of MHC windows had statistically significant (FDR- $p < 0.05$) increased Indigenous American ancestry was associated with anti-dsDNA autoantibodies. Significant windows spanned MHC Class I and II genes. These findings provide evidence that local ancestry, specifically within the MHC region, contributes to SLE manifestations in admixed populations.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3492. Loci for Cognitive Preservation in the Midwestern Amish

Authors:

L. Main¹, Y. Song², R. Laux², K. Miskimen², M. Cuccaro³, P. Ogrocki², A. Lerner², J. Vance⁴, M. Fuzzell², S. Fuzzell², M. Osterman², S. Hochstetler², A. Lynn², D. Dorfsman⁴, L. Caywood⁴, M. Prough⁵, L. Adams⁶, J. Clouse⁷, S. Herington⁸, W. Scott⁴, M. Pericak-Vance⁹, J. Haines¹⁰; ¹Case Western Reserve Univ., Lakewood, OH, ²Case Western Reserve Univ., Cleveland, OH, ³John P. Hussman Inst. for Human Genomics, Miami, FL, ⁴Univ. of Miami, Miami, FL, ⁵Univ. of Miami, Bristol, TN, ⁶Univ Miami, Miami, FL, ⁷Univ. of Miami, Miami, OH, ⁸Univ. of Miami Miller Sch. of Med. Hussman Inst. for Human Genomics, Elkhart, IN, ⁹Univ. of Miami Miller Sch. of Med., Miami, FL, ¹⁰Case Western Reserve Univ, Cleveland, OH

Abstract Body:

Alzheimer's disease (AD) is a significant cause of morbidity and mortality in the U.S. and continues to grow in prevalence as the population ages. There have been numerous risk loci associated with AD; however, the alternative of identifying loci that decrease AD risk or delay onset may lead to novel therapeutic targets. The Midwestern Amish are an ideal population for studying the genetic underpinnings of cognitive preservation, as they have a lower rate of dementia than the general population, a relatively homogeneous environment, and lower genetic diversity resulting from a founder effect. We define cognitive preservation as the combination of both cognitive resilience and cognitive resistance. We examined 946 individuals (age 75+) at high risk of developing AD; 626 were cognitively unimpaired (CU). We performed both a genome-wide association study (GWAS) and linkage analyses across the autosomes (SNPs = 256,978) and the X chromosome (SNPs = 9,853). GWAS analyses were corrected for sex, age, and relatedness. After adjusting for multiple testing using the simpleM method, suggestive loci ($p \leq 6.4 \times 10^{-4}$) were identified on chromosomes 1, 3, 5, 6, 11, and 13-16. Linkage was performed by generating an all-connecting pedigree leading to a 14-generation 8,222-person pedigree containing all our CU individuals. To be computationally tractable, this pedigree was divided into sub-pedigrees for the linkage analyses. In non-parametric analyses, 27 suggestive ($LOD \geq 1.86$) SNPs were identified on six chromosomes in the two-point and multipoint analyses; no loci reached genome-wide significance. In the parametric analyses, under the dominant model, 72 SNPs reached significance ($LOD \geq 3.3$) across 19 chromosomes in either two-point or multipoint analyses, while 44 loci were significant in both analyses on seven chromosomes. One known AD risk locus, RHOH, was significant under the two-point dominant model (suggestive in multipoint). Under the recessive model, 37 SNPs were significant across 12 chromosomes for either two-point or multipoint analyses, while 30 loci were identified as significant by both analyses on chromosomes 2, 7, and 9, including the known AD gene CNTNAP2 on chromosome 7. Of interest, one locus on chromosome 1 (~165-167Mb, GRCh38) was detected by both the GWAS and recessive parametric linkage analysis. This locus is upstream of POGK, which has shown to have enriched expression in the brain, but the protein function remains elusive. The ability to use both association and linkage in this unique kindred may help illuminate new pathways involved with cognitive preservation, as we have found both known AD risk loci, as well as a plethora of new loci to investigate.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3493. Longitudinal, multi-modal EHR prediction of incident CKD and ESRD for risk stratification and genetic studies

Authors:

N. Banerjee¹, X. Ding², A. Averitt³, J. Mower¹, M. Haas¹, M. Cantor¹, GHS-RGC DiscovEHR Collaboration, Regeneron Genetics Center; ¹Regeneron Genetics Ctr., Tarrytown, NY, ²Univ. of Washington, Seattle, WA, ³Regeneron Genetics Ctr., New York, NY

Abstract Body:

Rich longitudinal phenotype data from electronic health records (EHR) provides great opportunities for genetic research, especially for chronic conditions such as chronic kidney disease (CKD). Traditional phenotyping algorithms can generate high-quality case and control labels for specific diseases, but they often require expert knowledge, do not consider long range dependencies and ignore that diseases reside on a continuum.

Deep sequence models and attention mechanisms that consider subtleties of EHR (e.g., the irregularity of inter-visit intervals, and the temporal order of events) have recently seen success in predicting incident disease risk. Using temporal, multi-modal EHR data of 1.9 million patients from Geisinger Health Systems (GHS), we applied BEHRT, a transformer-based deep learning risk model to predict future disease diagnosis. In this work, we have two main objectives: (1) derive quantitative clinical risk scores (CRS) for a given disease given a unified representation of EHR diagnosis, lab and medication data, and (2) perform genetic association studies on the CRSs in 175K GHS patients to facilitate target identification and genetic studies. We illustrate the benefits of the model using CKD and end stage renal disease (ESRD) which often suffers from under-diagnosis. We assessed the model's performance and results show that the model makes incident disease risk predictions with relatively high AUROC/AUPRC in CKD (0.93/0.68) and ESRD (0.97/0.58). Individuals with risk factors for CKD (e.g., diabetes, hypertension and heart failure) are enriched in the higher deciles of CKD and ESRD CRSs. CKD CRS is most correlated with CRSs of cardiometabolic diseases (comorbidities) when compared to ~100 disease CRSs. Preliminary GWAS analysis of the CKD and ESRD CRS showed significant rare variant hits in many previously characterized genes associated with kidney disease (e.g., *PKD1*, *PKD2*). Interestingly, CRS GWAS results showed ~2-fold increase in significant variant hits compared to related binary phenotypes, some of which were in genes with pleiotropic effects (*SH2B3*, *FTO*). These significant CRS hits were directionally consistent with the binary phenotype GWAS. Low genetic correlation (-0.3) between eGFR and CKD CRS suggests that CRS might be finding medical factors contributing to CKD independent of eGFR. Using ESRD CRS we identified a locus ($p < 5E-08$) in *PARP8* involved in type 2 diabetes and a locus in *HCN1* involved in kidney disease. Our findings highlight the utility of temporal, multi-modal EHR data and deep learning approaches to reappraise factors contributing to disease risk, improve phenotype development and facilitate genetic studies.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3494*. Low-coverage sequencing imputation with 150,119 UK Biobank reference samples

Authors:

S. Rubinacci^{1,2}, R. Hofmeister^{1,2}, B. Sousa da Mota^{1,2}, O. Delaneau^{1,2}; ¹Univ. of Lausanne, Lausanne, Switzerland, ²Swiss Inst. of Bioinformatics, Lausanne, Switzerland

Abstract Body:

Low-coverage whole genome sequencing (lc-WGS) followed by genotype imputation is a cost-effective technology that can enhance genome-wide analysis and access to personalized medicine. Recent studies show that a shift from SNP arrays to lc-WGS is beneficial, although **with prohibitive computational costs for large reference panels**. Therefore, sequencing projects like the recent UK Biobank (UKB) WGS, bring improved reference panels, but also new computational challenges. We present GLIMPSE2, **a new lc-WGS imputation method**, designed to handle millions of reference samples with an accuracy inaccessible to other methods at sustainable costs. **Using 150,119 UKB samples as a reference panel**, we obtain a reduction of running time of 100-1000 times compared to existing methods, **achieving imputation of a genome for less than 0.10\$**.

Lc-WGS greatly benefits from the UKB reference panel, mainly for very low depth (<0.5x) and at rare variants (<0.1% MAF). For 0.25x and 0.5x, our method imputes variants at 0.01% MAF with $r^2=0.8$ and $r^2=0.89$ respectively. As a comparison, imputation using the UKB Axiom array reaches an accuracy of $r^2=0.71$. By looking at non-homozygous reference calls, we show that **imputation increases “in-silico” the coverage of lc-WGS samples by an order of 10x**, resulting in a striking reduction of sequencing costs.

We evaluated the performance of the UKB reference panel against 1000 Genomes by imputing 276 genomes from the Simons Genome Diversity Project. As expected, the UKB reference panel **substantially improves the imputation of European individuals**, especially those with Northern European ancestry. As a proof of concept, we imputed four ancient Viking samples and show that UKB reference panel brings a significant boost of accuracy (for 0.1x, 1000G $r^2=0.83$; UKB $r^2=0.92$), **implying that Viking ancestry is recovered remarkably well from the UKB**.

Furthermore, in order to **quantify the impact of these results for disease association**, we performed GWAS using 10,000 UKB individuals across 100 quantitative traits, comparing lc-WGS and SNP array to high coverage data. **We found that the UKB Axiom array is quantitatively similar to 0.25x data** (beta $r^2=0.9$, p val $r^2=0.87$), **and inferior to 1x data** (beta $r^2=0.97$, p val $r^2=0.95$). Looking at precision and recall of the hits, **0.5x captures high coverage GWAS signals better than the UKB Axiom array**, suggesting it is better suited for fine mapping. These results are also confirmed in a **polygenic risk score** analysis.

Overall, we demonstrate the remarkable performance of lc-WGS imputation from the UKB reference panel with a competitive financial cost, that could lead to a boost of sample size and signal of future genomic studies.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3495. Low-coverage sequencing induces uncertainty in polygenic scoring

Authors:

E. Petter¹, Y. Ding¹, K. Hou¹, A. Gusev², N. Zaitlen¹, B. Pasaniuc¹; ¹UCLA, Los Angeles, CA, ²Dana-Farber Cancer Inst., Boston, MA

Abstract Body:

Polygenic scores (PGS) have emerged as the standard approach to predict phenotypes from genotype data in a wide array of applications in personalized medicine. PGS have the potential to identify patients at high risk of disease for personalized monitoring and treatment. The standard approach for PGS computation estimates risk as a linear combination of alleles that a given individual carries weighted by their effect size. Most PGS applications assume the genotype data to be error-free, ignoring possible uncertainties introduced by sequencing or imputation. Genotyping error has the potential to bias the PGS accuracy and implementation. This carries implications for clinical use of PGS, particularly for applications in which individuals in the target population are sequenced at different sequencing-depth (as coverage impacts error rate) as well as for applications in which individuals in the target data are genotyped using different genotyping arrays.

In this work, we investigate the impact of genotyping error on PGS performance, and present a procedure for estimating PGS uncertainty due to genotyping error in low-coverage whole-genome sequencing (lcWGS) data. Our approach probabilistically samples genotypes based on their allele count posterior probabilities, and approximated the distribution of possible PGS for every given individual. The variance of the individual PGS distribution quantifies the uncertainty level of the PGS estimate.

We leverage paired germline SNP array data and lcWGS (median coverage 0.1x) of 802 patients from the Dana-Farber PROFILE cohort to investigate the impact of coverage on PGS uncertainty and accuracy.

We found that even in this cohort, which employed the same sequencing technology and similar sequencing depths for all individuals, there is a strong correlation between the individual coverage level with the PGS uncertainty ($P < 2.2e-16$) and the accuracy of the PGS estimate compared to an array-based PGS of the same individuals ($P < 1.2e-7$). We propose a probabilistic procedure for risk stratification that incorporates the individual level of uncertainty in the PGS estimation, and show that in simulations it improves classification precision by up to 6% compared to classification that does not account for PGS uncertainty.

Our results illustrate the importance of considering PGS uncertainty due to genotyping error, especially in cohorts with varying sequencing depths. We show PGS uncertainty can be quantified and used to improve risk stratification in a manner that could impact patient care.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3496. LT-Free: A novel method for leveraging family history in genetic association studies of arbitrarily complex diseases.

Authors:

J. Matthews, M. Thompson, N. Zaitlen; Univ. of California, Los Angeles, Los Angeles, CA

Abstract Body:

Rich phenotypes collected in biobanks have provided opportunities to improve genetic association study power via alternative analysis methods (1). For example, the LT-FH method leveraged family history data to estimate genetic liability for a given trait, thereby improving study power (2). However, this method imposes assumptions about underlying trait genetic architectures and requires external estimates of heritability and generational prevalence. Here we introduce LT-Free as a freer and interpretable model that does not require external information and does not assume any underlying model of genetic architecture. Briefly, LT-Free allows an arbitrary latent phenotype value for each combination of case-control status, sex, and family history data observed in the data. LT-Free then alters these phenotypes to optimize association test statistic values at SNPs known to be associated with the target phenotype. In data simulated under the LT-FH model, we find that LT-Free needs around 50 known GWAS SNPs to train on before it achieves comparable association statistics at held-out SNPs compared to LT-FH. Further, once the simulated genetic architecture deviates from the LT-FH model, fewer SNPs are needed. For example, under a strong sex effect, as few as 5 SNPs are needed before LT-Free obtains higher association statistics than LT-FH. We then apply LT-Free to 9 diseases in the UKBB and similarly meet or exceed the performance of LT-FH in 8 of the 9 diseases with significant improvement in 6 diseases: heart disease ($p = 6.2e-5$), COPD ($p = 1.2e-6$), bowel cancer ($p = 7.0e-5$), depression ($p = 2.5e-4$), hypertension ($p = 9.4e-4$), and lung cancer ($p = 2.2e-4$). In the one disease in which LT-Free did not perform as well as LT-FH, type II diabetes, the difference in association statistics between methods was not significant. In heart disease, we observe evidence of different latent phenotypes estimated across sexes and parental origin. We then run our method on 5 psychiatric phenotypes from the Denmark iPSYCH project. For all target phenotypes, we use our latent phenotypes as outcomes in GWAS to find more associations. Finally, we introduce a method of using LT-Free to obtain disease specific prevalence and heritability estimates without the need of an outside study, such as a twin study. Family history data is a powerful way to improve genetic association studies. By modeling this data with more general models, we can both improve study power and gain insights into genetics architectures. References 1. <https://doi.org/10.1038/ng.3975> 2. <https://doi.org/10.1038/s41588-020-0613-6>

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3497. LTPI: A generalized liability threshold model combining dichotomous and continuous phenotypes in EHR increases the association power

Authors:

C. Lee, A. Khan, W. Chen, K. Kiryluk, I. Ionita-Laza; Columbia Univ., New York, NY

Abstract Body:

Electronic health record (EHR) data can provide valuable information for case-control association studies. However, it is unclear how best to combine case-control status with individual health records. Recently, an association test based on posterior mean genetic liability conditional on the disease status and family history has been proposed (LTFH) (Hujoel et al., 2020). Despite its importance, the lack of family history data often limits its application. We propose a model that can additionally leverage rich phenotypic data (including dichotomous and continuous traits) provided by the EHR data. We have developed an association framework (LTPI) based on posterior mean genetic liability and maximum likelihood estimator under a liability threshold model conditional on the case-control status of the disease or phenotype of the interest and diverse EHR phenotypic features. Through simulations we show increased statistical power of LTPI relative to LTFH, and further show applications to several diseases in the UKB data.

We validated the performance of LTPI by simulating 100,000 genotypes and 100,000 phenotypes for parent-offspring trios and conducting association analyses using LTPI, LTFH, and the conventional GWAS methods (case-control). We confirmed that all three methods control FPR under the null. We assumed random effects of 500 causal SNPs for five dichotomous and continuous traits and generated liability and disease case-control status of the parent-offspring trios. We fixed heritability (0.5), prevalence (0.05), and genetic correlations (0.1) of the traits. The cross-validation analysis of 10 simulations showed that the statistical power of LTPI, 37.66% (1.49%; standard error), was more significant than LTFH, 33.08% (2.2%), by 13.84% and GWAS, 25.86% (1.64%), by 45.63%. We also combined LTPI and LTFH using the Cauchy combination method. The power of the combined method was 47.72% (2.21%), which is 44.27% greater than LTFH and 84.53% greater than the GWAS.

Reference

Hujoel, M. L. A., Gazal, S., Loh, P. R., Patterson, N., & Price, A. L. (2020). Liability threshold modeling of case-control status and family history of disease increases association power. *Nature Genetics*, 52(5), 541-+.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3498. Machine learning approach for identifying diabetes nephropathy and non-diabetic chronic kidney disease cases and controls using urinary high-throughput transcriptomic data.

Authors:

M. Traglia¹, W. Ruan², R. Y. Liu³, R. Thomas¹, A. Wu²; ¹Gladstone Inst.s, San Francisco, CA, ²Univ. of California San Francisco, San Francisco, CA, ³Impetus Bioscientific Inc, Millbrae, CA

Abstract Body:

Diabetic nephropathy (DN) is a common kidney complication, which affects 20-25% of the diabetic population and the leading cause of renal failure. Timely diagnosis and treatment are critical but biased by confounding factors including non-diabetes related chronic kidney disease (CKD). Invasive kidney biopsy, though considered as the golden standard, is not favored due to the 2% mortality rate, leaving DN under and misdiagnosed. Here, we modeled a non-invasive machine learning classifiers that distinguish DN, CKD and control subjects leveraging high-throughput transcriptomic data obtained from urine samples of 1,941 subjects.

We implemented a supervised learning pipeline using random Forests (randomForestSRC R package) that uses ethnicity, age, sex, BMI, retinopathy, albumin-to-creatinine ratio, eGFR, in addition to expression levels of genes assayed with bulk RNA-seq in the urine sediments of the subjects. We applied feature selection using Recursive Feature Extraction (msvmRFE [Duan et al., 2005]) on the clinical variables and the top 200 differentially expressed genes from three contrasts (DN-controls, CKD-controls and DN-CKD). We trained three pairwise classifiers on 70% of the datasets (480 DN, 427 CKD and 451 controls) using the top 100 ranked features.

The three classifiers performed with strong predictive value and sensitivity in the remaining 30% of the dataset (DN vs NEG: sens=1, spec=0.98, AUC ROC=0.99; CKD vs NEG: sens=0.96, spec=0.94, AUC ROC=0.95; DN vs CKD: sens=0.93, spec=0.89, AUC ROC=0.91).

The ongoing collection of 200 additional subjects will allow the development of a unique classification model which may improve early detection and treatment to prevent end-stage kidney diseases.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3499. Machine Learning Approaches to Genome-wide association studies

Authors:

D. Enoma; Covenant Univ., Ota, Nigeria

Abstract Body:

Genome-wide Association Studies (GWAS) are conducted to identify single nucleotide polymorphisms (variants) associated with a phenotype within a specific population. These variants associated with diseases have a complex molecular aetiology with which they cause the disease phenotype. The genotyping data generated from subjects of study is of high dimensionality, which is a challenge. The problem is that the dataset has a large number of features and a relatively smaller sample size. However, statistical testing is the standard approach being applied to identify these variants that influence the phenotype of interest. The wide applications and abilities of Machine Learning (ML) algorithms promise to understand the effects of these variants better. The aim of this work is to discuss the applications and future trends of ML algorithms in GWAS towards understanding the effects of population genetic variant. It was discovered that algorithms such as classification, regression, ensemble, and neural networks have been applied to GWAS for which this work has further discussed comprehensively including their application areas. The ML algorithms have been applied to the identification of significant single nucleotide polymorphisms (SNP), disease risk assessment & prediction, detection of epistatic non-linear interaction, and integrated with other omics sets. This comprehensive review has highlighted these areas of application and sheds light on the promise of innovating machine learning algorithms into the computational and statistical pipeline of genome-wide association studies. This will be beneficial for better understanding of how variants are affected by disease biology and how the same variants can influence risk by developing a particular phenotype for favourable natural selection.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3500. Machine learning based methods for predicting guide RNA effects on cell fitness and gene expression in CRISPR epigenomic experiments

Authors:

T. Luo¹, W. Mu¹, A. Barrera^{2,3}, L. Bounds⁴, T. S. Klann^{2,3,4}, J. Bryois⁵, G. E. Crawford^{2,3,6}, P. F. Sullivan^{7,8,9}, C. A. Gersbach^{2,3,4}, M. I. Love^{1,8}, Y. Li^{1,8}; ¹Dept. of Biostatistics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ²Ctr. for Genomic and Computational Biology, Duke Univ., Durham, NC, ³Ctr. for Advanced Genomic Technologies, Duke Univ., Durham, NC, ⁴Dept. of BioMed. Engineering, Duke Univ., Durham, NC, ⁵Roche, Basel, Switzerland, ⁶Dept. of Pediatrics, Div. of Med. Genetics, Duke Univ. Med. Ctr., Durham, NC, ⁷Dept. of Med. Epidemiology and Biostatistics, Karolinska Inst.t, Stockholm, Sweden, ⁸Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁹Dept. of Psychiatry, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract Body:

Noncoding cis regulatory elements (CREs) are important in controlling gene expression and governing biological processes. Despite the large number of putative CREs reported, we are still far from comprehensive understanding of the functional regulatory mechanisms that drive complex phenotypes at molecular and cellular levels. CRISPR-Cas9 system and its variants have proven to be powerful tools for performing genetic or epigenetic perturbations. A critical part of a successful CRISPR experiment is the selection of highly efficient guide RNAs (gRNA).

Here we leveraged machine learning methods to predict the impact of gRNAs. Specifically, we utilized data from a whole-genome CRISPR-dCAS9-based epigenomic regulatory element screening (wgCERES), involving >1 million gRNAs targeting >100,000 candidate CREs. We leveraged two methods, XGBoost and Convolutional Neural Networks (CNN), to predict gRNA effects on overall cell fitness (cell survival and proliferation) in K562 cells, and also on expression of nearby genes across multiple cell lines including K562, induced pluripotent stem cells and neural progenitor cells. Although several machine learning models have been developed for gRNA efficiency prediction in CRISPR genetic perturbations, to our best knowledge, ours is the first attempt in the context of CRISPR epigenetic experiments.

We used both gRNA sequence and functional annotations (containing information regarding thermodynamic properties, epigenetic marks, gene essentiality characteristics) as inputs to the prediction models. We built separate models for gRNAs targeting promoter and enhancer regions. For predicting whether a gRNA has a significant effect on cell fitness, we achieved an AUC of 0.817 for promoter regions and 0.723 for enhancer regions on independent test data. For gRNAs having significant effects, we further predicted the effect size (measured by log fold-change) and achieved a Spearman correlation of 0.520 for promoters and 0.342 for enhancers. To identify and rank top features in predicting gRNA effect, we used the Shapley Additive exPlanations (SHAP) method to investigate feature importance in our models; gRNA-DNA hybridization free energy was the most influential feature overall, while epigenetic marks such as H3K27ac and H3K4me3 were critical for gRNAs targeting promoters, whereas gRNA sequence information was critical for enhancer regions. We additionally attempted to predict gRNA effects on gene expression, quantified by single cell RNA-seq, in both whole genome and MHC regions and compared with activity-by-contact model.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3501. Machine learning guided transcriptomic analyses reveal a novel 9-gene module associated with schizophrenia risk

Authors:

F. Guan¹, C. Shen¹, L. Zhu¹, J. Xiao¹, H. Wang¹, W. Zhang¹, Y. Ma¹, H. Gui²; ¹Xi'an Jiaotong Univ., Xi'an, China, ²Henry Ford Hlth.System, Detroit, MI

Abstract Body:

Introduction: Schizophrenia (SCZ) is one of the most severe psychiatric disorders. Machine learning techniques are considered powerful tools for precision medicine in medical applications. However, there are still lacking studies that apply machine learning approaches to transcriptomic or other omics data to classify SCZ.

Methods: RNA-seq data was downloaded from the PsychEncode synapse portal. We integrated traditional bioinformatics analyses, e.g. differentially expressed gene analysis and enrichment pathway analysis, and feature selection with machine learning to mine risk genes that are highly correlated with SCZ. We first systematically characterized risk genes in postmortem dorsolateral prefrontal cortex (DLPFC) tissue of controls and SCZ patients using the available biological background, i.e. pathways known to be highly associated with the disease. Then, we further screened for genes closely associated with disease phenotypes employing efficient feature derivation, e.g. polynomial identity derivation, and signature screening including recursive feature elimination, sequential feature selection, and mutual information methods. Through the joint analysis of multiple machine learning models, i.e. bagging machine learning and ensemble machine learning, these risk genes are then utilized to make a comprehensive assessment of SCZ. We also use polygenic risk score, transcriptome-wide association study (TWAS), and functional genomic data to annotate those genes in the final module. Finally, we verified the differential expression of the final risk biomarkers in the PFC of the SCZ mice model by quantitative real-time PCR (qRT-PCR).

Results: In total 138 different genes were selected by WGCNA and bioinformatics. Among them, a set of 9 genes was identified to best predict SCZ risk (AUC=0.9 in the training dataset and 0.81 in the validation dataset). Among them, only 5 were differentially expressed genes in original datasets. Three of them were also supported by gene-based association tests. The qRT-PCR data showed that eight out of these 9 genes have the same direction of PFC expression changes in the SCZ mouse model.

Conclusion: We develop a novel pipeline of combining traditional bioinformatics analysis with a state-of-the-art machine learning algorithm to explore SCZ omics data. The final gene module contains not only known risk SCZ genes, but also novel ones. We provide proof of principle evidence showing an integrative pipeline not only helps to identify novel risk genes but also improves the prediction power for risk prediction of SCZ.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3502. Machine learning model discriminates false from true CNVs with a 98% accuracy.

Authors:

T. Renne¹, M. Loum², M. Jean-Louis¹, Z. Saci¹, A. Labbé³, G. HUGUET¹, S. Jacquemont¹; ¹Res. Ctr. CHU Sainte Justine, Université de Montréal, Montréal, QC, Canada, ²Université Iba Der Thiam de Thiès, UFR de Sci.s de l'ingénieur, Thies, Senegal, ³Dept. of decision, HEC Montréal, Montréal, QC, Canada

Abstract Body:

Background: Genomic Copy Number Variants (CNVs) are associated with a range of human psychiatric and medical diseases. To understand these associations, CNVs are being investigated in increasingly large cohorts of hundreds of thousands of individuals, which include genotyping array data. Using multiple calling methods and after applying well accepted quality control (QC) criteria (Huguet et al 2018), datasets still include a significant proportion of false CNVs (~ 10%). For example, CNVs calling algorithms performed on good QC GSA Illumina arrays generate a large multigenic CNV in 2% of samples even after applying classic QC methods. Currently, these artifacts have important consequences and delay all downstream analysis. New automated quality control methods like DeepCNV are significant advances, but require large computing resources. Their scalability is problematic and often work only with one CNV caller output.

Knowledge gap: There is currently no automated, efficient and scalable quality control algorithm that applies to multiple CNV callers. Our aim is to develop a machine learning model to discriminate true from false CNVs identified by multiple algorithms on genotyping arrays without the necessity of having access to data on probe intensity.

Methods: We called CNVs using 2 algorithms (PennCNV and QuantiSNP) in 396,000 individuals from 8 cohorts genotyped using Affymetrix and Illumina genotyping array technologies. We manually visualized and classified as true or false approximately 35,000 CNVs (by 3 different experts). The machine learning model was trained on 70% and tested on 30% of the manually curated CNV using multiple features related to the quality of the array (CallRate, Waviness Factor, ...), as well as the characteristics of the CNVs (Probe density, likelihood score, the percentage of overlap between algorithms, ...). To balance the sample sizes of true and false CNVs, we simulated false CNVs using the "imbalanced-learn" package.

Results: Manual curation classified 3,500 of the 35,000 CNVs as false. After training and parameter tuning, the machine learning model could predict CNV class (true vs false) with 98% accuracy. The model classified 99% of true CNVs as true and 80% of false CNVs as false. A pre-trained model can be used directly with new data instantly with no additional training. It was designed to be used with any type of array technology thanks to the feature "Number of probes by technology".

Conclusion: This novel tool, DigCNV, accelerates significantly the processing of large datasets and improves signal-to-noise, especially while analyzing rare non-recurrent CNVs.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3503. MagicalRsq: Machine learning based genotype imputation quality calibration.

Authors:

Q. Sun¹, Y. Yang², J. Rosen¹, M. Jiang¹, J. Chen¹, W. Liu¹, J. Wen¹, L. Raffield¹, R. Pace¹, Y-H. Zhou³, F. Wright³, S. Blackman⁴, M. Bamshad⁵, R. Gibson⁵, G. Cutting⁴, M. Knowles¹, D. Schrider¹, C. Fuchsberger⁶, Y. Li¹; ¹Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ²Yale Univ., New Haven, CT, ³North Carolina State Univ., Raleigh, NC, ⁴Johns Hopkins Univ., Baltimore, MD, ⁵Univ. of Washington, Seattle, WA, ⁶EURAC, Bolzano, Italy

Abstract Body:

Whole-genome sequencing (WGS) is the gold standard for fully characterizing genetic variation but is still prohibitively expensive for large samples. To reduce costs, many studies only sequence a subset of individuals or genomic regions, and genotype imputation is used to infer genotypes for the remaining individuals or regions without sequencing data. However, not all variants can be well imputed, and more importantly, the current state-of-the-art imputation quality metric given by imputation software, denoted as standard Rsq, is poorly calibrated for lower frequency variants. Here, we propose MagicalRsq, a Machine learning based imputation quality calibration, by using the eXtreme Gradient Boosted trees method to incorporate various variant-level imputation and population genetics statistics, to provide a better calibrated imputation quality metric. Leveraging WGS data from the Cystic Fibrosis Genome Project (CFGP), and whole exome sequence data from UK BioBank, we performed comprehensive experiments to evaluate the performance of MagicalRsq compared to standard Rsq for partially sequenced studies. We found that MagicalRsq aligns better with true R2 (the squared Pearson correlation between imputed dosages and true genotypes) than standard Rsq in almost every situation evaluated, for both European and African ancestry individuals. Thus, MagicalRsq can serve as an improved post-imputation quality metric. For example, when applying models trained from 1992 sequenced CFGP samples to 3103 independent samples with no sequencing data but with imputation using TOPMed freeze 8 reference panel, compared to the standard Rsq, MagicalRsq improved squared Pearson correlation with true R2 by 2.0% - 56.8% for common variants [minor allele frequency (MAF) > 5%], by 16.3% - 91.7% for low frequency variants (MAF in [0.5%, 5%]) and by 3.0% - 6.2% for rare variants (MAF < 0.5%), across different chromosomes. It also achieved net gains of 1.4 million rare variants, 117k low frequency variants and 18k common variants, where net gains were gained numbers of correctly distinguished variants by MagicalRsq over standard Rsq. Even when trained with only 10k variants, MagicalRsq still provided net gains of 787k rare variants, 94k low frequency variants, and 6k common variants, and such gains were robust to the inclusion of different sets of variants in training models. We anticipate MagicalRsq will benefit downstream analysis by better distinguishing well-imputed variants from poorly-imputed variants. MagicalRsq is freely available from <https://github.com/quansun98/MagicalRsq/>.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3504*. MANOCCA : An innovative powerful test to detect predictors of the gut microbiome variability

Authors:

C. Boetto¹, V. Basten Romero¹, E. Patin¹, M. Bredon², D. Duffy¹, L. Quintana-Murci³, H. Sokol^{4,5,6,7}, H. Aschard^{1,8}; ¹Inst. Pasteur, Paris, France, ²Sorbonne Université, Paris, France, ³Pasteur Inst. and Coll. de France, Paris, France, ⁴Hôpital Saint-Antoine, Paris, France, ⁵Inserm, Paris, France, ⁶Paris Ctr. for Microbiome Med., Paris, France, ⁷INRAE Micalis & AgroParisTech, Jouy-en-Josas, France, ⁸Harvard TH Chan Sch. of Publ. Hlth., Boston, MA

Abstract Body:

Multivariate analysis is becoming central in studies investigating high-dimensional omics data. However, some important characteristics of those data have been seldom explored. Here we present MANOCCA (Multivariate Analysis of Conditional Covariance), a powerful method to test the effect of both continuous and categorical predictors on the covariance matrix of a multivariate outcome. The proposed test is by construction orthogonal to tests based on mean (MANOVA) and variance (Levene), and can capture effects missed by both approaches. We first compared the performances of MANOCCA against existing correlation methods (BoxM, Mantel, Fisher Z-score, Jennrich) and show that the latter display severe type I error rate inflation in all simulations mimicking omics data. Only MANOCCA was correctly calibrated. As a case study, we then applied our test to assess the effect of environmental and host genetic predictors on 16S derived gut microbiome covariance matrix, in 1,000 healthy participants from the Milieu Interieur cohort while adjusting for age, sex and BMI. We systematically compared results from our test with standard univariate (linear regression) and multivariate (MANOVA) mean-based tests. MANOCCA strongly outperformed those tests, confirming associations with age and sex but with higher power (in average 200% power increase), and detected additional signals with smoking and appendectomy operation. Host genetic genome wide association study (GWAS) using MANOCCA identified multiple significant associations (P value < 1e-8). This includes a signal with *RGS3*, a gene previously reported associated with blood cell count, as well as a signal with *TCTB* which is known to be associated with cholesterol. In comparison, mean-based methods GWAS screening did not detect any signal. Altogether, these analyses demonstrate the strong capabilities of our novel approach to identify predictors associated with a multivariate outcome covariance matrix, and more generally to detect new signals that cannot be captured by standard mean-based approaches.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3505. MaSk-LMM: a matrix sketching-based fast and scalable linear mixed model for association studies in large biobanks

Authors:

M. Burch¹, A. Bose², P. Drineas¹, L. Parida²; ¹Purdue Univ., West Lafayette, IN, ²IBM T J Watson Res., Yorktown Heights, NY

Abstract Body:

Large-scale biobanks spanning hundreds of thousands of individuals offer unprecedented opportunities to discover novel genetic loci associated with complex human traits and disease risk. Linear mixed models (LMMs) have been widely used in genome-wide association studies (GWAS) to control for population stratification and relatedness. Implementation of LMMs is computationally intensive, particularly in building the genetic relatedness matrix (GRM) and computing the residual maximum likelihood (REML) estimation of genomic variance components. Conventional REML methods involve matrices that are dependent on the number of markers and individuals, thus making LMMs intractable for large biobanks. We developed a fast and efficient LMM method, MaSk-LMM (Matrix Sketching-based LMM), by applying both sample and marker sketching to reduce the dimensions of individuals and markers to speed up the estimation of REML parameters and the GRM computations without too much loss in accuracy. Matrix sketching introduces randomization to the genotype matrix by using a sketching matrix which is significantly smaller than the genotype matrix, but still preserves pairwise distances. We applied MaSk-LMM on a simulated cohort of 100,000 individuals and 600,000 genotypes imitating relatedness and fine-grained population stratification. We set 5% of the genotypes as causal or contributing to the binary or continuous trait of interest. We evaluated three sketching methods (Gaussian, random signs and fast cosine transform) and set the sketching dimensions to 1%, 5%, 10% and 15% of the number of markers and evaluated each scenario ten times to report running times and accuracy. We observed that the proportion of variance explained as the sketching dimension increases, while the false positive rate (FPR) decreases and correlates well with the genomic inflation factor. The p-value for MaSk-LMM increased relative to an exact LMM solution but it was twice as fast as current state-of-the-art (SOTA) methods such as FastGWA and BOLT-LMM. MaSk-LMM can be run in conjunction with grid search as done in Grid-LMM and FastGWA-GLMM and is “embarrassingly parallel,” permitting further gains in computational time. We also applied MaSk-LMM to real phenotypes from UKB such as height, body mass index, LDL cholesterol levels, myocardial infarction, etc. across 456K European samples and reported similar heritability measures and test statistics as in SOTA methods. Thus, sketching-based approaches can be very useful in speeding up intractable LMM computations for large biobanks and accelerate discovery of biomarkers with an efficient, fast, and accurate estimation of the variance components.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3506. MATS: A novel multi-ancestry transcriptome-wide association study to account for heterogeneity in the effects of cis-regulated gene expression on complex traits

Authors:

W. Pan, K. A. Knutson; Univ. of Minnesota, MINNEAPOLIS, MN

Abstract Body:

The Transcriptome-Wide Association Study (TWAS) is a widely used approach integrating gene expression and GWAS data to study the role of cis-regulated gene expression (GEx) in complex traits. However, the genetic architecture of GEx varies across populations, and recent findings point to possible ancestral heterogeneity in the effects of GEx on complex traits, which may be amplified in TWAS by modeling GEx as a function of cis-eQTLs. Here, we present a novel extension to TWAS to account for heterogeneity in the effects of cis-regulated GEx which are correlated with ancestry. Our proposed Multi-Ancestry TwaS (MATS) framework jointly analyzes samples from multiple populations and distinguishes between shared, ancestry-specific, and/or subject-specific expression-trait associations. As such, MATS amplifies power to detect shared GEx associations over ancestry-stratified TWAS through increased sample sizes, and facilitates detection of genes with subgroup-specific associations which may be masked by standard TWAS. Our simulations highlight the improved Type-I error conservation and power of MATS compared to competing approaches. Our real-data applications to AD case-control binary phenotypes from the Alzheimer's Disease Sequencing Project (ADSP) and quantitative phenotypes from the UK Biobank (UKBB) identify a number of unique gene-trait pairs which were not discovered through standard and/or ancestry-stratified TWAS. Ultimately, these findings promote MATS as a powerful method for detecting and estimating significant gene expression effects on complex traits within multi-ancestry cohorts, and corroborates the mounting evidence for inter-population heterogeneity in gene-trait associations.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3507. mBAT-combo: a gene-based association test to decipher masking effects

Authors:

A. Li¹, S. Liu¹, A. Bakshi², L. Jiang³, W. Chen⁴, Z. Zheng¹, P. Sullivan⁵, P. Visscher¹, N. Wray¹, J. Yang⁶, J. Zeng¹; ¹The Univ. of Queensland, Brisbane, Australia, ²Monash Univ., Melbourne, Australia, ³New York Genome Ctr., New York, NY, ⁴Garvan Inst. of Med. Res., Sydney, Australia, ⁵Univ North Carolina, Chapel Hill, NC, ⁶Westlake Univ., Hangzhou, China

Abstract Body:

Gene-based association tests aggregate multiple SNP-trait associations into sets defined by gene boundaries. Since genes have a direct biological link to downstream function, gene-based test results are widely used in post-GWAS analysis. A common approach for gene-based tests is to combine SNPs associations by computing the sum of chi-squared statistics. However, this strategy ignores the directions of SNP effects, which could result in a loss of power for SNPs with masking effects (i.e., the product of two SNP effects and their genotype correlation is negative). Here, we introduce mBAT-combo a new gene-based test that is better powered than other methods (fastBAT, MAGMA, multiple linear regression model) to detect multi-SNP associations in the context of masking effects. We validate the method through theoretical and empirical simulations and applications to real data. We find that of 35 blood and urine biomarker traits in the UK Biobank, 34 traits show evidence for masking effects in a total of 4,516 gene-trait pairs, implicating a ubiquitous presence of masking effects in complex traits. We further validate the improved power of our method in height, BMI and schizophrenia with different GWAS sample sizes, and show that on average 95.7% (sample size increased by 1.7-fold) of the genes detected only by mBAT-combo with smaller sample sizes can be identified by the single-SNP approach with larger sample sizes. As a more powerful gene-based method, mBAT-combo is expected to improve the downstream pathway or tissue and cell-type enrichment analysis that takes genes identified from GWAS data as input to understand the biological mechanisms of the trait or disease. Despite we focus on mapping genes in this study, the framework of mBAT-combo is general and can be applied to any set of SNPs to refine trait-association signals hidden in complex linkage disequilibrium.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3508. Mendelian randomization and colocalization of amino acids in the UK Biobank NMR metabolomic profiles

Authors:

C. Willis, A. Deaton, A. Holleman, L. Krohn, R. Hoffing, P. LoGerfo, M. Plekan, P. Nioi, L. Ward; Alnylam Pharmaceuticals, Cambridge, MA

Abstract Body:

Mendelian randomization (MR) studies offer a theoretically unbiased approach to estimating causal relationships between risk factors and disease outcomes. Studies commonly use protein and expression quantitative trait loci (QTL) as straightforward instruments in MR analysis; however, we sought to characterize and explore the use of metabolite QTLs (metaboQTLs) for target finding and biomarker discovery. We analyzed whole exome sequencing, NMR metabolomic profiling (Nightingale), and disease diagnosis data from the UK Biobank unrelated European ancestry population ($n = 81,112$) to identify disease risks linked to levels of amino acids. We computed LD score regression (LDSC) between 10 amino acids (alanine, creatinine, glutamine, glycine, histidine, phenylalanine, and branched-chain amino acids—leucine, isoleucine, and valine) and 68 disease outcomes revealing nominally significant genetic correlation for 76 metabolite-disease pairs. We performed colocalization and Mendelian randomization analysis on these amino acids and diseases using cis-metaboQTLs to minimize the risk of pleiotropy. Of 222 independent metaboQTLs identified, 18 were cis-acting for nine genes involved in amino acid catabolism. Nine cis-metaboQTLs in BCAT2, HAL, PAH, and SLC1A4 were colocalized ($PP > 0.5$) with at least one disease outcome. MR analysis revealed evidence supporting several causal relationships between amino acids and diseases. Leucine levels (modeled by a metaboQTL at BCAT2 encoding BCAA transaminase 2) were suggested to be an influencing risk for gallstones ($OR = 1.31$, $P = 8.87 \times 10^{-5}$), hyperlipidemia ($OR = 0.86$, $P = 1.84 \times 10^{-4}$), and hypercholesterolemia ($OR = 0.90$, $P = 1.93 \times 10^{-4}$). We also observed a potential protective effect of histidine levels (modeled by a metaboQTL at HAL encoding histidase) on skin cancer ($OR = -0.18$, $P = 5.04 \times 10^{-4}$). Expansion of this study to incorporate additional metabolomic biomarkers could offer vital insights into common disease risks as well as novel drug targets.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3509. Mendelian randomization implicates GDF15 blood levels as a causal factor in inflammatory disease.

Authors:

P. Timmers, E. Morgen, K. Fortney, Z. O'Brown, P. Leong, N. Shah; BioAge Labs, Richmond, CA

Abstract Body:

Growth/differentiation factor 15 (GDF15) blood levels are strongly associated with mortality and have been linked to myriad age-related diseases, including cardiovascular disorders, cancers, cognitive impairment, and chronic inflammation. Whether GDF15 plays a causal role in aging or simply functions as a biomarker of disease remains underexplored. Here, we combine three large genetic studies of GDF15 blood protein levels ($N = 60,486$), measured using SomaLogic and Olink technologies, and use phenome-wide colocalization and bidirectional Mendelian randomization to infer causality of GDF15 blood levels across hundreds of complex traits and disease outcomes. We reveal GWAS signals of body mass index and Cystatin C levels colocalize with the *GDF15* locus, but find genetically higher levels of these traits unidirectionally increase GDF15 protein levels and therefore act as confounding variables. We then use Bayesian GWAS to estimate direct genetic effects of GDF15 blood levels independent of these confounders, and find evidence for causal effects of GDF15 on blood levels of 12 proteins in *trans*. Gene set enrichment reveals the genes encoding these proteins are overrepresented in the hallmark IL6 JAK-STAT signaling pathway ($FDR < 0.05$). In addition, GDF15 cis-pQTL are strongly associated with the prevalence of several inflammatory diseases ($P < 1 \times 10^{-10}$), including inflammatory bowel disease and asthma, with moderate confidence for colocalization ($ColocPP > 50\%$). Finally, we find robust evidence that genetically determined GDF15 blood levels are associated with prevalence of prevotella in the gut ($P < 1 \times 10^{-60}$; $ColocPP > 80\%$), highlighting the microbiome for further study. In conclusion, we show phenome-wide bidirectional Mendelian randomization and colocalization can disentangle cause from confounding in a highly complex locus and use these methods to reveal the causal role of GDF15 blood protein levels in inflammatory disease.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3510. Mendelian randomization indicates sex-specific causal effects of estradiol levels on kidney function.

Authors:

M. Nasr^{1,2,3}, C. Schurmann⁴, E. Bottinger^{5,2}, A. Teumer^{1,3}; ¹Univ. Med. Greifswald, Greifswald, Germany, ²Hasso Plattner Inst., Potsdam, Germany, ³DZHK (German Ctr. for Cardiovascular Res.), Partner Site Greifswald, Greifswald, Germany, ⁴Bayer AG, Pharmaceuticals, Berlin, Germany, ⁵Mount Sinai Sch. of Med., New York, NY

Abstract Body:

Chronic kidney disease (CKD) is a public health burden and a leading cause of major cardiovascular disorders as well as a contributor to morbidity and mortality worldwide. Previous research has attempted to identify sexual disparity in CKD observed through sex-associated differences in exposure and prognosis of kidney disease. Epidemiological studies observed an association between sex hormones, including estradiol, and kidney function that may contribute to the different prevalence of CKD in men and women. However, it is difficult to assess the causality of this relationship because it requires randomized controlled trials, which are lengthy and costly. As an alternative approach, we conducted a Mendelian randomization (MR) study to test for a possible causal effect of estradiol on kidney function using observational data and genetic information. We performed unidirectional two-sample MR using published genetic associations results of serum estradiol levels in men (n=4,191 and n=206,927) and women (n=2,607), and of kidney function traits represented by estimated glomerular filtration rate (eGFR, n=567,460), urine albumin-to-creatinine ratio (UACR, n=547,361), and CKD (n=41,395 cases and n=439,303 controls) obtained from the CKDGen Consortium. Additionally, we applied linear mixed models to genome-wide individual-level data from the UK Biobank (n=11,798 men and n=6,835 women) to identify novel genetic associations with estradiol, and subsequently used these variants as instruments in a one-sample MR adjusted for age, body mass index (BMI) and diabetes status. Two-sample MR indicated that genetically predicted estradiol levels are significantly associated with eGFR in women (beta = -0.005, p-value = 0.015). Through the genome-wide association study with UK Biobank data, we identified a single locus at chromosome 14 associated with estradiol levels in men. Using the index variant of this locus (rs7151019, beta = -0.026, p-value = 6E-22) as an instrument in the one-sample MR in 11,798 men with both estradiol and kidney function data available revealed a significant association with eGFR (beta = 0.199, p-value = 0.017). There were no significant causal estimates for UACR or CKD. Based on these results, we conclude that serum estradiol levels may have a causal effect on kidney function characterized by eGFR. Further efforts are needed to identify a strong and reproducible genetic association with estradiol levels in both genders. However, our MR results provide starting points for subsequent studies to develop therapeutic and intervention strategies to reduce kidney diseases and related disorders.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3511*. MENDEL-modified segregation analysis of 168 *TP53*-positive families estimates age-specific risks for cancer types beyond the established spectrum for Li-Fraumeni syndrome.

Authors:

C. Fortuno¹, B-J. Feng², C. Carroll², W. Kohlman², C. Lazaro³, L. Feliubadalo³, M. Menendez³, J. Brunet³, M. Ballinger⁴, D. Thomas⁴, A. Campbell⁵, M. Field⁶, M. Harris⁷, J. Kirk⁸, N. Pachter⁹, N. Poplawski¹⁰, R. Susman¹¹, K. Tucker^{12,13}, M. Wallis^{14,15}, R. Williams^{13,16}, E. Cops¹⁷, D. Goldgar², P. James^{17,18}, A. Spurdle¹; ¹Genetics and Computational Biology, QIMR Berghofer Med. Res. Inst., Brisbane, Queensland, Australia, ²Univ. of Utah, Salt Lake City, UT, ³Hereditary Cancer Program, Catalan Inst. of Oncology, IDIBELL, Hospitalet de Llobregat, Barcelona, Spain, ⁴Garvan Inst. of Med. Res., Darlinghurst, New South Wales, Australia, ⁵Dept. of Clinical Genetics, Austin Hlth., Melbourne, Victoria, Australia, ⁶Familial Cancer Service, Royal North Shore Hosp., St Leonards, New South Wales, Australia, ⁷Monash Hlth.Familial Cancer Service, Melbourne, Victoria, Australia, ⁸Familial Cancer Service, Crown Princess Mary Cancer Ctr., Westmead Hosp., Westmead, New South Wales, Australia, ⁹Genetic Services of Western Australia, King Edward Mem. Hosp., Perth, Western Australia, Australia, ¹⁰Adult Genetics Unit, Royal Adelaide Hosp., Adelaide, South Australia, Australia, ¹¹Genetic Hlth. Queensland, Royal Brisbane and Women's Hosp., Brisbane, Queensland, Australia, ¹²Hereditary Cancer Clinic, Prince of Wales Hosp., Randwick, New South Wales, Australia, ¹³Prince of Wales Clinical Sch., UNSW Med. and Hlth., UNSW Sydney, Sydney, New South Wales, Australia, ¹⁴Tasmanian Clinical Genetics Service, Tasmanian Hlth. Service, Royal Hobart Hosp., Hobart, Tasmania, Australia, ¹⁵Sch. of Med. and Menzies Inst. for Med. Res., Univ. of Tasmania, Hobart, Tasmania, Australia, ¹⁶Prince of Wales Hereditary Cancer Ctr., Prince of Wales Hosp., Randwick, New South Wales, Australia, ¹⁷Parkville Familial Cancer Ctr., Peter MacCallum Cancer Ctr. and Royal Melbourne Hosp., Melbourne, Victoria, Australia, ¹⁸Sir Peter MacCallum Dept. of Oncology, Univ. of Melbourne, Melbourne, Victoria, Australia

Abstract Body:

Establishing accurate age-related penetrance figures for the broad range of cancer-types that occur in individuals harbouring a pathogenic variant in *TP53* is essential to determine the most effective risk management strategies in clinical practice. These figures are also needed to allow co-segregation data to be used optimally in the classification of *TP53* variants of unknown significance. However, penetrance estimation can easily be affected by ascertainment, particularly where families have been selected for testing in the past based on Li-Fraumeni syndrome (LFS) clinical criteria.

We performed a maximum likelihood penetrance estimation using full pedigree data from an international multi-centre study of *TP53* positive families identified through dedicated clinics and research studies, incorporating the ability to adjust for the effect of ascertainment and population-specific background cancer risks. The analysis included 163 unique pedigrees from Australia, Spain, and USA, with phenotypic information for 4,899 individuals. In addition to the expected strong associations with classical LFS cancers (breast, osteosarcoma, soft tissue sarcoma, brain cancer, and adrenocortical carcinoma), the analysis also detected a significantly increased lifetime risk for colorectal, lung, and ovarian cancer. Current results show a cumulative risk of any cancer type by age 50 of 44.2% (95% CI 41.9 - 46.3) for males, and 96.6% (95% CI 95.8 - 97.2) for females, higher due to the incidence of breast cancer.

Our findings demonstrate the utility of a MENDEL-modified segregation approach to generate new age-specific cancer risk estimates for *TP53* pathogenic variant carriers. The results confirm the known high lifetime risk for the classical-LFS associated cancer types, and indicate significantly increased lifetime risks for several additional cancer-types where the association with LFS has not been previously established. Accurate cancer risk estimates will help with the refinement of management

recommendations for *TP53* pathogenic variant carriers, and can also be incorporated into *TP53* variant classification modelling.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3512. Metabolite prediction models in UK Biobank: A metabolome-wide association study (MWAS)

Authors:

L. Huang¹, A. Tapia², B. Rowland³, J. Rosen^{3,4}, J. Wen⁴, Y. Li^{3,4,5}, L. Raffield⁴; ¹Curriculum in Bioinformatics and Computational Biology, UNC - Chapel Hill, Chapel Hill, NC, ²Dept. of Psychiatry, Univ. of Pittsburgh, Pittsburgh, PA, ³Dept. of Biostatistics, UNC - Chapel Hill, Chapel Hill, NC, ⁴Dept. of Genetics, UNC - Chapel Hill, Chapel Hill, NC, ⁵Dept. of Computer Sci., UNC - Chapel Hill, Chapel Hill, NC

Abstract Body:

As potential disease risk factors, multi-omics readouts including transcripts, proteins, and metabolites are all of interest. High-throughput multi-omics profiling has been recently applied to various cohorts, discovering new disease risk factors and molecular mechanisms. However, due to cost and tissue availability, sample sizes for omics data have severely lagged behind genetic data. To fill this gap, we use genetic and metabolomic data from UK Biobank (UKBB) to perform a metabolome-wide association study (MWAS). Our initial analyses included 101,349 European ancestry participants (EUR) with 161 post-quality control metabolites measured using the Nightingale platform. We identified metabolite quantitative trait loci (mQTL) with a subset of 45,581 individuals, built metabolite prediction elastic net models with these mQTLs ($p < 1e-6$) in an independent subset of 45,466 individuals, and assessed the performance of the prediction models in the remaining 10,302 individuals. Our results showed that a number of metabolites can be reasonably predicted: median testing R^2 0.08, but maximum R^2 0.16, largely consistent with SNP heritability estimates for these metabolites (median H^2 0.18 and maximum H^2 0.24). We then proceeded with our MWAS for the 134 metabolites (testing $R^2 > 0.05$) and tested associations with 29 blood cell traits in the 349,956 EUR UKB samples without measured metabolites. We also tested the associations between measured metabolites and blood cell traits, using the 101,349 samples with measured metabolites. Using either measured or predicted metabolites, we found numerous significant associations. At a Bonferroni threshold of $1.3e-5$, we had an average of 114 (range: 13-133) significant associations per blood cell trait using measured metabolites. Similarly, we had an average of 73 (range: 0-129) significant associations per trait using predicted metabolites. The findings are largely expected because most of the Nightingale metabolites are lipid-related, and lipid traits have well-established relationships with blood cell indices. Association results are reasonably consistent between measured and predicted: 1241 (37.5%) of associations are significant and in the same direction using measured and predicted metabolites. We plan to perform PheWAS with the predicted metabolites, which we anticipate will reveal signals missed by using the much smaller sample size measured metabolites. Although much smaller sample sizes are available, we will also explore MWAS approaches in UKB non-EUR samples. In summary, MWAS in biobank samples can be a powerful approach to reveal molecular mechanisms and discover novel therapeutic targets.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3513. Meta-imputation combining the East Asian and multi-ethnic reference panels enhanced the imputation performance of rare variants for East Asian population

Authors:

Y. Kim¹, M. HWANG¹, N-H. Choi¹, D. Shin¹, K. Yoon¹, H-Y. Park², B-J. Kim¹; ¹Div. of Genome Sci., NIH, Cheongju-si, Korea, Republic of, ²Dept. of Precision Med., NIH, Cheongju-si, Korea, Republic of

Abstract Body:

Genotype imputation has become an essential analysis to enhance association mapping power and discover rare and structural variants which are hardly obtained by most genotyping arrays. Imputation analysis can be more accurate with a population specific reference panel or a multi-ethnic reference panel with a large number of samples. A recent population specific reference panel, the China Metabolic Analytics Project (ChinaMAP, N=10,155), emphasized an importance of a population specific reference panel by showing increased imputation accuracy over the Trans-Omics for Precision Medicine (TOPMed, N=97,256) reference panel, which has about 10 times more multi-ethnic samples. However, it is not feasible to construct a population specific reference panel with a large number of sequenced samples. Alternatively, a meta-imputation approach was developed to comprehensively increase imputation performance by merging imputation results from difference reference panels without assessing individual-level genotype data of the reference panels. Therefore, meta-imputation combining a population specific and multi-ethnic reference panels would be an efficient strategy for accurate imputation of genotypes. In this study, we constructed an East Asian reference panel, the Korean Reference Genome (KRG) panel, from 1,490 whole genome sequenced Korean samples. We analyzed the imputation performance of KRG and further studied an increase in imputation accuracy via meta-imputation. Sequenced reads were aligned and genotypes were called using DRAGENTM. For comparison analysis, three genotype panels of Korean (KOR), Japanese (JPT), and Chinese (CH) were created by selecting variants available in the Korea Biobank Array from 208 whole genome sequenced samples which were not used for the KRG, 104 Japanese and 208 Chinese samples from 1KGP3, respectively. Imputation performance of KRG was compared with previously introduced reference panels of 1,000 Genomes project phase 3 (1KGP3), GenomeAsia Sequencing project (GASP), ChinaMAP, TOPMed, and meta-imputation results using MetaMinimac2 combining KRG and TOPMed (META). For KOR and JPT, KRG showed increased imputation performance (aggregate $R^2 > 0.88$) in overall compared to those of other reference panels ($R^2 < 0.86$). As expected, ChinaMAP showed the best performance for CH. However, for rare variants (Minor Allele Frequency < 1%), META showed enhanced imputation accuracy than other reference panels. This study demonstrated the importance of a population specific reference panel and meta-imputation combining difference reference panels for assessing substantial number of accurately imputed rare variants.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3514. Methods and software for empirical Bayes multivariate multipletesting and effect size estimation

Authors:

Y. Yang¹, P. Carbonetto¹, D. Xie¹, G. David², M. Stephens¹; ¹Univ. of Chicago, Chicago, IL, ²American Univ., Washington D.C., DC

Abstract Body:

The problem of estimating how genetic effects are shared across different conditions or different treatments is an important aspect of many genomic analyses, such as the joint analysis of expression QTL (eQTLs) in multiple tissues. The patterns of sharing in these analyses are often very heterogeneous; for example, different eQTLs may act in different combinations of tissues via distinct mechanisms. To flexibly model these heterogeneous sharing patterns, Urbut et al (2019) recently proposed the multivariate adaptive shrinkage (mash) method to jointly analyze genetic effects across multiple conditions/treatments. However, the question of how to learn these heterogeneous patterns of sharing from data remains unresolved, so here we propose new statistical methods to tackle this question. Our new method is called “Ultimate Deconvolution” because we developed new algorithms as well as implemented the existing Extreme Deconvolution (ED) algorithm of Bovy et al (2011). Compared with ED, UD has two key advantages. First, our new algorithms find better estimates of sharing patterns, and more efficiently, these methods are easy to be applied to large genomic data sets. Second, we address deficiencies of maximum-likelihood estimation in more challenging high-dimensional settings by implementing adaptive regularization for the sharing patterns and cross-validation to control for model complexity. The combination of adaptive regularization, cross-validation, and better algorithms leads to better estimates of sharing in high-dimensional settings with many conditions, and therefore improved accuracy to detect the true underlying effects. We have quantified the benefits of these improvements extensively in simulated genetic data sets. We also illustrate the benefits of these new methods through an analysis of eQTLs in 49 human tissues, using data from the Genotype Tissue Expression (GTEx) project. We have implemented these new methods in an R package, `udr` (“Ultimate Deconvolution in R”), available on GitHub.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3515. MethylPipeR: Multi-omic Prediction of Complex Traits.

Authors:

Y. Cheng; Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract Body:

DNA methylation (DNAm)-based predictors have been developed for a wide range of traits and have shown promise in improving disease prediction. Whilst previous studies have been largely constrained by linear assumptions and the use of CpGs one-at-a-time, we have adopted a more flexible approach based on a range of linear and tree-ensemble survival models for incident disease prediction. Using the Generation Scotland cohort (training $n_{\text{cases}}=374$, $n_{\text{controls}}=9,461$; test set $n_{\text{cases}}=255$, $n_{\text{controls}}=4,546$) we show that the use of CpG DNAm information leads to a significant improvement in the prediction of 10-year T2D incidence (ascertained through linkage to electronic health records) over and above common risk factors ($p=4.2 \times 10^{-12}$). We replicated this finding in an external test dataset (the German-based KORA study). For systematic development of complex trait and incident disease predictors, we have also developed MethylPipeR, an R package with accompanying user interface. MethylPipeR is designed with a focus on reproducibility and is applicable to a wide range of omics data and target traits.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3516. Metric projection—AUCs, correlations, and heritability—for biobank scale data: using sparse prediction in the UK Biobank to plug diversity gaps in future biobanks.

Authors:

T. Raben¹, E. Widen^{1,2}, L. Lello^{2,1}, S. Hsu^{1,2}; ¹Michigan State Univ., East Lansing, MI, ²Genomic Prediction, Inc., North Brunswick Township, NJ

Abstract Body:

Sparse machine learning algorithms regularly achieve optimal performance at complex trait prediction when compared to other methods. Additionally, enforcing sparsity can greatly reduce computational memory requirements. Here we discuss *new results* that simple L_1 penalized algorithms regularly perform just as well as, or better than, more widely used methods (Bayesian regression with continuous shrinkage priors, elastic nets, etc.) when error factors and uncertainties are taken into account (both empirical and theoretical uncertainty sources). This is demonstrated for a wide variety of case/control and continuous complex traits in the UK Biobank (UKB) including: breast cancer, asthma, hypertension, type 1/2 diabetes, CAD, lipoprotein a, bilirubin, height, and BMI. Polygenic scores (PGS) for these traits cover a wide range in sparsity from as few as 50 to 25,000 selected SNPs. For all traits, the characteristic phase change behavior of compressed sensing can be identified by measuring performance metrics, variance accounted for, and sparsity. The above methods can now be used to identify where the most predictive signal can be found using the least amount of data. A *novel Monte Carlo method* is used to model the performance metrics (AUC, correlation, etc.) for all of these traits. This leads to projections, with confidence intervals, that can be used to estimate the performance of PGS trained in other biobanks. (1) For a case control trait like asthma, L_1 based PGS leads to an $AUC \sim 0.7$ using only genetic information. (Including covariate contributions—e.g., age and sex—leads to even larger AUCs). For asthma, these projections predict L_1 methods will find an $AUC \sim 0.55$ for African Americans for a predictor trained in the newly available All of Us biobank and an $AUC \sim 0.65$ for individuals of Han Chinese ancestry in the Taiwan Precision Medicine Initiative (TPMI). This TPMI result will surpass that found from predictors trained with European ancestry populations. (2) For a continuous trait such as BMI we find an asymptotic correlation ~ 0.4 . This is competitive with heritability estimates from linear mixed model approximations (e.g., GCTA) and indicates that L_1 methods can capture all of the linear SNP heritability. Similar projections for other traits and ancestry groups are also studied. These predictions will help identify how the largest equity gaps in PGS science can be rectified, for example by helping to determine recruiting targets for diverse biobanks.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3517. miRNA Metabolome - wide association study: A multi-omics integrative approach to Asthma

Authors:

R. Sharma¹, K. Mendez^{1,2}, S. Begum¹, M. Huang¹, A. Tiwari¹, J. Celedon³, C. Clish⁴, S. Weiss¹, J. Lasky-Su¹, M. McGeachie¹; ¹Channing Div. of network Med., Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, ²Dept. of Chemistry, Edith Cowan Univ., Perth, Australia, ³Children's Hosp of Pittsburgh of UPMC, Univ. of Pittsburgh, Pittsburgh, PA, ⁴Broad Inst., Cambridge, MA

Abstract Body:

Introduction Both microRNAs (miRNAs) and metabolites have been identified as significant biomarkers for many conditions. Multiple studies have shown that various metabolic stimuli alter miRNA expression. Conversely, miRNAs regulate most cellular processes, impacting metabolism. We postulated that a systems integration of serum miRNAs and metabolites in a large childhood cohort could shed light on the combined synchronized role of miRNAs and metabolites in asthma and in broader metabolomic regulation. **Methods** We performed miRNA sequencing of 1121 serum samples from the Genetic Epidemiology of Asthma in Costa Rica Study (GACRS) and combined these data with targeted LC-MS metabolomic profiling. We performed a global miRNAome-metabolome-wide association (miR-metabo-WAS) analysis using a generalized linear model with adjustment for age, gender, height, and BMI. We applied WGCNA (weighted co-expression network analysis) to identify clusters (modules) of metabolites significantly correlated with miRNA modules and clinical features of asthma. **Results** We identified 2800 significant association between 214 miRNAs and 228 metabolites at 5% FDR. In WGCNA analysis, seven metabolite modules were associated with clinical features of asthma at a 10% FDR (adjusted p-values 8.5×10^{-2} to 2.7×10^{-14}). One of these metabolite modules showed strong significant correlation (P-value: 3.0×10^{-5} to 2.0×10^{-35}) with five miRNAs modules which were also significantly associated with asthma clinical features (adjusted P-value: 8.6×10^{-2} to 4.5×10^{-3}). This module comprised 20 metabolites, including diHOME's, sebacate, and cortisol, that were strongly correlated with six serum miRNAs (mir-143-3p, mir-22-3p, mir-320b/c/d and mir-483-5p) previously reported as asthma-related miRNAs. This module also contained many of the strongest individual miRNA-metabolite associations identified in the miR-metabo-WAS. miRNAs and metabolite modules were associated with airway hyper-responsiveness, airflow obstruction, forced expiratory volume in 1 second, eosinophil count and serum IgE, and asthma-related hospitalizations. **Conclusion** The relatively concentrated effect of miRNAs on metabolite clusters shows that while miRNAs may target and regulate hundreds or thousands of genes, their impact on the resulting metabolome is fairly constrained in asthma. This study of serum miRNA and metabolites demonstrates the value of a combined microRNA-Seq and metabolomics approach to identify meaningful clinical associations and broader genomic regulation of cellular metabolism in Asthma.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3519. MTClass: Identification and annotation of multi-tissue cis-eQTLs using machine learning.

Authors:

R. Li, Z. Qin; Emory Univ., Atlanta, GA

Abstract Body:

Dysregulation of gene expression contributes to the pathophysiology of many human diseases. Mapping genetic variants that affect the expression of one or more genes, termed expression quantitative trait loci (eQTLs), is a growing area of interest. However, to this date, eQTL analysis is typically done one tissue at a time, despite known statistical correlations and biological relationships among tissues. Multi-tissue eQTL mapping identifies variants that affect the expression of the same gene in multiple tissues, thereby potentially recognizing variants with stronger and broader functional impacts that are more likely to lead to phenotypic consequences. However, measuring the effect of a variant on gene expression in multiple tissues is a challenge. To solve this problem, we propose a model-free strategy named MTClass. Instead of calculating a p-value under the statistical testing framework, we attempt to classify vectors of multi-tissue expression values in terms of donor genotype for the variant of interest. We then rank variant-gene pairs based on classification performance. We have tested a support vector classifier and a random forest classifier in this study. We utilized our strategy in a lower-dimensional example (9 tissues, 103 donors) with no missing data and a higher-dimensional example (48 tissues, 948 donors) with missing data. To accommodate missing expression measures in donor-tissue combinations in the latter case, we use multiple imputation with predictive mean matching. We demonstrate in the 9-tissue case that MTClass can identify variants with stronger functional impacts than competing methods such as multi-phenotype association tests. The top 100 and 250 variants identified by MTClass, according to weighted F1 measure, each had 38% more known GWAS SNP hits in their neighborhood according to the GWAS Catalog compared to a competing method. Additionally, we found evidence of biological importance of the top genes identified by MTClass but not by other methods. Using gene set enrichment analysis, we observed that many of these genes serve immune-related functions, especially *HLA-DRB5*, a histocompatibility complex antigen that plays an important role in asthma and allograft rejection. Together, these results emphasize that multi-tissue eQTLs from our method are likely to have functional consequences, and they suggest the value of further pursuing such variants.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3520. Multi trait GWAS for diverse ancestry : Mapping the knowledge gap

Authors:

H. Julienne; Inst. Pasteur, Paris, France

Abstract Body:

Approximately 95% of samples analyzed in univariate genome-wide association study (GWAS) are from European ancestry (EUR) individuals. As recently illustrated in the context of polygenic risk scores, existing data in non-European populations, which often present modest sample sizes, can benefit from innovative methods. Here we conducted a multi trait GWAS using JASS (Joint Analysis of Summary Statistics) on blood traits across four super populations.

We detected 357 new genome-wide significant associations in non-European populations (14 in American (AMR), 72 in African (AFR) and 271 in East Asian (EAS)) and conducted an eQTL-informed functional annotation of all lead SNPs. New associations detected represented respectively 7%, 25% and 21% of all associations in the AFR, AMR and EAS populations.

Between EAS and EUR populations, 179 newly detected genes are shared. Many of these shared genes mapped to blood pathways like *MYB* Proto-Oncogene playing a key role in the regulation of hematopoiesis and Neurobeachin Like 2 (*NBEAL2*) involved Gray Platelet Syndrome. Overall, in EAS, top go terms are substantially more enriched when including genes detected with the joint analysis : the enrichment p-value for the term regulation of immune system decreased from 1.5×10^{-4} to 2×10^{-11} .

Interestingly, gene annotation was poorer for the AMR ancestry suggesting a lower multi-trait or eQTLs annotation signal quality which need further elucidation.

Focusing on the AFR population, six new associations mapped to genes annotated for hemoglobin subunit (*HBG1*, *HBG2*, *HBE1*, *HBB* and *HBD*) involved in Beta Thalassemia and Fetal Hemoglobin Quantitative Trait Locus 1 diseases. At least six other newly detected associations for AFR are annotated for blood trait related pathways including Tubulin Beta 1 Class VI (*TUBB1*) expressed in platelets and *CD36* Molecule (*CD36*) gene protein located at the platelet surface.

By applying JASS on non-European ancestries, we mapped 357 new associations and observed a relevant functional annotation for new signals in the EAS and AFR populations. We argue that multi trait GWAS methods can be a valuable tool to narrow the genetic knowledge gap between European and non-European populations.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3521. Multi-ancestry and sex-specific phenome-wide association meta-analysis of genetically-predicted height.

Authors:

E. Marouli¹, A. Papadopoulou¹, S. Raghavan², E. LITKOWSKI³, M. Graff⁴, Z. Wang⁵, R. Smit⁵, W. Zhu⁶, J. Below⁷, C. Avery⁸, G. Chittoor⁹, A. Justice¹⁰, S. Vedantam¹¹, L. Yengo¹², S. Berndt¹³, J. Arias¹⁴, D. Kim¹⁵, A. Wood¹⁶, T. Assimes¹⁷, Y. Sun¹⁸, J. Hirschhorn¹⁹, R. Loos²⁰, K. North⁴, P. Deloukas¹, GIANT MR-PheWAS-Pleiotropy Working Group; ¹Queen Mary Univ. of London, London, United Kingdom, ²VA Eastern Colorado Hlth.Care System, Aurora, CO, ³Univ. of Colorado, Aurora, CO, ⁴Univ North Carolina, Chapel Hill, NC, ⁵Icahn Sch. of Med. at Mount Sinai, New York, NY, ⁶Vanderbilt Univ., Nashville, TN, ⁷Vanderbilt Univ Med Ctr., Nashville, TN, ⁸Dept. of Epidemiology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁹Dept. of Population Hlth.Sci., Danville, PA, ¹⁰Geisinger, Danville, PA, ¹¹Boston Children's Hosp, Boston, MA, ¹²The Univ. of Queensland, Brisbane, Australia, ¹³Natl. Cancer Inst., Rockville, MD, ¹⁴Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, MD, ¹⁵UNC-Chapel Hill, Chapel Hill, NC, ¹⁶Coll. of Med. and Hlth., Exeter Univ., Exeter, United Kingdom, ¹⁷Stanford Univ Sch Med., Palo Alto, CA, ¹⁸Atlanta VA Hlth.Care System, Decatur, GA, ¹⁹Boston Children's Hosp., Boston, MA, ²⁰Univ. of Copenhagen, Copenhagen, Denmark

Abstract Body:

We performed ancestry-specific and sex-specific Phenome Wide Association Studies (PheWAS) analyses to explore traits and disease related outcomes associated with genetically predicted height. We used a Polygenic Score (PGS) for increased height using summary statistics from the latest Genetic Investigation of ANthropometric Traits (GIANT) trans-ethnic meta-analysis for height and assessed its association with a series of phecodes, which are aggregated ICD-10 codes. We examined health-related phecodes from clinical records in more than 700,000 diverse ancestry individuals. We explored data from European ancestry individuals in the UK Biobank (UKB) (N= 330,965), the Million Veteran Project (MVP) (N=217,225), BioVU (N=71,976) and BioMe (N=7,987). We performed a PheWAS meta-analysis (meta-PheWAS) per ancestry, and applied a Bonferroni adjusted significance threshold. The meta-analysis yielded significant associations with 200 phecodes ($p < 4.66E-05$). Height PGS was associated with an increased risk of many circulatory diseases, including atrial fibrillation and flutter (OR = 1.17, 95% CI [1.16, 1.18], $p = 2.87E-212$). We also report that higher height PGS was associated with a higher risk of chronic venous insufficiency (OR = 1.19, 95% CI [1.16, 1.22], $p = 6.01E-53$). We further explored sex-specific meta-PheWAS in the UKB, MVP and BioVU cohorts. 154 and 57 traits reached phenome-wide significance in males and females respectively. Comparing males and females, the meta-PheWAS yielded 110 significant traits only in males and 19 only in females; 95% of the traits were concordant, and their effect sizes were larger for males. Comparing males with the sex-combined meta-PheWAS, 10 traits were found to be significant only in males, specifically, height PGS was associated with a decreased risk of alcoholic liver damage (OR=0.93, 95% CI [0.90, 0.96], $p = 1.05E-05$) and hyperpotassemia (OR=0.95, 95% CI [0.93, 0.97], $p = 8.37E-06$) in males. Comparing females with the sex-combined meta-PheWAS, no traits were found to be significant in females only. We further performed ancestry specific meta-PheWAS analyses. We investigated individuals of African ancestry in the UKB (N= 6,682), MVP (N=57,065), BioVU (N=15,581) and BioMe (N=6,240) cohorts; we identified 17 significant traits. Higher height PGS was associated with a higher risk of osteoarthritis NOS (OR=1.04, 95% CI [1.02, 1.07], $p = 2.75E-05$) in the African meta-PheWAS and presented an attenuated non-significant effect in the European meta-PheWAS. Furthermore, we explored East-Asian ancestry individuals in the UKB (N= 1,936) and MVP (N=4,843) cohorts where no significant traits were identified.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3522. Multi-ancestry polygenic risk scores for venous thromboembolism

Authors:

Y. Jee¹, F. Thibord², C. Kabrhel³, N. Smith⁴, P. Kraft¹, International Network against Venous Thrombosis (INVENT) consortium; ¹Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ²Div. of Intramural Res., Population Sci. Branch, Framingham, MA, ³Massachusetts Gen. Hosp., Boston, MA, ⁴Univ of Washington, Seattle, WA

Abstract Body:

Venous thromboembolism (VTE) is a significant contributor to morbidity and mortality, with large disparities in incidence rates across ancestry populations. Polygenic risk scores (PRSs) comprised of genome-wide significant variants have been demonstrated to identify European ancestry individuals at the highest risk of VTE. However, there is limited evidence on whether high-dimensional PRS constructed using more sophisticated methods can enhance the predictive ability and their utility in populations of non-European ancestry. We developed PRSs for VTE using summary statistics from the International Network against Venous Thrombosis (INVENT) consortium GWAS meta-analyses of European (71,771 cases and 1,059,740 controls) and African ancestry samples (7,482 cases and 129,975 controls). We used LDpred2, stacked clumping and thresholding, and PRS-CSx to construct multiple PRSs and evaluated their performance in an independent European ancestry sample (2,222 cases and 2,201 controls). LDpred2 trained using European ancestry summary statistics performed the best with OR of 1.51 (95% confidence interval [CI] 1.38-1.67) and area under the curve (AUC) of 0.62 (0.60-0.65). In the European ancestry test set, a multi-ancestry PRS constructed using a linear combination of PRS-CSx scores trained in European and African ancestry populations (AUC=0.60, 0.58-0.63) did not perform better than PRS trained in European ancestry samples (AUC=0.60, 0.58-0.63) or African ancestry samples alone (AUC= 0.58, 0.55-0.60). The highest fifth percentile of the LDpred2 distribution was associated with 2-fold increased risk for VTE (OR=2.04, 1.73-2.40). Results were similar for PRS trained using PRS-CSx in European or African ancestry samples or the multi-ancestry PRS combining European and African scores (OR=1.76 (1.74, 2.42), OR=1.59 (1.67, 2.26), OR=1.87 (1.76, 2.46), respectively). These findings suggest that PRS may be used to identify individuals at highest risk for VTE event and provide guidance for the most effective treatment strategy. We are currently validating these PRSs in African ancestry population.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3523. MultiSuSiE: Multi-population fine-mapping under the sum of single effects model

Authors:

J. Rossen¹, H. Shi¹, M. Kanai², Z. R. McCaw³, L. Liang¹, O. Weissbrod¹, A. L. Price¹; ¹Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, ²Broad Inst. of MIT and Harvard, Cambridge, MA, ³Insitro, South San Francisco, CA

Abstract Body:

Statistical fine-mapping aims to identify biologically causal genetic variants at disease-associated loci. The sum of single effects (SuSiE) model has provided a powerful and versatile approach for fine-mapping causal variants in a single population (Wang et al. 2020 JRSSB) and accommodating functional priors (Weissbrod et al. 2020 Nat Genet). Incorporating data from multiple populations can greatly improve fine-mapping due to differences in patterns of linkage disequilibrium across populations (Schaid et al. 2018 Nat Rev Genet), strongly motivating multi-population methods.

We propose MultiSuSiE, an extension of SuSiE to multiple populations. SuSiE sums across multiple single-effect models (each involving a single causal variant), fitting and residualizing phenotypes for each single-effect model in turn. In MultiSuSiE, each single-effect model still assumes a single causal variant, but effect sizes are allowed to vary across populations via a multivariate normal prior informed by cross-population genetic correlations; MultiSuSiE fits and residualizes phenotypes for population-specific effect sizes of each single-effect model in turn. We additionally adapt the PolyFun framework (Weissbrod et al. 2020 Nat Genet) to multiple populations, integrating genome-wide functional annotations to construct functional priors.

We evaluated MultiSuSiE using simulations involving real genotypes from 93,000 British and 7,000 African individuals from UK Biobank. MultiSuSiE identified 25% more true causal variants with posterior causal probability >0.5 compared to SuSiE applied to 100,000 British individuals, while maintaining correct calibration; MultiSuSiE attained similar improvements relative to existing methods for multi-population fine-mapping. We will present results of applying MultiSuSiE to 49 diseases and complex traits from UK Biobank and incorporating functional priors.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3524. Nearest Neighbor Simulated Annealing Cohort Selection for Improved PCA Genome Matching.

Authors:

R. Laboulaye, V. Borda, T. D. O'Connor; Univ. of Maryland Sch. of Med., Baltimore, MD

Abstract Body:

Mining existing genomic data from biobanks to find controls for GWAS can increase study viability and power. These benefits are further pronounced for underrepresented populations such as Latin Americans, whose studies, compared to those of Europeans, typically have smaller sample sizes and whose fine-scale variation in Native American ancestry can be dominated by continental admixture. Prior work on genome matching focuses on finding the optimal genotype embedding space and distance metric, failing to address the construction of an optimal control cohort and instead defaulting to bipartite matching. Our algorithm focuses on not only finding controls that individually match the querying user's cases and controls but also combine into a cohort that reflects the variation exhibited by the query.

Our method employs local search to find the optimal cohort from a set of reasonable candidates. Using the principal components of both the query and potential controls, we use the variance-weighted Mahalanobis distance to find the α ; nearest neighbors of each query genome from the potential controls, which we then merge into a candidate set. Given a desired control cohort size m , we sample m controls from the candidate set and do so β ; times to generate β ; control cohorts. We use the genomic control λ ;, calculated between a control cohort and the query, to evaluate the β ; control cohorts. The λ ; values are then used to select the optimal starting control cohort and a function of their standard deviation is used to initialize our simulated annealing temperature. We perform simulated annealing for n iterations, randomly swapping γ ; genomes between our control cohort and the candidate set at each iteration, evaluating the control cohort by its genomic control λ ;

We evaluate our control cohort selection algorithm on Latin American genomes from 32 dbGAP cohorts, using both the cases and controls as a query. For our biobank, we use the Genetics of Latin American Diversity (GLAD) database, which contains 52,237 genomes. We filter all genomes to a set of 246799 LD-pruned SNPs and embed them in a PCA space. We construct a greedy baseline in which bipartite matching is applied iteratively until a control cohort of size n has been selected. We first evaluate our algorithm on control cohort size $n=500$ and find that it yields an mean improvement (decrease) in the genomic control of 0.252 and a median improvement of 0.0314 when compared with the baseline. We then evaluate our algorithm on values of n from 500 to 3000 incrementing by 500 for our 10 largest PHS cohorts. We find a gradual decline in the improvement over the baseline as n increases, with a median improvement of 0.0125 for $n=1500$ and 0.000152 for $n=3000$.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3525. Network-based cross-phenotype risk scoring models for non-linear compositing multiple disease risks using biobank-scaled PheWAS data

Authors:

Y. Nam, V. Sriram, S-H. JUNG, B. Xiao, M. Shivakumar, A. Verma, D. Kim; Univ. of Pennsylvania, Philadelphia, PA

Abstract Body:

Background: The polygenic risk score (PRS) can help to identify high-genetic risk by combining individual genetic profiles with single-nucleotide polymorphisms identified through genome-wide association studies. Even though multiple diseases will usually afflict a patient at once or in succession, the conventional PRS focuses on only a single disease of interest. Phenome-wide association studies (PheWAS) can successfully identify associations between multiple phenotypes and genetic variants, allowing us to develop risk scores considering cross-phenotype relationships for index diseases of interest. Genetic associations between phenotypes discovered through PheWAS can be easily implemented through network structure. **Methods:** We developed a network-based cross-phenotype risk score algorithm (netCRS) to predict individual disease risk by utilizing cross-phenotype associations identified from PheWAS. To observe the underlying structure between phenotypes and variants, we designed a multi-layered phenotype-variant relational network using biobank-scaled PheWAS summary statistics. The network consists of a cross-phenotype association network and variant network. Label propagation was applied to predict individual risk scores on the multi-layered network. Individual genetic profiles give initial label information on variant networks, and label propagation diffuses individual genetic profiles into cross-phenotype associations. The final aggregation of propagated results on cross-phenotype association represents a measurement of the possible risk scores across phenotypes focused on an index disease. **Results:** UK Biobank PheWAS summary statistics were used for constructing multi-layered networks, and the individual-level genetic profiles for the European population were collected from the Penn Medicine BioBank. We obtained the netCRS for three dichotomous traits: type 2 diabetes (T2D; 4,400 cases vs. 24,219 controls), obesity (OBS; 4,185 vs. 22,971), and coronary atherosclerosis (CAD; 6,064 vs. 21,900). The number of cross-phenotypes was considered for each trait as follows: 192 phenotypes for T2D, 65 for OBS, and 184 for CAD. To investigate the utility of netCRS, we compared the disease risk prediction between netCRS and PRS (LDpred). The combined model (netCRS + PRS + Sex + Age + PC1~5) achieved an AUC improvement compared to the (PRS + Sex + Age + PC1~5) model; improvement of 3.78% for T2D, 6.57% for OBS, and 3.54% for CAD. **Conclusion:** We expect that using these risk prediction models by incorporating cross-phenotype relationships will allow for the development of prevention strategies and reduction of disease mortality.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3526. New paradigm for uncovering the clinical consequences of genetic variation.

Authors:

M. Xiong¹, R. Fan², L. Luo³, T. Xu⁴, X. Sun⁵, J. Zhao⁶, E. Boerwinkle⁷; ¹Univ Texas Sch. of Publ. Hlth., Houston, TX, ²Georgetown Univ Med Ctr, Washington, DC, ³Univ New Mexico, Albuquerque, NM, ⁴Univ. of Florida, Gainesville, FL, ⁵Univ. of Miami Miller Sch. of Med., Miami, FL, ⁶Univ Florida, Gainesville, FL, ⁷Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Abstract Body:

Uncovering the clinical relevance of genetic variants has a great potential to transform healthcare and medicine. Classical methods for mechanical interpretation of variants are association analysis or supervised prediction, based on sparse, imbalanced and noisy clinical labels. Accumulating evidence indicates that a large proportion of association signals is irrelevant to clinical phenotypes. Moreover, the accuracy estimated from the training data may not hold for out-of-training samples. To overcome these limitations, here we propose a new paradigm and analytical pipeline for uncovering clinical consequences of genetic variations with intelligent genomics and causal inference as major tools. We start with sequence evolution and then follow the path: MSA transformer → variational autoencoder (VAE), estimating effects of mutations via grammatical and semantic analysis of sequence embeddings → estimating causal QTL, eQTL scores (assess the causal relationship between genetic variation, gene expression, and phenotypes) → inferring protein structure and function → finally performing multicriteria score optimization and testing the causal role of genetic alterations in disease pathogenesis. The proposed deep latent generative models with nonlinear interactions capture higher-order, context-dependent constraints in sequence evolution. The proposed paradigm is applied to learning the evolution of SARS-Cov-2 variants, forecasting of potential high risk SARS-CoV-2 mutations, and evaluating their fitness potential and immune escape property. We discover numerous virus mutations that increase fitness potential and immune escape capacity. We demonstrate that the proposed paradigm is notably superior to the state-of-art methods in forecasting potentially harmful future SARS-CoV-2 variants and estimating their immune escape ability.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3527. New perspectives in the genetics of persistent opioid use: reexamining candidate gene studies and presenting new results from the Michigan Genomics Initiative

Authors:

A. Annis¹, V. Gunaseelan¹, A. Smith², G. Abecasis³, D. B. Larach¹, M. Zawistowski⁴, C. Brummett¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Univ MICHIGAN, Ann Arbor, MI, ³Regeneron Pharmaceuticals, Tarrytown, NY, ⁴Univ Michigan, Ann Arbor, MI

Abstract Body:

The opioid epidemic is a major health crisis in the United States, with mortality rates from opioid overdoses rising to tens of thousands of deaths each year. Most opioid-related genetic studies focus on addiction and opioid dependence; however, persistent opioid use stemming from opioid prescriptions is common, especially after surgery. Understanding genetic factors contributing to prescription-engendered persistent opioid use would be a huge advance in the apprehension of morbidity and pathways to opioid use disorder and overdose.

In a comprehensive literature review, we collated over fifty genes and genetic loci from sixty-four peer-reviewed publications that posit genetic contributors to opioid-use phenotypes. ~70% of these publications are candidate gene studies (median case count: 303, median control count: 202); of the few studies classified as genome-wide association studies (GWAS), only two of them (Polimanti et al 2020 and Sanches-Roige et al 2021) have sample sizes $\geq 10,000$.

We determined to systematically replicate reported associations for opioid-use phenotypes in the Michigan Genomics Initiative (MGI), which has ~40,000 surgery patients with opioid prescription data. We did association analyses for persistent opioid use with complete data (3,198 cases who filled ≥ 2 opioid prescriptions in the 30 days prior to 180 days after surgery, 36,321 controls) and with opioid naive participants who have no record of prescription opioid exposure in the 120 to 31 days prior to surgery (794 cases, 32,656 controls). Both analyses replicated ($p < 0.05$) known associations in *OPRM1*, the μ -opioid receptor, but did not replicate most other signals, calling into question the validity of small-sample candidate gene studies for opioid use. Importantly, the MGI analyses showed nominal significance for two recently-reported loci, rs640561 ($p = 0.02$) and rs9291211 ($p = 0.04$), which are associated with opioid dependence and problematic opioid use in the Polimanti et al and Sanches-Roige et al GWAS, respectively.

While MGI did not replicate most previously-reported associations, finding signals for known *OPRM1* variants and loci from two sizable GWAS affirms the persistent opioid use phenotype in MGI, gives hope for the tractability of genetic contributors to opioid-use phenotypes, and demonstrates the need for genome-wide association analyses to understand complex traits.

[1] <https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates> [2]

<https://jamanetwork.com/journals/jamasurgery/fullarticle/2618383> [3]

<https://www.nature.com/articles/s41380-020-0677-9> [4] <https://www.nature.com/articles/s41380-021-01335-3>

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3528. Next generation phenotyping of DEGCAGS syndrome.

Authors:

R. Freeman¹, J. Woods²; ¹Valley Children's Hosp., Fresno, CA, ²Valley Children's Hosp., Madera, CA

Abstract Body:

Introduction: Developmental Delay with Gastrointestinal, Cardiovascular, Genitourinary, and Skeletal Abnormalities (DEGCAGS) is a recently discovered, autosomal recessive disorder with a distinct multi-systemic phenotype (MIM # 619488). *ZNF699* has been proposed as an etiology of DEGCAGS, though its function in human development and its role in disease have yet to be elucidated. Fourteen patients with DEGCAGS have been reported in the literature. Herein we report on the fifteenth patient using Next Generation Phenotyping (NGP) for quantitative determination of the core DEGCAGS phenotypes. **Methodology:** Phenotypic data present in pdf files were derived from the patient chart and literature review. These data were imported into R, combined into one file and converted to txt files (pdftools package). The txt files were then processed in ClinPhen for Human Phenotype Ontology (HPO) term extraction. The HPO terms were sorted by frequency according to the HPO hierarchy of terms (ontologyX). A co-occurrence matrix of HPO terms was generated to create an interactive heat map of HPO term co-occurrence frequency (heatmaply). **Results:** A total of 264 unique HPO terms were extracted from the patient record and literature review. The most frequent HPO terms within the top third were hypotonia (19), syndactyly (18), abnormal facial shape (17), microcephaly (12), failure to thrive (6), global developmental delay (6), hearing impairment (5), intellectual disability (5), intestinal atresia (5) and jejunal atresia (5). The count of the top half of HPO systemic hierarchical categories was head (82 HPO terms), heart morphology (48), kidney (25), limb bone deformity (25), cardiovascular system physiology (19), muscle tone (19), hearing abnormality (17) and neurodevelopmental abnormality (17). The top quartile of co-occurring phenotypes were syndactyly/abnormal facial shape (7 co-occurrences), syndactyly/hypotonia (6), hypotonia/abnormal facial shape (5), syndactyly/jejunal atresia (5) and abnormal facial shape/failure to thrive (5). **Discussion:** NGP offers a quantitative approach to phenotyping to generate phenotypic data in an objective and reproducible manner. Co-occurrence matrices provide insight into the core phenotypes associated with disorders. Systemic categorization of phenotypes within the HPO hierarchy provides insights into the biomedical systems involved in disorders that may go unrecognized if only individual phenotypes are reported. NGP methodology may be widely applied to existing and novel genomic disorders to provide systemically sorted quantitative data for those interpreting genomic data within the context of patient phenotypes.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3529. NGS Variant Detection as a Sequence-to-Sequence Modeling Problem

Authors:

B. O'Fallon¹, J. Durtschi², A. Bolia¹, L. Yang¹, H. Best³; ¹ARUP Labs, Salt lake city, UT, ²ARUP, Salt Lake City, UT, ³Univ. of Utah/ARUP Lab., Salt Lake City, UT

Abstract Body:

NGS variant detection tools use many statistical techniques to identify candidate variants, including logistic regression, gaussian mixture models, pair Hidden Markov Models (HMMs), and multiple hand-picked parameter values and heuristic thresholds. Here, we describe an alternative variant detection approach that does not involve any statistical modeling, and instead uses a single deep learning model to directly predict the two germline haplotypes present in the sample. Our model recasts the variant detection problem as a sequence-to-sequence modeling task, akin to language translation, in which the input sequence tokens are the collection of bases aligning to a single reference genome coordinate, and the predicted output sequences are the two haplotypes present in the individual. The model architecture utilizes stacked transformer encoder layers and a simple, fully connected decoder modified to predict two output sequences. We use training data from 17 Genome-in-a-Bottle cell lines sequenced as full genomes on Illumina sequencing instruments, and investigate how model size and training data size affect sensitivity and specificity of the resulting variant calls. We show that our model has the ability to reconstruct haplotypes in regions with complex or ambiguous read mappings, indicating that it learns a HaplotypeCaller-like ability to perform de novo local reassembly of reads. In addition we demonstrate that the model uses local context information when assessing the likelihood of a given variant. Overall, genotyping and phasing accuracy exceeds 99% for most variant classes, indicating that both haplotypes are reconstructed with high accuracy. When compared to HaplotypeCaller, DeepVariant, and Strelka2, our approach yields similar sensitivity for both SNVs and indels, but slightly reduced precision for SNVs.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3530. Non-additive outlier effects of gene expression on anthropometric traits

Authors:

A. Brown¹, T. Dupuis², A. Sartori³, A. Real⁴, D. Davtian¹, E. Pearson¹, E. Dermitzakis⁵, A. Viñuela⁶; ¹Univ. of Dundee, Dundee, United Kingdom, ²Univ. of Dundee, Sch. of Med., Dundee, United Kingdom, ³Université de Genève, Genève, Switzerland, ⁴Univ. of Geneva, Geneva, Switzerland, ⁵GSK, Geneva, Switzerland, ⁶Newcastle Univ., Newcastle, United Kingdom

Abstract Body:

Approaches that use genetic data to produce causal links between molecular phenotypes and disease risk usually assume that additive effects on the molecular trait translate to additive effects on disease risk. However, many complex biological processes involve non-additive mechanisms such as buffering, pathway substitution and feedback loops, which would be expected to reduce the impact of extreme expression phenotypes. In the opposite direction, some rare variants with extreme expression effects are known to cause monogenic disease. By considering combinations of eQTL variants with predicted extreme effects on expression, we test the effect of extreme expression on 6 anthropometric traits. We used eQTLs from DIRECT and GTEx v8 to calculate predicted expression in 44 non-sex specific tissues for 405,719 individuals (UK Biobank). For every gene across all tissues, we defined low (LEEI) and high extreme expression individuals (HEEI) as the persons whose predicted expression levels fell in the bottom or top percentile. We tested for trait differences between these individuals and the rest of the population for 6 traits: systolic and diastolic blood pressure, BMI, grip strength, height, and waist-hip ratio. By controlling for additive effects, we are searching for instances where the combined effects of the variants on the trait, mediated by expression, differs from the sum of the individual variant effects. We found respectively 2, 5, 48, 22, 617 and 35 associations (FDR=0.05) between these traits and the extreme expression of a gene in a tissue. In most cases, genes with non-additive associations between extreme expression and phenotype also showed significant additive correlations (609/729). We observed that extreme expression generally attenuates the contribution of the additive component, suggesting that compensatory mechanisms are in action. We find considerably more evidence for additive expression effects on traits (77,960 total associations), consistent with a dramatically higher power to discover these relationships. However, these non-additive associations provide other insights into how molecular phenotypes affect traits and potentially also provide a way of linking genes to disease with less scope to be confounded by pleiotropy and linkage contamination.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3531. Novel Approach to Age Estimation of Rare Variants Enables Prioritization of Functional Variants

Authors:

K. Liao¹, H. Kang², S. Zollner³, Y. Si¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Univ Michigan, Ann Arbor, Ann Arbor, MI, ³Univ Michigan, Ann Arbor, MI

Abstract Body:

Due to the abundance of rare variation, a priori evidence into their functionality is key to developing successful testing strategies for trait associations. However, annotating the consequences of rare variants (especially non-coding variants) based on genomic features is a challenging problem. A piece of evidence into a rare variant's functionality is whether it is undergoing selection. Currently, this is typically considered using conservation scores, which reflect purifying selection over phylogenetic time scales. Evidence of selection within the human population can complement this with more recent signals. A useful proxy for such selection is the age of the variant where variants under selection are typically much younger than other variants of the same allele frequency. Here, we propose a novel method to accurately infer the relative age of a rare variant and highlight how allele age reveals insights into functional consequence.

In this method, we model that carriers of a younger allele also share a more recent common ancestor in the region surrounding the variant site, resulting in longer shared haplotypes nearby. Moreover, due to the more recent common ancestry the density of rare variants on these haplotypes is lower. We leverage these observations to estimate the pairwise time to common ancestor between carrier haplotypes. We combine the pairwise estimates of all carriers to arrive at the age of the variant. As the numerical value of this estimate strongly depends on the population genetic model, we primarily focus on the relative age of a variant. We evaluate our approach in coalescent simulations, using published models of human population history. Across all simulated populations, our approach provides accurate relative allele age estimates with rank correlation above 0.8 with the true ages among variants with the same allele counts. The correlation ranges from 0.86 for doubletons to 0.82 for variants occurring 10 times in the sample in samples between 1,000 and 10,000 individuals. We validate our method in sequencing data by observing that known loss of function variants are younger than synonymous variants of the same allele frequency. We further compare our approach to two recent methods, GEVA and tsdate in both simulation and real data.

Our method is computationally efficient to apply to whole genome sequencing data of >100,000 individuals. These estimated ages provide insights into evolutionary constraints of rare noncoding variants and may thus help identify rare risk variants.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3532. Omics profiling in overweight and obese children in the California Bay Area during childhood development and the effects of a multi-modal health intervention program tailored to a majority Latinx community

Authors:

J. Li-Pook-Than¹, E. Wei¹, B. Lee-Mcmullen¹, F. Haydel¹, S. Banuelos¹, T. Robinson¹, M. Snyder²; ¹Stanford Univ., Palo Alto,, CA, ²Stanford Univ., Stanford, CA

Abstract Body:

Childhood development and health risks associated to overweight and obesity stresses are understudied particularly in diverse populations. We describe preliminary multi-omic results of a five-year weight loss intervention program on ~241 overweight and obese children (average 7- 11 year olds) recruited from primarily LatinX communities in the Bay Area of Northern California. Children in the study were randomized into either a community based multi-component, multi-level, multi-setting (MMM) intervention group or active placebo group (control). Using traditional clinical measurements along with comprehensive -omic profiling (including transcriptomes, metabolome and O-link cardiovascular-related proteins) we describe longitudinal changes over the course of the study. Among our findings, are key biological pathways (KEGG) associated to genes that decreased in expression when comparing MMM intervention from control groups (deltas), including genes associated to inflammation (chemokine), insulin, adrenergic signaling in cardiomyocytes, as well as pathways related to cancer and long-term depression. Furthermore, Ingenuity Pathway Analysis (IPA), reveal a crosstalk of networks between these pathways, -omic analytes and components related to childhood development. This unprecedented work also shows that inclusivity and equitable research supports improved precision health science understanding.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3533. Omnibus proteome-wide association study (PWAS) method identified 28 novel risk genes for Alzheimer's disease.

Authors:

T. Hu¹, R. L. Parrish¹, A. S. Buchman², N. T. Seyfried³, P. L. De Jager⁴, D. A. Bennett², M. P. Epstein¹, J. Yang¹; ¹Ctr. for Computational and Quantitative Genetics, Dept. of Human Genetics, Emory Univ. Sch. of Med., Atlanta, GA, ²Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, ³Dept. of Biochemistry, Emory Univ. Sch. of Med., Atlanta, GA, ⁴Columbia Univ Irving Med Ctr, New York, NY

Abstract Body:

Alzheimer's disease (AD) is a neurodegenerative disorder with polygenic inheritance. The most recent Genome-wide association studies (GWAS) of AD identified 38 risk loci, the majority of which have unknown underlying mechanisms. Proteome-wide association study (PWAS) integrating proteomics data with GWAS summary data has been shown as a powerful tool to identify risk genes associated with AD, with genetic effects potentially mediated through protein synthesis. Combining multiple PWAS methods is expected to improve power by modeling the unknown genetic architectures of protein abundances with multiple models.

Analogous to the transcriptome-wide association study (TWAS), PWAS first trains protein abundance imputation models using nearby SNPs as predictors with a reference dataset, and then tests gene-based association with GWAS summary data taking the estimated SNP effect sizes from the reference data as SNP weights (i.e., pQTL weights). Existing TWAS methods such as PrediXcan (Elastic-Net penalized regression), FUSION, and TIGAR (nonparametric Bayesian Dirichlet process regression) can be easily adapted for PWAS. We applied both PrediXcan and TIGAR to train protein abundance imputation models for 8874 proteins obtained from prefrontal cortex samples of European descent (n=395). We first conducted PWAS using the most recent GWAS summary data of AD (n=762,917) with pQTL weights by PrediXcan and TIGAR, respectively. Next, we used the omnibus aggregated Cauchy association test (ACAT-O) to combine PWAS p-values by PrediXcan and TIGAR, as well as the PWAS p-values using the previously published pQTL weights (by FUSION).

We obtained 6673 protein abundance imputation models by TIGAR/DPR and 1835 by PrediXcan, with 5-fold cross validation $R^2 > 0.5\%$. PWAS by TIGAR detected 11 significant risk genes, and PWAS by PrediXcan detected 7 significant risk genes, with $FDR < 0.05$. The omnibus ACAT-O method identified 32 significant risk genes with $FDR < 0.05$, where 4 genes overlapped with previous PWAS by FUSION (which detected 13 risk genes) and 7 genes were identified by previous GWAS of AD. Interestingly, our identified risk genes *FNBP4* and *CDC42* were reported for general cognitive function; *CCDC86* and *KIF18B* were found involved in the development of neurofibrillary tangles. In conclusion, 28 novel PWAS risk genes were identified for AD, whose genetic effects are potentially mediated through the protein synthesis processes. These results highlight the effectiveness of using existing TWAS methods and the omnibus ACAT-O test for PWAS. This omnibus PWAS method can be applied to study complex polygenic diseases and provide new insights into their genomic architecture.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3534*. On the polygenic trait model in population-based human genetics studies: What is random and what is fixed.

Authors:

Y. Tang, J. D. Storey; Lewis-Sigler Inst. for Integrative Genomics, Princeton Univ., Princeton, NJ

Abstract Body:

The polygenic trait model (PTM) forms the basis of the genetic analysis of complex quantitative traits. The PTM has wide and reliable applications in experimental settings where researchers control the genetics of the experimental units. These scenarios satisfy the assumption of a fixed kinship in the original PTM model. A well-established way to construct the PTM under these conditions is using the sampling distribution to develop a marginal polygenic trait model (mPTM) to conduct inferences based on the marginal likelihood. When the mPTM is utilized in population-based human genetics studies where the kinship is unknown in advance, there is a trend in the literature to use another probabilistic model with fixed genotypes and random genetic effects (FGRE) to perform inference on the mPTM. However, the FGRE adopts a biased kinship estimator and a nontrivially different sampling distribution due to its assumption of fixed genotypes. This results in a different marginal likelihood, thereby affecting the estimation of genetic variance components and heritability. Following the original assumptions of the PTM -- random genotypes and fixed genetic effects (FPTM) -- we describe how to utilize the PTM in population-based studies by equipping the original PTM with a recently developed unbiased kinship estimator. We provide evidence from theory and simulations to suggest that the FPTM yields more accurate estimates of variance components and heritability. Applying the FPTM to the analysis of real data with complex population structures, we found that the FPTM has the ability to modulate the ancestral population when estimating genome-wide heritability. The FPTM also leads to a more accurate characterization of the effects of the choice of ancestral population on statistical power in genome-wide association studies. Our findings may assist in choosing appropriate model assumptions and methods to construct the PTM for human genetics studies with arbitrary population structures, thereby leading to a better understanding of the molecular basis of complex human traits and diseases.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3535. One-sample and two-sample mendelian randomization studies consistently showed that an increase in adiposity raises serum uric acid level in the East Asian populations.

Authors:

S. Lee¹, M. Ryu¹, N. Kim¹, S. Yoon¹, S-H. Jee²; ¹Basgenbio, Seoul, Korea, Republic of, ²Yonsei Univ., Seoul, Korea, Republic of

Abstract Body:

Previous mendelian randomization studies (MR) suggested that the increase in serum uric acid causes adiposity, not vice versa. However, most findings came from the data of the European ancestry. Among East Asian populations, both directions in the causal relationship between adiposity and uric acid have been suggested by observational studies and a few MR studies. Therefore, in this study, two separate MR studies were conducted to test the hypothesis of the causal effect of adiposity, measured by body mass index (BMI), on uric acid in East Asian populations. In the first study, summary statistics from two large, publicly available biobanks from South Korea (72,299 participants in the Korean Genome and Epidemiology Study; KoGES) and Japan (158,284 participants in the Biobank Japan; BBJ) were used to conduct a two-sample MR study. In the second study, a one-sample MR study was conducted using the BBJ only. In both studies, the fixed-effect inverse variance weighted method was the primary analysis, and as sensitivity analyses, leave-one-out analyses, the weighted median method, and the MR-Egger regressions were performed. Twenty-six single nucleotide polymorphisms (SNPs) in the KoGES and 52 SNPs in the BBJ were selected as genetic instruments. In the two-sample MR study using the KoGES and the BBJ, genetically determined BMI had a positive causal effect on uric acid (IVW: β [95%CI] = 0.029 [-0.006, 0.065]; weighted median: β [95% CI] = 0.042 [0.019, 0.065]). While the causal effect estimate from the MR-Egger regression was not statistically significant (β [95%CI] = -0.002 [-0.113, 0.110]), there was also no evidence of horizontal pleiotropy from the MR-Egger regression intercept (β [95%CI] = 0.004 [-0.010, 0.018]). In the one-sample MR study using only the BBJ, the positive causal effect of BMI on uric acid was also observed (IVW: β [95%CI] = 0.197 [0.129, 0.264]; weighted median: β [95% CI] = 0.170 [0.095, 0.244]). Leave-one-out analyses confirmed that the pooled causal effect estimate from various SNPs was not driven by any single SNP. Our findings from the South Korean and Japanese populations are consistent with the causal effect of BMI on uric acid previously found in the European population, suggesting that such pattern may be universal across different populations. When analyzed in both two-sample MR and one-sample MR settings, the causal effect estimates differed in magnitude because the two MR designs tend to be biased in the opposite directions. Nonetheless, the directions of the causal effects were consistent in both studies, increasing the confidence in our findings.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3536. OTTERS: A powerful TWAS framework leveraging summary-level reference data

Authors:

Q. Dai¹, G. Zhou², H. Zhao³, U. Võsa⁴, A. Battle⁵, A. Teumer⁶, T. Lehtimäki⁷, O. Raitakari⁸, T. Esko⁹, M. Epstein¹, J. Yang¹; ¹Emory Univ., Atlanta, GA, ²Yale Univ., New Haven, CT, ³Yale Univ. Sch. of Publ. Hlth., New Haven, CT, ⁴Univ. Med. Ctr. Groningen, Groningen, Groningen, Netherlands, ⁵Johns Hopkins Univ., Baltimore, MD, ⁶Univ. Med. Greifswald, Greifswald, Germany, ⁷Tampere Univ., Tampere, Finland, ⁸Univ. of Turku, Turku, Finland, ⁹Univ. of Tartu, Tartu, Tartumaa, Estonia

Abstract Body:

Transcriptome-wide association studies (TWAS) identify genes that influence complex traits through genetic regulation of gene expression. Existing TWAS tools like FUSION require individual-level eQTL reference data from a source like GTEx (n~100s) to impute gene expression in a test GWAS. Therefore, these TWAS tools are not applicable to enormous summary-level reference eQTL datasets like those generated by the eQTLGen consortium (n~32K). Development of TWAS methods that can harness summary-level reference data like eQTLGen is valuable not only to enable TWAS in more general settings but also to permit more powerful analyses, since we expect expression prediction to improve with increasing reference sample size. To fill this important gap, we develop a TWAS framework called OTTERS (Omnibus Transcriptome Test using Expression Reference Summary data) that can use summary-level reference eQTL datasets to perform TWAS. OTTERS adapts a variety of published polygenic risk score (PRS) methods (P+T, lassosum, SDPR, PRS-CS) to train eQTL effect sizes based on a multivariate regression model. For each PRS method, OTTERS uses the estimated eQTL weights to impute gene expression per gene and tests for association between imputed expression and outcome for a GWAS dataset. OTTERS then combines these PRS-based association tests together to create an omnibus TWAS p-value. Simulation studies using real genotype data to simulate expression data illustrated that OTTERS based on summary-level reference eQTL datasets achieved comparable results with FUSION using individual-level reference data. We also demonstrated that TWAS performance based on each of these PRS methods substantially depends on the underlying genetic architecture for gene expression. Since such architecture is unknown a priori in real studies, we believe the omnibus TWAS test implemented in OTTERS will be preferred in practice as this test had near-optimal performance across all of our simulation scenarios. We applied OTTERS to blood eQTL summary-level data (n=31,684) from the eQTLGen consortium and GWAS summary data of cardiovascular disease from the UK Biobank. Comparing OTTERS results to those of FUSION using individual-level GTEx V6 reference data (n=338 for blood tissue), we found that OTTERS revealed 11 potential risk genes missed by FUSION using smaller individual-level reference datasets. In conclusion, our novel OTTERS framework not only provides a practical and powerful tool for TWAS analysis, but also provides the opportunity to leverage other emerging summary-level -omic data, such as methylation, histone marks, and proteins. Free software implementing OTTERS is available on Github.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3537. PathWAS: Combining transcriptomics with proteomics to predict pathway functionality and associations with complex traits

Authors:

S. May-Wilson¹, J. Wilson², N. Pirastu³; ¹Univ. of Edinburgh, Edinburgh, United Kingdom, ²Univ Edinburgh, Edinburgh, United Kingdom, ³Human Technopole, Milan, Italy

Abstract Body:

Rationale: GWAS are a staple in the analysis of complex traits and multifactorial disease. They are often combined with pathway-enrichment analyses to try and elucidate biological mechanisms and relationships, but unfortunately this methodology may miss important individual interactions due to small effect sizes caused by focusing on individual genes acting in isolation. However, combining multiple genes into one broader pathway score may allow us to gain greater insight into the aetiology of complex traits while improving power of discovering novel associations. Our aim was thus to create a statistical framework for creating genetic scores predictive of pathway functionality and then use them to understand the impact of the pathway on complex traits.

Method: PathWAS involves the use of measured protein biochemically at the end of a pathway, as a proxy for pathway functionality. For this we select all available cis-eQTLs for each gene within a pathway from the eQTLgen consortium. We then combine all predicted gene expression in a single pathway score by evaluating the weight of each eQTL on the end-point protein. Overall scores are then derived as the weighted sum of the PRS for each transcribed gene level, where the weights correspond to the MVMR coefficients. For end-point protein summary statistics we used 4907 proteins from the DeCODE proteomics (in 35,000 Icelandic individuals), while pathways were taken from the KEGG database. Finally, each PRS_{pathway} was used for a PheWAS of 60 traits in UK Biobank.

Results: Using the available proteomics data we were able to create 1109 pathway-protein pairs. From these, PheWAS identified more than 3000 significant associations between pathways and different traits (FDR < 0.05). For example, higher activity of the SNAP Receptor (SNARE) interactions pathway increased platelet count while higher activity of NOD-like receptor pathway and reduced BMI.

Conclusion: Our results show that it is possible to use publicly available data to create PRS predictive of pathway functionality and that these can be used to understand their impact on human biology. With the integration of single cell eQTL data which is gradually becoming more widely available it will potentially be possible to create tissue and cell-type specific scores. In conclusion our method extends TWAS analysis to entire pathways increasing power to detect associations. This will be extremely important in understanding the underlying physiology of complex traits while providing potential targets for drug development.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3538. Pathway analysis identifies novel non-synonymous variants contributing to extreme vascular outcomes in Williams-Beuren syndrome.

Authors:

D. Liu¹, C. J. Billington², N. Raja³, Z. C. Wong¹, M. D. Levin⁴, W. Resch⁵, C. Alba⁶, D. N. Hupaloo⁷, E. Biamino⁸, M. F. Bedeschi⁹, M. Digilio¹⁰, G. M. Squeo¹¹, R. Villa⁹, P. Parrish¹, S. Osgood¹², C. L. Dalgard¹³, G. Merla¹⁴, B. Pober¹⁵, C. Mervis¹⁶, A. Roberts¹⁷, C. Morris¹⁸, L. Osborne¹⁹, B. Kozel²⁰; ¹Natl. Heart, Lung, and Blood Inst., Bethesda, MD, ²Univ. of Minnesota, Minneapolis, MD, ³Natl. Heart, Lung and Blood, Bethesda, MD, ⁴Natl. Heart, Lung and Blood Inst., Bethesda, MD, ⁵Ctr. for Informational Technology, NIH, Bethesda, MD, ⁶The Henry M Jackson Fndn., Bethesda, MD, ⁷Uniformed Services Univ., Bethesda, MD, ⁸Univ. of Turin, Turin, Italy, ⁹Fondazione IRCCS Ca Granda Ospedale Maggiore Policlinico Med. Genetic Unit, Milan, Italy, ¹⁰Bambino Gesù Children's Hosp., Rome, Italy, ¹¹Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, ¹²Natl. Heart, Lung and Blood Inst., Bethesda, MD, ¹³USUHS - TAGC, Bethesda, MD, ¹⁴IRCCS Casa Sollievo Della Soff, San Giovanni Ro, Italy, ¹⁵Massachusetts Gen. Hosp., Boston, MA, ¹⁶Univ. of Louisville, Louisville, KY, ¹⁷Boston Children's Hosp., Boston, MA, ¹⁸Univ Nevada Sch Med, Las Vegas, NV, ¹⁹Univ Toronto, Toronto, ON, Canada, ²⁰NIH, Bethesda, MD

Abstract Body:

Introduction: Supravalvar aortic stenosis (SVAS), a narrowing of the aorta above the aortic valve, is a characteristic feature of Williams-Beuren syndrome (WBS). SVAS is present in 67% of those with WBS, but severity varies; 21% have SVAS requiring surgical intervention while 46% have milder SVAS and 33% have no SVAS. Finding genetic modifiers contributing to SVAS severity in WBS has been elusive since 1993 when hemizyosity of the ELN locus at 7q11.23 was recognized as the major driver of vascular pathology in WBS.

Methods: We collaboratively phenotyped 473 individuals with WBS and performed the largest whole genome sequencing (WGS) study to date for this condition. Most of statistical packages developed for genome-wide association study (GWAS) require thousands of cases and controls, which is unattainable for most rare disease studies, including WBS. Accordingly, we developed a set of strategies for identifying pathway-level modifiers using WGS data. The strategies start with extreme phenotyping (surgical SVAS vs. no SVAS) and pre-selection of non-synonymous variants with increased predicted functional impact (mild or high CADD score) and an allele frequency (AF) difference between the extreme phenotype groups (5% for common variants and present in at least 1% of one extreme phenotype and 0% in the other extreme for the less frequent variants). These filtered variants account for 1-2% of the total variants. We then identified pathways enriched in common or less frequent variants using the separate allele frequency filters, followed by association testing of SVAS severity with the enriched pathways. We also compared the genes in associated pathways in our analysis to the genes detected from large GWAS studies on aortic related traits/diseases.

Results: The common variant analysis identified key pathways including the extracellular matrix and the innate immune system, while pathways encompassing adaptive immunity, ciliary function, lipid metabolism and PI3KAKT were captured by both the common and less frequent variant analyses. Cell cycle and estrogen responsive pathways were among those identified through the less frequent variant analysis. Eleven genes implicated by this study, including PCSK9 and ILR6, were previously identified in large GWAS assessing aortic traits, suggesting overlapping disease mechanisms. The set of novel strategies presented here were useful for the identification of disease modifiers in WBS and can also be applied to other rare conditions using WGS data.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3539. Pathway analysis using expression factors reveals biological processes underlying genetics of complex traits

Authors:

X. Sun, X. He; The Univ. of Chicago, Chicago, IL

Abstract Body:

GWAS have mapped a large number of loci associated with common human traits. To translate these associations into biological insights, one strategy is to analyze GWAS data of a trait across the genome to identify important gene pathways. The pathway analysis is usually based on enrichment of association signals near genes of a certain pathway. It is often unclear, though, whether the enriched pathways have any causal roles in the trait. For instance, if a large number of genes in a causal pathway of a trait also appear in other pathways, then all these pathways may show enrichment of genetic signals.

We propose a new type of pathway analysis. We hypothesize that genetic variants of a trait may perturb activities of a relatively small number of “core pathways”, which affect the phenotype. This hypothesis is supported by the observations that trait-associated variants may affect expression of many related genes in trans. To map these pathways, we propose to use factor analysis on gene expression data. The factors of genes in a pathway capture correlated expression of these genes, thus can be viewed as a proxy of the pathway activity. Importantly, these factors can be treated as if they were measured data, so that one can test association of variants with these factors, and assess their roles in phenotypes.

We applied this approach to an expression QTL dataset of immune cells. We obtained expression factors of KEGG pathways, and tested association of these factors with all variants in the genome. This test was, however, under-powered. Thus we limited our analysis to variants already associated with phenotypes of interest, blood and immune traits in our case. To assess the role of a pathway in a trait, we tested if the trait-associated variants tend to be associated with the expression factor of this pathway. This analysis revealed numerous significant pathway-trait pairs.

For these pairs, we performed additional analysis to assess possible causal roles of pathways. Borrowing ideas from Mendelian Randomization, we reason that if a pathway factor has a causal effect on a phenotype, then the effect sizes of the variants on the phenotype will be correlated with the effect sizes on the factor. Using this “effect consistency test”, we found 99 pathway-trait pairs. As an example, the top pathway of Crohn’s disease, “Phosphatidylinositol signaling”, has a known role in autoimmune diseases. Interestingly, 55 (53.4%) of these pairs also show enrichment under classic pathway analysis (MAGMA), providing independent support. In conclusion, we have developed a novel approach to pathway analysis of complex human traits. The pathways we found may offer important clues of the etiology of common diseases.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3540*. Performance and accuracy of imputation panels for genetic association studies in sub-Saharan African populations

Authors:

D. Sengupta¹, G. Botha², A. Meintjes², M. Mabiavanga², S. Hazelhurst¹, N. Mulder², M. Ramsay¹, A. Choudhury¹; ¹Univ. of the Witwatersrand, Johannesburg, South Africa, ²Univ. of Cape Town, Cape Town, South Africa

Abstract Body:

Background: The quality of genotype imputation largely relies on the size and genetic proximity of a reference panel to the population for which data are being imputed. Given several fold differences in size and varying degrees of representation of populations from the African continent in the current panels, the selection of an optimal panel is a major challenge for sub-Saharan African (SSA) genome-wide association studies (GWAS).

Methods: A cohort of ~10,900 SSA participants was genotyped using the H3Africa SNP array. This dataset was imputed using five widely used reference panels hosted at the Sanger (African Genome Resource (AGR), KGP and HRC panels), TOPMed and Michigan (KGP) Imputation services and the imputation quality was assessed and compared. High-coverage (>30x) whole-genome sequence (WGS) data from 95 of the SSA samples was used as the "truth-set" to evaluate the accuracy of imputation.

Results: The comparisons show that the two best performing panels, TOPMed and AGR, are able to accurately capture over 75% of the SNPs present in these African genomes. Despite being about 20-fold smaller, AGR not only performs comparably to TOPMed in imputing SNPs (other than extremely rare SNPs, MAF<0.005) but also shows lower discordance with the WGS data (2.2±0.5% for AGR and 3.6±1.9% for TOPMed). Each panel imputed some unique content. TOPMed uniquely imputed ~18M SNPs while AGR uniquely imputed ~9.5M SNPs. Meta-imputation is emerging as an approach to further enhance the quality and quantity of imputation. Our comparisons suggest that such an approach could lead to further improvements in increasing the size of imputed dataset and reducing imputation errors. However, comparison with WGS showed that each panel imputed a considerable number of SNPs (7% for AGR and 9% for TOPMed) that were not observed in WGS, underlining the necessity for cautious scrutiny of results. We observed notable differences in imputation of datasets from three African geographic regions (East, West and South). Characterization of admixture proportions in participants further showed that a considerable portion of the variability was driven by ancestral differences. For instance, individuals with higher Khoe-San ancestry imputed significantly more SNPs than other Southern Africans.

Conclusions: Our results highlight the need for larger reference panels and inclusion of geographically and ancestrally diverse WGS data to improve the imputation of SSA genotype datasets in preparation for GWAS. We also underline the value of developing resources for harmonization and meta-imputation of datasets across platforms/services.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3541. Personality traits are consistently associated with blood mitochondrial DNA copy number estimated from genome sequences in two genetic cohort studies.

Authors:

R. Oppong¹, A. Terracciano², M. Picard³, Y. Qian⁴, T. J. Butler¹, T. Tanaka⁴, A. Z. Moore¹, E. M. Simonsick¹, K. Opsahl-Ong¹, C. Coletta¹, A. Sutin⁵, M. Gorospe¹, S. M. Resnick¹, F. Cucca⁶, S. W. Scholz⁷, B. J. Traynor⁸, D. Schlessinger⁹, L. Ferrucci¹, J. Ding⁴; ¹Natl. Inst. on Aging, Baltimore, MD, ²Dept. of Geriatrics, Florida State Univ., Tallahassee, FL, ³Dept. of Neurology, Columbia Univ., New York, NY, ⁴Natl. Inst Aging, Baltimore, MD, ⁵Florida State Univ., Tallahassee, FL, ⁶IRGB CNR, monserrato, Italy, ⁷NIH, Bethesda, MD, ⁸Natl. Inst Aging, Bethesda, MD, ⁹NIA, Baltimore, MD

Abstract Body:

Mitochondrial DNA copy number (mtDNA_{cn}) in tissues and blood can be altered in conditions like diabetes and major depression and may play a role in aging and longevity. However, little is known about the association between mtDNA_{cn} and personality traits linked to emotional states, metabolic health, and longevity. This study tests the hypothesis that blood mtDNA_{cn} is related to personality traits and mediates the association between personality and mortality. We assessed the big five personality domains and facets using the Revised NEO Personality Inventory (NEO-PI-R), assessed depressive symptoms with the Center for Epidemiologic Studies Depression Scale (CES-D), estimated mtDNA_{cn} levels from whole-genome sequencing, and tracked mortality in participants from the Baltimore Longitudinal Study of Aging (722 participants with complete data, mean age 75 (48 - 100), 48% women). Results were replicated in the SardiNIA Project (587 participants with complete data, mean age 57 (15 - 96), 62% women). We found that mtDNA_{cn} was negatively associated with the Neuroticism domain and its facets and positively associated with facets from the other four domains (Extraversion, Openness, Agreeableness, and Conscientiousness). The direction and size of the effects were replicated in the SardiNIA cohort and were robust to adjustment for potential confounders such as age, sex, platelet count, and white blood cell parameters in both samples. Consistent with the Neuroticism finding, higher depressive symptoms (CES-D) were associated with lower mtDNA_{cn}. Finally, mtDNA_{cn} mediated the association between personality and mortality risk. To our knowledge, this is the first study to show a replicable association between mtDNA_{cn} and personality. It establishes a new connection between mitochondrial biology and emotional traits and states relevant to human health. Furthermore, we found that mtDNA_{cn} mediates the association between personality and mortality risk, thereby providing the first evidentiary support for the hypothesis that mtDNA_{cn} may be a biomarker of the biological process that explains part of the association between personality and mortality.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3542. Phenome-wide Mendelian randomization study of plasma triglycerides and 2,600 disease traits.

Authors:

J. Park¹, S. Bafna¹, I. Forrest¹, A. Duffy¹, C. Marquez-Luna¹, B. Petrazzini¹, H. Vy¹, D. Jordan¹, M. Verbanck², J. Narula¹, R. S. Rosenson¹, G. Rocheleau¹, R. Do¹; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²Université Paris Cité, Paris, France

Abstract Body:

Background: Triglycerides (TG) play vital roles in human physiology, ranging from energy storage and mobilization to inflammation and thrombosis. However, controversy mires whether the relationship between plasma TG levels and disease risk are causal or associative in most human diseases. This topic remains contentious even in atherosclerotic cardiovascular disease (ASCVD), as reflected by the conflicting results of two recent landmark trials, STRENGTH and REDUCE-IT. Establishing causality between TGs and phenome-wide disease risk could inform drug development of the novel repurposing opportunities and adverse effects of approved and emerging TG-lowering therapies, including icosapent ethyl and *APOC3* inhibitors. Clinical trials are currently evaluating TG-lowering therapies targeting *ANGPTL3*, *ANGPTL4* and *APOC3* for dyslipidemias.

Methods: We conducted a phenome-wide, two-sample Mendelian randomization (MR) analysis using an inverse-variance weighted (IVW) estimator to infer the causal effects of plasma TG levels on 2,600 disease traits in the European ancestry population of UK Biobank. We then externally tested 221 nominally significant associations ($p < 0.05$) for replication in an independent population from FinnGen. To account for potential horizontal pleiotropy and the influence of invalid instrumental variables, we performed sensitivity analyses using MR-Egger, weighted median, and MR-PRESSO methods. Finally, we applied multivariable MR controlling for correlated lipid fractions to isolate the independent effects of plasma TG levels.

Results: Our results identified 7 disease traits reaching Bonferroni-corrected significance in both the discovery ($p < 1.92 \times 10^{-5}$) and replication analyses ($p < 2.26 \times 10^{-4}$), supporting a causal relationship between plasma TG levels and ASCVDs, including coronary artery disease (OR 1.33, 95% CI 1.24-1.43, $p = 2.47 \times 10^{-13}$). We also identified 12 disease traits that were Bonferroni-significant in the discovery or replication analysis and at least nominally significant in the other analysis ($p < 0.05$), identifying plasma TG levels as a novel risk factor for 9 non-ASCVD disease traits, including uterine leiomyoma (OR 1.19, 95% CI 1.10-1.29, $p = 1.17 \times 10^{-5}$).

Conclusion: Using a phenome-wide Mendelian randomization approach, we identified 19 disease traits with positive causal associations with plasma TG levels, many of which are outside the context of cardiovascular disease. This suggests rationale for repurposing approved and emerging TG-lowering agents toward novel indications or reveals their adverse side effects.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3543. Phenotype integration improves power and preserves specificity in biobank-based genetic studies of MDD.

Authors:

A. Dahl¹, M. THOMPSON², U. An³, K-L. Georgii Hellberg⁴, L. Huang⁵, R. Border⁶, A. Pazokitoroudi⁷, M. Krebs⁴, V. Appadurai⁴, S-A. Bacanu⁸, T. Werge⁴, A. Schork⁹, J. Flint², K. Kendler⁸, S. Sankararaman³, N. Cai¹⁰; ¹Univ. of Chicago, Chicago, IL, ²Univ. of California, Los Angeles, Los Angeles, CA, ³UCLA, Los Angeles, CA, ⁴Inst. of Biological Psychiatry, Copenhagen, Denmark, ⁵Helmholtz Pioneer Campus, Munich, Germany, ⁶Univ. of California Los Angeles, Los Angeles, CA, ⁷UCLA, LA, CA, ⁸Virginia Commonwealth Univ, Richmond, VA, ⁹Inst. for Biological Psychiatry, Copenhagen, Denmark, ¹⁰Helmholtz Zentrum München, Neuherberg, Germany

Abstract Body:

Biobanks play an increasing role in human genetics because their large size substantially increases power to detect genetic associations. Biobanks often contain several phenotypes relevant to a given disorder, and researchers face complex trade-offs between shallow phenotypes (high sample size, low specificity) and deep phenotypes (low sample size, high specificity). Here, we focus on an extreme case: Major Depressive Disorder (MDD) in UK Biobank. Previous studies found that shallow and deep MDD phenotypes have qualitatively distinct genetic architectures, but it remains unclear which are optimal for scientific study or clinical prediction. We propose a new framework to get the best of both worlds by integrating information across MDD-relevant measures phenome-wide. First, we use phenotype imputation to increase sample size for deep MDD phenotypes, which dramatically improves GWAS power (increases #loci ~10 fold) and PRS accuracy (increases R² ~2 fold). Further, we find that the genetic architecture of the imputed deep MDD phenotypes remains highly specific to MDD by (1) validating our results using data from PGC, iPSYCH, and ATLAS and (2) leveraging recent and novel measures of pleiotropy based on GWAS summary statistics or PRS. We also develop a reciprocal approach to improve MDD-specificity of GWAS on shallow MDD phenotypes by adjusting for phenome-wide PCs. Finally, we study phenotype integration at the level of GWAS summary statistics, which improves GWAS and PRS power but introduces non-MDD-specific signals. Our work provides a simple and scalable recipe to improve genetic studies in biobanks by combining the large sample size of shallow phenotypes with the power and specificity of deep phenotypes, which addresses a crucial barrier to characterizing the etiology of heterogenous complex disorders.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3544. Phenotypic and genetic relationships between physical activity, sleep, and brain health.

Authors:

Y. Guo, B. Zhao, P. Paschou; Purdue Univ., West Lafayette, IN

Abstract Body:

Lifestyle choices are associated with brain structure, function, and health. Physical activity and sleep, two essential aspects of lifestyle, can be easily intervened and have been under clinical practice to prevent the incident or slow the progression of various brain dysfunctions. In addition, changes in brain correlates of mental health can potentially affect sleep regularity, willingness to physical activity, and physical performance. Using multiple data types from the UK Biobank, we aim to examine the phenotypic and genetic relationships between sleep, physical activity, and brain health. We expect to identify brain regions and networks that are strongly associated with the variation in physical activity and sleeping habits. We also seek to quantify possible genetic correlation and causality between physical activity, sleep, and brain health. In general, we hope this study can contribute to the current knowledge of the genetic architecture and biological networks involved in the beneficial effects of developing physically active and balanced lifestyle behaviors on brain health.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3545*. Phenotypic subtyping via contrastive learning reveals heterogeneity in the genetic architecture of traits.

Authors:

A. Gorla¹, J. Mefford², V. Ravi³, Y. Di^{4,5}, J. Wang³, T. Zhu^{4,5}, A. Alwan³, E. Halperin^{6,7,8,9}, J. Flint¹⁰, N. Zaitlen¹¹, E. Rahmani¹²; ¹Bioinformatics InterDept.al Program, Univ. of California, Los Angeles, Los Angeles, CA, ²Dept. of Neurology, Univ. of California, Los Angeles, Los Angeles, CA, ³Dept. of Electrical and Computer Engineering, Univ. of California, Los Angeles, Los Angeles, CA, ⁴CAS Key Lab. of Behavioral Sci., Inst. of Psychology, Beijing, China, ⁵Dept. of Psychology, Univ. of Chinese Academy of Sci., Beijing, China, ⁶Dept. of Computer Sci., Univ. of California, Los Angeles, Los Angeles, CA, ⁷Dept. of Computational Med., Univ. of California, Los Angeles, Los Angeles, CA, ⁸Dept. of Anesthesiology and Perioperative Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, ⁹Dept. of Human Genetics, Univ. of California, Los Angeles, Los Angeles, CA, ¹⁰Dept. of Psychiatry and Biobehavioral Sci., Brain Res. Inst., Univ. of California, Los Angeles, Los Angeles, CA, ¹¹Ctr. for Neurobehavioral Genetics, Semel Inst. for NeuroSci. and Human Behavior, Univ. of California, Los Angeles, Los Angeles, CA, ¹²Dept. of Electrical Engineering and Computer Sci., Univ. of California, Berkeley, Berkeley, CA

Abstract Body:

BACKGROUND: Unmodeled phenotypic heterogeneity reduces statistical power and reproducibility in GWAS, and it has been posited as a possible explanation for small observed effect sizes. Existing phenotype subtyping methods primarily rely on clinically observed heterogeneity or metadata clustering. However, these approaches are typically unsuccessful in mapping phenotypic variation to heterogeneity in the genetic basis of traits.

OBJECTIVE: We aimed to develop a method to define *de novo* phenotypic subtypes that are more likely to reveal genetic heterogeneity.

METHOD: We introduce Phenotype Aware Component Analysis (PACA), a novel model-based contrastive learning algorithm. Given case-control data of any modality that may harbor meaningful sub-phenotypic signals, PACA learns a gradient of variation unique to cases, while accounting for variation and imbalances of biological/technical confounders between cases and controls.

RESULTS: We compared PACA to several commonly used dimensionality reduction and contrastive learning algorithms. To do so, we pooled individuals with different traits from the UK Biobank and evaluated the methods in discriminating the presence of subtypes. We found that PACA is the only method that is both well-powered to capture phenotypic heterogeneity and calibrated in reporting no heterogeneity under the null. We then verified that PACA can be applied to data modalities other than genetics. Using the PsychENCODE data (N=1,321), we grouped transcriptomes from bipolar (BP) and schizophrenia (SZ) cases. PACA, which was blind to the BP/SZ labels, isolated a gradient that differentiates BP from SZ ($r=0.3$, $p=2.7e-14$) and reflects heterogeneity in the genetic architecture ($p=1.4e-6$, Subtest). Finally, we used PACA to define subtypes for traits in which heterogeneity is suspected but has not so far been demonstrated. Applying PACA to voice recordings of women with recurrent major depressive disorder (MDD; N=6,142) we identified two MDD subtypes. A regression analysis, conditioned on 10 genetic principal components, geographic location and age, demonstrated that the variation between the putative subtypes is associated with psychiatric illness and handedness. A follow up GWAS on one of the subtypes revealed an association with rs1430259 ($p=4.5e-07$), an intronic variant in *CNTNAP5*; importantly, neither this variant ($p=0.66$) nor any other was revealed in a standard case/control GWAS of MDD.

CONCLUSION: PACA is a calibrated and well-powered modality-agnostic method for learning sub-phenotypic variation that reflects genetic heterogeneity.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3546. PheWAS-based clustering of Mendelian Randomization instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment.

Authors:

L. Darrous^{1,2}, Z. Kutalik^{1,2,3}; ¹Univ. Ctr. for Primary Care and Publ. Hlth., Lausanne, Switzerland, ²Swiss Inst. of Bioinformatics, Lausanne, Switzerland, ³Dept. of Computational Biology - Univ. of Lausanne, Lausanne, Switzerland

Abstract Body:

Motivation:

Mendelian Randomization (MR) is a statistical method that estimates the causal effect between risk factors and common complex diseases using genetic variants as instruments. It has several assumptions, some of which are prone to violation when the genetic variants are pleiotropic. For example, within-sibling MR analysis reveals a small negative causal effect estimate between body mass index (BMI) and educational attainment (EDU) (-0.05 [95% CI: $-0.09, -0.01$]), however, MR analysis based on GWAS of unrelated samples yields a substantially larger estimate (-0.19 [$-0.22, -0.16$]) possibly due to confounding and pleiotropy.

Methods:

We investigated pleiotropic effects of the BMI-associated instruments across 407 traits in the UK Biobank and grouped these SNPs using K-means clustering based on their association profile. We then used MR to estimate the cluster-specific causal effects of BMI on EDU. Enrichment analysis enabled meaningful labelling of the obtained clusters. As a complementary approach, we also systematically scanned the 407 traits for potential confounders and mediators (via bi-directional MR) and included them in a stepwise multivariable MR (MVMR) to obtain the multivariable causal effect of BMI on EDU.

Results:

The clustering of 322 BMI-associated genetic instruments yielded six groups with distinct causal effect estimates on EDU. Notably, a cluster strongly enriched for socio-economic status (SES)-related traits (job type, time spent outdoors) yielded the largest BMI-on-EDU causal effect estimate (-0.48 [$-0.55, -0.41$]) whereas the cluster enriched for largest impact on lean-mass had the smallest estimate (-0.09 [$-0.13, -0.05$]), agreeing with within-sibling MR results. Repeating the analysis using a proxy trait for childhood BMI revealed four clusters with homogenous near-null causal effects on EDU. Reassuringly, none of these clusters were enriched for SES traits. Our follow-up analysis uncovered 19 potential confounder traits (including time spent watching television and past tobacco smoking). Four of these traits survived in the stepwise MVMR, massively attenuating the causal effect estimate of BMI on EDU (from -0.19 to -0.01).

Conclusion:

Well-powered GWAS studies have not only revealed wide-spread pleiotropy, leading to biased MR estimates, but also enabled us to identify and annotate subsets of exposure instruments with similar functions. Using our approach (of instrument clustering and MVMR with systematic confounder search), we have uncovered distinct causal effects and the presence of confounders of the BMI-educational attainment relationship.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3547. PLEIOVAR - Assessing Pleiotropic Effects of SNPs on Traits in Genome-wide Association Studies.

Authors:

O. Meirelles, L. Launer, D. Li, K. Tsai, D. Schlessinger; Natl. Inst. on Aging, Baltimore, MD

Abstract Body:

Genome-Wide Association Studies (GWAS) explore associations between individual DNA variants, usually single nucleotide polymorphism (SNPs), and a trait. Analyses have recently been extended to assess association between multiple SNPs and a trait, or between a single SNP and multiple traits. With “PLEIOVAR”, we add analyses for pleiotropic effects between groups of SNPs, defined for each gene region, and multiple traits. Using principal component analysis (PCA) of multiple SNPs vs. multiple traits, associations are assessed by a summary statistic for each gene, and its significance efficiently evaluated by the chi-squared distribution. We applied the program to analyses of levels of lipids -- LDL, HDL and triglycerides (TG) -- in 6,289 participants in the SardiNIA population study and replicated the results for the lipid traits in the Coronary Artery Risk Development in Young Adults (CARDIA) study (2,726 individuals) and the InCHIANTI cohort (1,151 individuals). Computational speed is fast, taking under 5 minutes on a single CPU to run SardiNIA on the three lipid traits and on 22,397 genes. After meta-analysis of all cohorts, PLEIOVAR identified a total of 40 significant gene associations, with 23 previously reported by GWAS (14 of them pleiotropic) and 13 candidate novel gene associations, mostly involving lipid metabolism and cardiovascular disease. Finally, we used our method as a novel way to score gene networks from Gene Ontology (GO) to effectively detect pleiotropic enriched gene networks, generating association scores more significant than GO and other gene network databases, which like GO, also base their p-values on Fisher’s exact test.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3548*. Polygenic prediction across populations is influenced by ancestry, genetic architecture, and methodology

Authors:

Y. Wang^{1,2}, M. Kanai^{2,1,3,4}, T. Tan⁵, M. Kamariza³, K. Tsuo^{3,1,2}, K. Yuan^{1,2}, W. Zhou^{1,2}, Y. Okada^{4,6}, the BioBank Japan Project, H. Huang^{1,2}, P. Turley⁷, E. Atkinson⁵, A. Martin^{1,2}; ¹Massachusetts Gen. Hosp., Boston, MA, ²Broad Inst. of MIT and Harvard, Cambridge, MA, ³Harvard Univ., Cambridge, MA, ⁴Osaka Univ., Suita, Japan, ⁵Baylor Coll. of Med., Houston, TX, ⁶RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, ⁷Univ. of Southern California, Los Angeles, CA

Abstract Body:

With large-scale efforts underway to increase diversity in genomics research, there is an urgent need to understand how best to leverage such resources to reduce gaps in polygenic risk score (PRS) performance between ancestries. However, few studies have comprehensively investigated how various factors, such as the ancestry composition and trait genetic architecture, affect the generalizability of PRS built from multi-ancestry GWAS (PRS_M). In this study, we analyze large-scale simulated and empirical data to investigate this issue.

Specifically, we simulated 560K individuals from European (EUR), East Asian (EAS), and African (AFR) ancestry, respectively. We used 365K EUR and 155K EAS from UK Biobank (UKBB) and Biobank Japan (BBJ) for real data analyses, including anthropometric and blood-panel traits. We constructed single-ancestry PRS (PRS_S) using GWAS with varying numbers of individuals from each ancestry and PRS_M using meta-analyzed EUR and EAS or AFR GWAS with various ancestry compositions. We also constructed local ancestry-based PRS using GWAS from the Tractor method on 4K African American from UKBB to utilize resources from recently admixed populations. PRS was initially built using P+T and evaluated in understudied non-EUR populations using partial R^2 relative to covariates only.

Using simulated data, we find that PRS_M including individuals of the target ancestry, outperform PRS_S. Notably, such accuracy gain is saturated at a lower proportion of target ancestry for less polygenic traits compared to more polygenic traits. We observe similar findings using real data. First, we find that using substantially smaller sample sizes (20%-80% fewer) in BBJ can achieve comparable accuracy to 320K EUR in UKBB. Second, we observe a significant improvement of average accuracy across traits of PRS_M over PRS_S using GWAS from EUR (0.027 vs. 0.021) and EAS (0.026 vs. 0.016), respectively. We note that such improvement is not always seen (~14% comparisons) especially when EAS only accounts for a relatively small proportion in the multi-ancestry GWAS. Finally, we find that local ancestry-based PRS can improve performance compared to using large-scale EUR-based PRS (0.025 vs. 0.010) especially for less polygenic traits when there are variants with large ancestry-specific effects.

Overall, our study provides insights into how ancestry composition and genetic architecture impact polygenic prediction across populations. With the benefits of leveraging understudied populations when constructing PRS using different strategies, we highlight the necessity of increasing genetic diversity to close the gap and achieve equitable use of PRS across ancestries.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3549. Polygenic risk scores for coronary heart disease in diverse populations using population-specific optimization.

Authors:

J. Smith¹, C. Tcheandjieu², O. Dikilitas¹, S. Clarke³, A. T. Hilliard⁴, S. Koyama⁵, K. M. Lee⁶, J. Lynch⁷, K. Ito⁸, K-M. Chang⁹, S. Kanoni¹⁰, P. Tsao¹¹, M. Kosel¹, D. Schaid¹, T. Assimes¹², I. Kullo¹; ¹Mayo Clinic, Rochester, MN, ²Gladstone Inst.s, San Francisco, CA, ³Stanford Universit, Stanford, CA, ⁴VA Palo Alto Hlth.care System, Palo Alto, CA, ⁵Broad Inst., Cambridge, MA, ⁶U.S. Dept. of VA, Bedford, MA, ⁷Salt Lake City VA Med. Ctr., Salt Lake City, UT, ⁸RIKEN Ctr. for Integrative Med. Sci., Yokohama City, Japan, ⁹Corporal Michael J Crescenz VA Med. Ctr., Philadelphia, PA, ¹⁰Queen Mary Univ. of London, Cambridge, United Kingdom, ¹¹VA Palo Alto Hlth.Care System/Stanford Univ., Palo Alto, CA, ¹²Stanford Univ Sch Med., Palo Alto, CA

Abstract Body:

Introduction: The performance of polygenic risk scores (PRSs) for coronary heart disease (CHD) varies across populations with the lowest performance observed among African (AFR) ancestry individuals. **Methods:** To improve CHD PRS performance across diverse populations, we used pruning and thresholding (P+T) to construct population specific as well as multi-population PRSs from GWAS summary statistics of European (EUR), African (AFR), Hispanic (HIS) and East Asian (EAS) populations including nearly a quarter of a million cases of CHD from the Million Veterans Program (MVP), the UK Biobank, Cardiogram+C4D, and Biobank Japan. The PRSs were tested and validated within each population using independent samples from the Million Veterans Program (MVP), the electronic Medical Records and Genomics (eMERGE) Network, and several NHLBI cohorts with data available through dbGaP. PRS performance was assessed using logistic regression analyses adjusted for age, sex, and principal components. The best performing PRS for each population was identified based on the point estimate of the highest OR observed per standard deviation (SD) increase in the PRS. These ORs were compared to the ORs obtained with the prior best performing PRS in Europeans (metaGRS). We also compared the predictive ability of P+T PRS with a multi-population PRS using a continuous shrinkage method implemented in PRS-CSx. **Results:** Population-specific P+T PRSs outperformed the metaGRS and the multi-population PRS with ORs (95% CI) of 1.24 (1.16-1.33) for AFR, 1.35 (1.31-1.40) for EUR, and 1.49 (1.27-1.40) for HIS populations in the MVP. However, confidence intervals of the multi-population P&T PRS and the population specific PRS had substantial overlap. Additional validation in multiple independent cohorts involving all populations combined, EUR, AFR, HIS, EAS, and South Asian (SAS) populations resulted in ORs (95% CI) per SD of 1.43 (1.40-1.47), 1.60 (1.55-1.66), 1.18 (1.13-1.22), 1.39 (1.25-1.54), 1.69 (1.61-1.76), and 2.75 (2.41-3.14), respectively. The analysis using the single multi-population PRS-CSx for CHD was also validated in AFR, EUR, and HIS populations with ORs (95% CI) of 1.20 (1.14-1.27), 1.41 (1.37-1.44), and 1.54 (1.35-1.75). **Conclusions:** Our population-specific CHD PRS outperformed metaGRS for AFR individuals, but the gap in predictive performance compared to EUR remains. Our work highlights the need for more diverse population representation at each stage of PRS development.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3550. Polygenic scores enable discovery of widespread genetic interactions associated with quantitative traits in the UK Biobank.

Authors:

L. Ferreira¹, S. Hu², S. Myers¹; ¹Univ. of Oxford, Oxford, United Kingdom, ²Novo Nordisk Ltd., West Sussex, United Kingdom

Abstract Body:

Although a huge number of human genetic associations have been discovered, genetic predictions overwhelmingly use additive polygenic scores (PGS) which ignore interactions between loci. Such interactions are expected based on evidence from model organisms but human examples are limited, with the vast number of potential interactions hampering discovery power.

We developed an approach to overcome this lack of power by testing for interactions between a SNP and *groups* of other variants, such as those in the PGS, and applied this to quantitative traits in the UK Biobank. We are motivated by functional networks, where a SNP altering behaviour of one gene can have downstream effects on many others. Our approach is robust to false positives due to nonlinear additive effects or locally clustered associations and is initialised by iteratively building a PGS accounting for all linear associations.

We find many distinct interaction signals: 148 loci interacting with the PGS for 53 traits. These include variants driving pathogenesis for diverse diseases at genes including *APOE*, *FTO*, *G6PC2*, *HFE*, *HLA-G*, *IL33*, *LDLR*, *PNPLA3*, *PRDM1*, *SHGB*, *TCF7L2* and *UBD*. As well as their well-studied direct impacts, our analysis therefore reveals that these SNPs also modify effects at other loci. Strikingly, some SNPs show no direct association with a trait but instead alter the predictive power of its PGS.

We developed a test to identify, at each variant interacting with the PGS, the subset of SNPs driving that interaction, revealing networks of interacting loci. This identified a known interaction between an *ABO* variant and a *FUT2* stop-gain mutation impacting alkaline phosphatase, and novel mutations involved in this interaction at *FUT6*, *TREH*, *PIGC*, *ASGR2* and *ZNF678*. A second highlight is an interaction between a variant upstream of *IL33* and a missense variant in *ALOX15* affecting eosinophil count. This may shed light on the aetiology of asthma: a recent study in mice found a role for *ALOX15* in reducing *IL33*-induced airway eosinophilic inflammation. Finally, we see interactions involving distinct traits (LPA and alanine aminotransferase) and distinct coding mutations at *APOE* that drive strongly reduced, and increased, Alzheimer's disease risk, respectively, and are interesting given recent suggestions of a link between Alzheimer's and liver function.

Our results challenge the paradigm of simply adding effect sizes across mutations and provide examples of functionally connected variants for many traits. They offer potential to improve PGS performance using non-linear terms and their cross-population transferability by identifying determinants of performance variability.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3551. Polygenic scores for insulin resistance are associated with brain volumes in the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program.

Authors:

C. Sarnowski¹, S. Gaynor², X. Jian³, F. Ammous⁴, T. H. Mosley⁵, S. Heckbert⁶, A. L. Fitzpatrick⁶, W. T. Longstreth⁶, J. C. Bis⁶, L. J. Launer⁷, J. Dupuis⁸, J. C. Florez⁹, M-F. Hivert¹⁰, J. A. Smith⁴, L. R. Yanek¹¹, P. Nyquist¹¹, D. C. Glahn¹², J. E. Curran¹³, J. Blangero¹³, R. A. Mathias¹¹, D. K. Arnett¹⁴, B. Psaty⁶, H. Lin¹⁵, A. Manning⁹, M. Fornage¹, J. I. Rotter¹⁶, S. S. Rich¹⁷, J. Meigs⁹, S. Seshadri³, A. C. Morrison¹, TOPMed Diabetes working group, TOPMed Neurocognitive working group; ¹Univ. of Texas Hlth.Sci. Ctr. Sch. of Publ. Hlth., Houston, TX, ²Regeneron Genetics Ctr., Regeneron Pharmaceuticals, Tarrytown, NY, ³Glenn Biggs Inst. for Alzheimer's and Neurodegenerative Diseases, Univ. of Texas Hlth.Sci. Ctr., San Antonio, TX, ⁴Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, ⁵Memory Impairment and Neurodegenerative Dementia Ctr. and Univ. of Mississippi Med. Ctr., Jackson, MS, ⁶Univ. of Washington, Seattle, WA, ⁷Intramural Res. Program Natl. Inst. on Aging, Bethesda, MD, ⁸Boston Univ. Sch. of Publ. Hlth., Boston, MA, ⁹Massachusetts Gen. Hosp., Boston, MA, ¹⁰Harvard Pilgrim Hlth.Care Inst., Boston, MA, ¹¹Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ¹²Boston Children's Hosp., Harvard Med. Sch., Boston, MA, ¹³Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX, ¹⁴Univ. of Kentucky, Lexington, KY, ¹⁵Univ. of Massachusetts Med. Sch., Worcester, MA, ¹⁶Lundquist Inst., Harbor-UCLA Med. Ctr., Torrance, CA, ¹⁷Univ. of Virginia, Charlottesville, VA

Abstract Body:

Background: Insulin resistance (IR) is a major risk factor for Alzheimer's Disease (AD) and has been associated with cognitive impairment, dementia, and neurodegeneration. Though the association between IR and AD has been explored genetically, the proportion of variance explained (PVE) by the IR genetic instruments has been limited, due to few genetic loci identified in genome-wide association studies (GWAS) of IR. **Methods:** We constructed polygenic scores (PSs) for fasting insulin (FI) based on the Trans-Omics for Precision Medicine (TOPMed) Freeze 9 whole-genome sequencing (WGS) data. We used PRS-CSx (Ruan *et al.*, *Nat Genet.* 2022) to generate ancestry-specific PSs (European, African, and Hispanic/Latino), with weights derived based on ancestry-specific UK Biobank reference panels and FI GWAS summary statistics adjusted for body mass index (BMI) (Chen *et al.*, *Nat Genet.*, 2021). We used MetaSubtract (Nolte *et al.*, *Eur J Hum Genet.* 2017) to remove the effect of TOPMed studies from the FI meta-analyses. We generated a multi-ancestry PS_{FI} by fitting a linear combination of the standardized ancestry-specific PSs that most accurately predicted HOMA-IR in five TOPMed cohorts (validation set, N~17k participants [34% European, 28% African, 38% Hispanic] without diabetes), adjusting for BMI. We then evaluated the association of the multi-ancestry PS_{FI} with HOMA-IR, AD, dementia, general cognitive function, and four brain volumes in eight TOPMed cohorts (testing set, ~14k participants [66% European, 22% African, 11% Hispanic]). Association analyses were performed using logistic or linear mixed-effect models adjusted for age, sex, study, 11 genetic principal components, and accounting for relatedness using a genetic relationship matrix. General cognitive function and brain volumes analyses were additionally adjusted for education and intracranial volume (ICV) respectively. We used a threshold of $P < 0.05 / N_{\text{traits}} / N_{\text{tests}} = 0.05 / 7 / 4 = 0.002$ to define an association as significant. **Results:** The multi-ancestry PS_{FI} was strongly associated with HOMA-IR (PVE_{Joint}=12%). No significant or suggestive association was detected for the multi-ancestry, the African, or the Hispanic PS_{FI} with any of the neurological outcomes. The European PS_{FI} was significantly associated with ICV ($P=7E-07$), and suggestively associated with lateral ventricular ($P=0.004$) and total brain volumes ($P=0.05$). **Conclusion:** By leveraging multi-ancestry and WGS data, we increased the PVE of the genetic instruments for IR and

ASHG 2022 Annual Meeting Poster Abstracts

confirmed the association of IR with brain volumes. The identified European-specific associations require further investigation. **Funding:** NIH K99AG066849-02

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3552. Polygenic transcriptome risk scores can translate genetic results between species.

Authors:

S. Sanchez-Roige^{1,2}, N. Santhanam³, Y. Liang³, A. Chitre⁴, D. Munro⁵, D. Chen⁶, R. Cheng⁴, J. Gao⁴, A. George⁷, A. Gileta⁴, K. Holl⁸, A. Hughson⁹, C. P. King⁷, A. C. Lamparelli⁷, C. D. Martin⁷, A. Garcia Martinez¹⁰, S. Mi³, C. St. Pierre^{11,4}, J. Tripi⁷, T. Wang¹⁰, H. Chen¹², S. Fligel⁹, K. Ishiwari⁷, P. Meyer⁷, L. Saba¹³, L. C. Solberg Woods¹⁴, O. Poleskaya⁴, A. A. Palmer⁴, H. Kyung Im³; ¹University of California San Diego, La Jolla, CA, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Univ. of Chicago, Chicago, IL, ⁴Univ. of California San Diego, La Jolla, CA, ⁵Scripps Res. and UC San Diego, San Diego, CA, ⁶Univ. of California, San Diego, San Diego, CA, ⁷Univ. at Buffalo, Buffalo, NY, ⁸Med. Coll. of Wisconsin, Milwaukee, WI, ⁹Univ. of Michigan, Ann Arbor, MI, ¹⁰Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, ¹¹Washington Univ. Sch. of Med., St. Louis, MO, ¹²Univ. of California Davis, Davis, CA, ¹³Univ. of Colorado, Aurora, CO, ¹⁴Wake Forest Univ., Winston-Salem, NC

Abstract Body:

Genome-wide association studies have demonstrated that most traits are highly polygenic; however, translating these polygenic signals into biological insights remains difficult. A lack of satisfactory methods for translating polygenic results across species has precluded the use of model organisms to address this problem. Here we explore the use of polygenic transcriptomic risk scores (PTRS) for translating polygenic results across species. Unlike polygenic risk scores (PRS), which rely on SNPs, PTRS use imputed gene expression for prediction, which allows cross-species translation to orthologous genes. We first developed RatXcan, which is a framework for transcriptome-wide association studies (TWAS) in outbred rats. Leveraging predicted transcriptome and genotype data from UK Biobank, and the genetically trained gene expression models from RatXcan, we scored more than 3,000 rats using human-derived PTRS for height and BMI. Strikingly, we found that these human-derived PTRS significantly predicted analogous traits in rats ($r = 0.08$, $P = 8.57 \times 10^{-6}$; $r = 0.06$, $P = 8.51 \times 10^{-4}$, respectively). The genes included in the PTRS were enriched for biological pathways including skeletal growth and metabolism and were over-represented in tissues including pancreas and brain. This approach facilitates experimental studies in model organisms that examine the polygenic basis of human complex traits and provides an empirical metric by which to evaluate the suitability of specific animal models and identify their shared biological underpinnings.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3553. PopMLvis: A Tool for Analysis and Visualization of Population Structure in Genome-Wide Association Studies

Authors:

M. Elsharif¹, K. Isufaj¹, K. Kunji¹, M. Saad²; ¹Qatar Computing Res. Inst. - Hamad Bin Khalifa Univ., Doha, Qatar, ²Qatar Computing Res. Inst., HBKU Research Complex, QCRI, Qatar

Abstract Body:

Background: Population genetics aims to analyze genetic differences within and among groups of individuals. These analyses are carried out using a variety of algorithms that assign an individual to a given population group based on its genetic background. Although there are several tools that perform population genetic analysis, there is a lack of a framework that allow an interactive visualization of several clustering and multi-dimension reduction tools such as Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), Admixture, PC-air, and K-means.

Methods: We propose *PopMLvis*, a user-friendly platform that offers a complete environment to interactively visualize population structure analysis of Genome-Wide Association Studies (GWAS) data. In particular, *PopMLvis* allows users to visualize their PCA and Admixture data combined, make interactive cluster inference, and outlier detection (e.g., Isolation Forest, Local Outlier Factor, and OneClassSVM). It also allows running a PCA that accounts for genetic relatedness between individuals (PC-air). *PopMLvis* supports dimensionality reduction (PCA and t-SNE), fuzzy C-means, and hierarchal clustering, as well as outlier detection algorithms. Figures and text files can be downloaded with all exploratory results and clustering inference with great flexibility. *PopMLvis* was mainly written in Python with the addition of the R package PC-air.

Availability: *PopMLvis* is freely available and can be accessed online from any web browser at <https://popmlvis.qcri.org>. In addition, for data privacy issues, the source code of *PopMLvis* can be downloaded and installed from GitHub, where a standalone version can run on user laptops or servers.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3554. Population diversity in global reference genome cohorts

Authors:

V. Skrahina¹, G. Oprea¹, N. Ameziane¹, H. Cheema², A. Rolfs^{1,3}; ¹Arcensus, Rostock, Germany, ²Childrens' Hosp., Lahore, Pakistan, ³Univ. of Rostock, Rostock, Germany

Abstract Body:

Whole Genome Sequencing (WGS) is the most comprehensive genetic test available with the highest diagnostic yield. The major diagnostic challenge is the “uniformity” of the currently used human genome reference (hg38), being mainly Caucasian- based, the genomic diversity is restricted to the European population (>85%).

To increase the diagnostic yield, we have gathered under-represented genomic and clinical datasets, especially originating from consanguineous families. We have analysed more than 1200 Pakistani (71%), Albanian (10%) and Saudi Arabian (19%) probands by WGS. In both cohorts, at least 93% of the patients are younger than 18 years and the most predominant symptom are neurodevelopmental delay, seizures, and intellectual disability.

A high diagnostic rate (pathogenic and/or likely pathogenic variants) was achieved for the Pakistani (78%), Saudi Arabia (66%) and Albanian cohort (65%), with the following percentage of autosomal recessive disorders 85%, 56% and 26%, respectively. The most frequently diagnosed diseases are metabolic disorders (14%), intellectual neurodevelopment disorders (9%) and epileptic encephalopathy (6%). The remaining disorders are representing a highly diverse group of more than 450 different diseases. Importantly, for most frequent diseases (affected genes: *BTD*, *GALT*, *ATP7B*, *CDKL5*, *NAGLU*) causative treatments exist. 4-10% of all patients from these ethnicities are suffering from two or more monogenic disorders.

Studying consanguineous cohorts speeds up the identification and characterization of novel genes, indicating clinically relevant or irrelevant genomic regions, and contribute simultaneously to the global genome reference setup by including genomic diversity. Interestingly, for our cohorts more than 20% of the clinical genome is represented by novel variants (not described elsewhere), of which 4.9% predicted to impact protein function. 80% of variants that impact protein structure are frequent in our Pakistani and Saudi-Arabia population and classified therefore as benign or likely benign. Without this valuable information, these patients would have been misdiagnosed.

The application of WGS as a routine first-line molecular diagnostic approach has a significant impact on clinical management of patients and if treatment is not available, the inclusion of such patients in ongoing clinical trials, will speed drug development. Understanding the global human genome diversity leads to increased knowledge about the genetic variants causing diverse phenotypes and those being part of the global “variome”.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3555*. Power Window: a power-based sliding window method to identify the rare and novel variants underlying gene-based association signals

Authors:

A. Bolze¹, K. Barrett¹, J. Grzymalski², W. Lee¹, N. Washington¹, E. Cirulli¹; ¹Helix, San Mateo, CA, ²DRI, Reno, NV

Abstract Body:

Systematic determination of which rare genetic variants are pathogenic for a given phenotype remains a major challenge, even when there is an established association between the phenotype and rare variants in the gene. We have developed Power Window, a novel technique to identify the regions of the gene where rare variants are statistically significantly associated with a trait using population-scale clinico-genomic datasets. Power Window is a sliding window technique for rare variants that slides not by the number of variants or the number of nucleotides, but rather by the number of variant carriers—and thus the statistical power. The method can be used to focus on specific types of variants, such as loss of function (LoF) or coding, or specific parts of the gene, such as those expressed in different tissues. It can also identify regions within a gene that have opposite directions of effect on a phenotype.

We use Power Window to build regional LoF and coding models for well-established gene-level disease associations in a training set of 300k exomes from the UKBiobank (UKB). We then test our models in two additional cohorts: an unrelated set of 128k exomes from the UKB and 30k exomes from the Healthy Nevada Project (HNP). Importantly, the variants tested in these replication datasets include novel variants that were not observed in the training set, but that occur in the same set of regions implicated in the training set. We find that 99% of these models show the same direction of effect in the test set, and 57% demonstrate statistically significant replication. The significant models retain a mean of 54% of the variant carriers in the gene (range 2-98%) and often drive a dramatic improvement in the effect size, especially for coding variants: 50% of the significant coding models for quantitative traits saw their effect size improve by at least 40% over the effect size observed when rare variants were aggregated across the whole gene; for binary traits, 50% of the significant coding models more than doubled the associated odds ratio. Power Window showcases that even in the absence of family data or functional tests, in some cases computational predictions alone can determine, with high accuracy, which novel coding variants will have high penetrance in a population, unlocking new potential for population genetic screening.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3556. Practical screening for differences in predictive value of polygenic scores across the phenotypic range

Authors:

J. Mefford¹, M. Smullen², F. Zhang¹, R. Border¹, A. Dahl³, N. Zaitlen¹; ¹Univ. of California Los Angeles, Los Angeles, CA, ²UMass, Worcester, MA, ³Univ. of Chicago, Chicago, IL

Abstract Body:

Polygenic scores (PGS) show promise as tools for basic and translational science. Genetic risk predictions for heritable diseases may finally bring knowledge gained from large genetic association studies to the clinic with personalized disease screening and care management. PGS have also been used as instruments in social science to help identify causal connections between biological or social factors and outcomes such as educational attainment.

An underlying assumption of applications of PGS is the uniformity of predictive accuracy across the phenotypic range. We demonstrate with simulations and analytic derivations how G×E and G×G affect PGS predictive value. For example, we know that height is quite heritable, and we can make PGS that are well correlated with height. However, if a portion of a target population faces food insecurity, the severity of under-nutrition may strongly affect height and limit PGS predictability - particularly in the low end of the height range. PGS with such variable predictability will bias downstream applications.

Quantile regression followed by trend tests of quantile-specific effect sizes has been used to test for non-uniform PGS accuracy. We show that such tests of trends or non-uniformity can be inflated under a range of analyses. Instead, we focus on the magnitude of differences in PGS predictive value. We use bootstrap confidence intervals and 1-sided significance tests to identify traits where quantile-specific PGS predictions have substantially different accuracy than the mean PGS accuracy.

We conduct a comprehensive survey of PGS predictability for traits in the UK Biobank and observe that this issue is severe and widespread, with 46% of traits having ratios of quantile specific to mean PGS predictability significantly greater than 1.2 ($p < 0.05$) and 54% of traits having quantile specific predictability significantly less than 1/1.2.

We see a number of qualitative patterns of PGS effect sizes across covariate adjusted phenotype quantiles. Some traits like BMI have a wide range of PGS predictability, with relative effects - ratios of quantile specific PGS effects to the standard, mean PGS effect - ranging from 0.38 (95% CI 0.36, 0.58) at the 0.05 BMI quantile to 1.90 (1.73, 2.34) at the 0.95 BMI quantile. That is, predicting BMI is much less accurate for lower BMI individuals than for higher BMI individuals using a global BMI PGS.

Such non-uniformity of PGS predictability could result in inequitable disease risk assessment or biased PGS-based research in health and social sciences. Our approach provides a straightforward routine screening for the magnitude of the non-uniformity of PGS effect sizes in current and future studies.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3557. Predicting cancer risk from germline next-generation sequencing data using a novel context-based variant aggregation approach.

Authors:

Z. Guan, C. B. Begg, R. Shen; Mem. Sloan Kettering Cancer Ctr., New York, NY

Abstract Body:

Twin studies have shown that most common cancer types have a substantial heritable component. However, known risk variants (SNPs from GWAS and pathogenic variants in known cancer predisposition genes) explain only a limited proportion of the estimated heritability of cancer. It has been hypothesized that much missing heritability lies in rare variants not captured by SNP arrays. Rare variants need to be aggregated to achieve sufficient statistical power for detection. We propose a novel context-based variant aggregation approach for extracting signals from rare variants detected through germline whole-exome and whole-genome sequencing. Many studies have shown that the distributions of the genomic, nucleotide, and epigenetic contexts of somatic variants in tumors are informative of cancer etiology and site of origin. Recently, a new direction of research has focused on extracting signals from the contexts of germline variants and evidence has emerged that patterns defined by the nucleotide contexts of germline variants are associated with oncogenic pathways, histological subtypes, and prognosis. It remains an open question whether aggregating germline variants based on these contexts can improve cancer risk prediction. Using germline whole-exome sequencing data from over 200,000 individuals from the UK Biobank and whole-genome sequencing data from over 2,000 individuals from the Pan-Cancer Analysis of Whole Genomes Consortium, we investigate the predictive value of meta-features aggregating rare variants based on their genomic, nucleotide, and epigenetic contexts for distinguishing cancer cases from controls and for predicting tumor subtypes (such as homologous recombination deficiency). We compare the performance of risk models based on known risk variants and risk models that additionally include the meta-features.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3558*. Predicting ExWAS results from GWAS data: A shorter path to causal genes.

Authors:

K. Liang, Y. Chen, S. Yoshiji, J. Farjoun, V. Forgetta, B. Richards; McGill Univ., Montreal, QC, Canada

Abstract Body:

Genome-wide association studies (GWAS), have identified thousands of loci associated with susceptibility to disease, yet the causal genes within these loci are largely unknown. Identifying these causal genes is necessary to enable for deeper understanding of the disease and for genetics-based drug development. Several algorithms have been developed to address this problem, such as the Effector Index (Ei), Locus-2-Gene (L2G), Polygenic Prioritization score (PoPs), and the Activity-by-Contact score (ABC). However, their efficacy as compared to the more expensive, but more precise Exome-wide Association Study (ExWAS) is unclear. Here, we quantified the performance of these algorithms by evaluating their ability to identify ExWAS significant genes in nine traits. We found that Ei, L2G and PoPs can identify ExWAS significant genes with an area under the precision recall curve much greater than their respective baselines (Ei: 0.53 vs 0.06, L2G: 0.46 vs 0.07, PoPs: 0.21 vs 0.06, ABC: 0.18 vs 0.08). Furthermore, we found that for every 1SD increase in the normalized scores, there is an associated 1.4-4.3 fold increase in the odds of a gene reaching exome-wide significance (Ei: 4.3, L2G: 2.7, PoPs: 2.1, ABC: 1.4). Overall, we see that Ei, L2G, and PoPs can accurately anticipate ExWAS results, making these promising techniques for drug target discovery when well-powered ExWAS data are not readily available. In addition, because these gene prioritization metrics do not rely on ExWAS data, they can serve as an independent source of evidence that can complement ExWAS results when they are available. These metrics are therefore cost-effective ways to use the large collection publicly available GWAS data to obtain additional evidence regarding the causality of a gene and help identify potential drug targets.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3559. Predicting how our genetic variants regulate splicing and impact human phenotypes.

Authors:

H. Jacobs¹, C. Burge¹, H. Finucane²; ¹MIT, Cambridge, MA, ²Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

A fundamental goal of human genetics is to predict which genomic variants impact molecular phenotypes to better understand disease etiologies. One molecular phenotype is pre-mRNA splicing, a process that determines which RNA sequences will be encoded into protein. Despite much progress, the functional impact of variants on splicing is under-studied compared to other molecular phenotypes, such as gene expression.

Here, we leverage fine-mapped splice quantitative trait loci (sQTLs) (Barbeira et al 2021) in the Genotype-Tissue Expression (GTEx) Consortium and construct a predictor that differentiates between SNPs with high vs low posterior probability of causally impacting splicing. This predictor, which we call splice modifier score (SMS), utilizes features such as epigenetic marks, conservation, evidence of splicing factor binding, splice site strength (via MaxEnt), and distance to nearest GTEx splice site. SMS predicts a held-out test set with an AUROC of 0.94. We find distance to nearest GTEx splice site greatly improves recall, while additional features generally improve precision, suggesting sQTL variants interact locally with core splicing machinery. Using SMS rather than sQTL PIP improves colocalization with UKBB GWAS variants. Using a probability cutoff of 0.6 (precision: 0.7), SMS prioritizes 57 putatively causal variants ($PIP \geq 0.5$, excluding protein truncating variants) in the UK Biobank (UKBB) as likely splice-modifying variants, compared to 43 identified by sQTL colocalization with the same probability cutoff. We also compared splice sites identified in GTEx with splice sites from reference transcript annotations (GENCODE). In GTEx, we determine there are ~450k biologically interpretable splice sites, of which ~170k do not exist in GENCODE. GTEx-specific splice sites have a similar distribution of MaxEnt scores to GENCODE splice sites. These novel splicing events tend to be both tissue-specific and rare in the GTEx population: 20% of GTEx-specific splice sites are unique to a single person in GTEx, compared to 5% of splice sites shared between GTEx and GENCODE. Incorporating GTEx-specific splice site annotations improves recall of UKBB high PIP ($PIP \geq 0.9$) variants in splice sites by 1.2x. GTEx splice sites intersect with high PIP UKBB variants in many tissues, demonstrating that leveraging more RNAseq data from the human population can help uncover splicing relevant to complex traits. This work demonstrates the utility of studying splicing to interpret non-coding variants in disease.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3560. Prediction of atrial fibrillation and stroke using machine learning models.

Authors:

A. Papadopoulos¹, D. Harding¹, G. Slabaugh¹, P. Deloukas², E. Marouli³; ¹Queen Mary Univ. of London, London, United Kingdom, ²Queen Mary Univ. London, London, United Kingdom, ³Barts and The London Sch. of Med. and Dentistry Queen Mary Univ. of London, London, United Kingdom

Abstract Body:

We employed machine learning (ML) approaches to evaluate 2,716 clinical features and disease phenotypes available in the UK Biobank as predictors for Atrial Fibrillation (AF) risk. After quality control, 308 features were selected for analysis in 23,095 prospective AF cases and equal number of controls. Different ML methods were employed, including LightGBM, XGBoost, Random Forests (RF), Deep Neural Networks (DNN,) and Logistic Regression with L1 penalty (LR). In order to eliminate the "black box" character of the tree-based ML models, we employed Shapley-values (SHAP), which are used to estimate the contribution of each feature to AF prediction. The area-under-the-curve (AUC) values and the 95% confidence intervals (CI) per model were: 0.751 (0.742, 0.759) for XGBoost, 0.750 (0.742, 0.759) for LightGBM, 0.740 (0.731,0.749) for SVM, 0.739 (0.730, 0.748) for LR, 0.725 (0.716, 0.734) for RF and 0.669 (0.659, 0.678) for DNN. Considering the running time, memory and stability of each algorithm, XGBoost was the best performing among those examined. De-Long's test showed that there is statistically significant difference in the AUCs between XGBoost and the non-tree-based models. The two most important features identified for XGBoost, using SHAP analysis, are the genetic risk score (GRS) of AF and age. As expected, the AF GRS had a positive impact on the model output, i.e. a higher AF GRS increased AF risk. Similarly, age also had a positive impact increasing AF risk. Secondary analysis was performed for the individuals who developed stroke after AF diagnosis, employing 322 features in 3,219 prospective cases of people who developed stroke after AF, and equal number of controls in UK Biobank. The AUC values and the 95% CI per model were: 0.618 (0.593, 0.643) for SVM, 0.613 (0.585, 0.638) for LightGBM, 0.611 (0.584, 0.637) for RF, 0.602 (0.576, 0.629) for XGBoost, 0.571 (0.544, 0.598) for DNN, 0.544 (0.517, 0.571) for LR. De-Long's test showed that there is evidence for statistically significant difference between the SVM and the LR models. Using SHAP analysis for the tree-based model that performed best, LightGBM, the two most important features are cystatin C and systolic blood pressure. De-Long's test showed that there is evidence for statistically significant difference between LightGBM and the current clinical tool for stroke prediction in AF patients, CHA2DS2-VASc, which has AUC and 95% CI of 0.608 (0.581, 0.634).

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3561. Prediction of congenital heart disease subgroups associated with copy number variants using a multinomial machine learning-based classifier.

Authors:

J. S. Penaloza^{1,2}, B. Moreland¹, B. J. Landis³, L. R. Elmore³, G. C. Geddes³, J-h. Lin⁴, S. Yatsenko⁴, C. W. Lo⁴, S. Wechsler⁵, S. R. Lalani^{6,7}, V. Garg^{1,2}, J. C. Hodge³, S. Ware³, K. L. McBride^{1,2}, P. White^{1,2}; ¹Nationwide Children's Hosp., Columbus, OH, ²Ohio State Univ. Coll. of Med., Columbus, OH, ³Indiana Univ. Sch. of Med., Indianapolis, IN, ⁴Univ. of Pittsburgh, Pittsburgh, PA, ⁵Emory Univ. Sch. of Med., Atlanta, GA, ⁶Baylor Coll. of Med., Houston, TX, ⁷Texas Children's Hosp., Houston, TX

Abstract Body:

Congenital Heart Disease (**CHD**) is a global health burden that is a major cause of infant mortality. Copy number variants (**CNV**) are a common cause of CHD. Some recurrent CNVs are associated with specific CHD, but frequently, patients who have the same CHD are diagnosed with distinct CNVs whose contribution to CHD pathogenesis remains unclear. To address this challenge, in this study we functionally annotated clinically validated CNVs associated with isolated CHD by applying machine learning.

Here we leveraged our access to data from the Cytogenomics of Cardiovascular Malformations (**CCVM**) Consortium, including information on demographics, diagnosis, and clinical chromosomal microarray analysis of 1353 patients with CHD. The isolated CHD subgroups we focused on are (Ventricular/Atrial Septal Defect (**VSD/ASD**), Right Ventricular Obstruction (**RVOTO**), Left Ventricular Outflow Tract Obstruction (**LVOTO**), Heterotaxy (**HTX**), Conotruncal Defect (**CTD**), Atrioventricular Septal Defect (**AVSD**), and Anomalous Pulmonary Venous Return (**APVR**).

We developed a multinomial machine learning classifier to identify CNV patterns, both distinct and shared, between isolated CHD subgroups, among features such as genomic region, biological processes related to genes within the CNV, and other attributes. Three tree-based algorithms were compared: decision trees, random forests, and XGBoost. The Random Forest (**RF**) model performed the best, attaining a F1 score of 0.60, which is 5X higher than the baseline performance threshold of 0.125 for an 8-way balanced classification. We used Shapley Additive Explanation (**SHAP**) values to analyze the features that most contribute to the classification of the CHD subgroups. From the global SHAP plots we found CNV type, brain related processes, cell migration, and cell adhesion are important parameters determining the overall classification. Class-specific SHAP plots, illustrate the positive and negative impact each attribute has on predicting a CHD subgroup. As an example, according to our RF model, CNVs that appear with HTX tend to involve CNV duplication, and impact to cell adhesion and cell projection with distinguishing features in tissue morphogenesis and positive regulation of intracellular signaling transduction. In contrast, for the VSD/ASD subgroup, we see larger weight placed on central nervous system development and neurogenesis biological processes. The class-specific features vary in their impact on the model with respect to the CHD subgroup, amongst a combination of shared and distinct attributes. These findings demonstrate machine learning is an effective tool for functional annotation of CNVs associated with CHD.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3562. Prioritization of causal genes from genome-wide association studies by Bayesian data integration across loci.

Authors:

Z. Mousavi¹, M. Arvanitis¹, T. Duong¹, J. Brody², A. Battle¹, N. Sotoodehnia³, A. Shojaie³, D. E. Arking⁴, J. S. Bader⁵; ¹Johns Hopkins Univ., Baltimore, MD, ²Univ of Washington, Seattle, WA, ³Univ. of Washington, Seattle, WA, ⁴Johns Hopkins Univ Sch. of Med., Baltimore, MD, ⁵Johns Hopkins Univ, Baltimore, MD

Abstract Body:

Genome-wide association studies (GWAS) have been effective in identifying genetic variants, usually single-nucleotide polymorphisms (SNPs) associated with human traits, including disease and disease risk. The variants identified by GWAS, or causal variants in linkage disequilibrium with the variants, often affect the regulation of nearby genes or the activity of their protein products. A GWAS locus can span many genes; however, and prioritizing which gene or genes in a locus are most likely to be causal remains a challenge. Better prioritization and prediction of causal genes could reveal disease mechanisms and suggest interventions. We describe a new Bayesian inference method to prioritize candidate genes across loci identified by large-scale GWAS. This method prioritizes genes as causal by integrating information about individual loci, including existence of Mendelian genes or expression quantitative trait loci (eQTL) in the region, and network connectivity as provided by protein-protein and gene regulatory interaction data. Our analysis of recent GWAS data for well-studied traits such as cardiovascular electrophysiology and cancer predisposition suggest that for 70-80% of genome-wide significant SNPs, the closest gene is most likely to be causal. For the remaining loci, our predictions improve upon the naive choice of the closest gene, assessed by better enrichment for relevant biological annotations, and provide predictions for loci that lack eQTL or other informative features. A strength of the method is the ability to prioritize genes based on cross-locus information at loci where there is no evidence for colocalization. Examples in which the gene is selected based on cross-locus information rather than distance to the SNP or colocalization include NKX2-5 for cardiovascular electrophysiology, and MRKN1 for breast cancer predisposition.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3563. Prospective analysis of disease incidence and progression with genetic, clinical and lifestyle risk factors

Authors:

W. Wang, N. Eriksson, M. McIntyre, R. Bagur Quetglas, 23andMe Research Team, A. Auton, S. Shringarpure; 23andMe, Sunnyvale, CA

Abstract Body:

Large and diverse databases and improved methodology allow advanced understanding in prediction of disease incidence and progression with genetic, clinical and lifestyle risk factors.

We constructed two non-overlapped cohorts, discovery cohort and prospective cohort, from 23andMe, Inc. research-consented participants, a large cohort of research-consented individuals with genotype data obtained via a direct-to-consumer product and phenotypic data obtained from online self-report surveys. We built polygenic risk score (PRS) models across multiple populations in the discovery cohort, and then compared cumulative incidence rates among different genetic risk groups for multiple phenotypes (12 in European population, 9 in Latino population and 3 in African American population) in the prospective cohort. We also investigated the interplay of PRS with clinical and lifestyle risk factors in disease incidence of type 2 diabetes (T2D) and coronary artery disease (CAD).

The cumulative incidence rates, in the first follow-up year, were higher in participants at high genetic risk compared to all participants across multiple populations. Among participants with clinical risk, we observed that those at high genetic risk had higher disease incidence rates. We also found that low lifestyle risk was associated with reduction in disease incidence among participants at high genetic risk. Take the first follow-up year incident T2D as an example, the cumulative incidence rate was 1.19% (95% CI [1.12%, 1.27%]) in participants within the top 10% PRS risk and 0.48% (95% CI [0.46%, 0.50%]) among all participants in European population; 1.18% (95% CI [1.04%, 1.32%]) in participants within top 20% PRS and 0.56% (95% CI [0.52%, 0.61%]) among all participants in Latino population; 1.52% (95% CI [1.20%, 1.85%]) in participants within top 20% PRS and 0.95% (95% CI is [0.83%, 1.06%]) among all participants in African American population. Among the participants who reported with high blood sugar in European population, the T2D incidence rate was 7.27% (95% CI [6.44%, 8.17%]) in the high genetic risk group, compared to 5.29% (95% CI [4.97%, 5.60%]) in the overall high blood sugar group. Within the high genetic risk group, participants at low lifestyle risk had a lower incidence rate, 0.42% (95% CI [0.31%, 0.53%]), compared to the overall incidence rate, 1.19% (95% CI [1.12%, 1.27%]), in European population.

In this study, we found that our PRS model predicted disease incidence across multiple populations, and could identify individuals with higher risk of disease progression. The analysis also provided insights on potential disease risk modifications based on lifestyle.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3564. Protecting privacy in polygenic risk score calculation by homomorphic encryption

Authors:

H. Kim¹, C. Kim¹, B. Han²; ¹Seoul Natl. Univ., Seoul, Korea, Republic of, ²Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Abstract Body:

Genome-Wide Association Studies (GWAS) have revealed many small effects of genetic variants on common complex diseases and opened an opportunity for calculating polygenic risk score (PRS) of individuals from their genotype data. Many individuals now have access to their genotype data due to the prevalent direct-to-customer genetic test, and numerous health companies are providing PRS calculation services conditioned on that the users upload their genetic data. However, this trend raises serious privacy and security concerns. Even if the genotype data is encrypted during transmission, the uploaded genetic data must ultimately be decrypted in order to calculate PRS. If the decrypted genotype data is hacked, disastrous situations can occur, such as the individual being uniquely identified. To overcome this privacy issue, we developed SecurePRS (Secure Polygenic Risk Score), a systematic framework to calculate PRS using homomorphic encryption technique. In our framework, even the PRS service provider needs not decrypt the data and does not have access to the PRS results, thereby providing a full privacy protection. For better performance, we applied four representative linear PRS models (PRScs, SBLUP, LDpred and P+T) in an unencrypted state and used the estimated weights to calculate PRS in an encrypted state.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3565. Proteome-Wide Association Analysis in the Women's Health Initiative Study

Authors:

B. Chen¹, C. Lee¹, A. Tapia¹, A. Reiner², H. Tang³, C. Kooperberg⁴, Y. Li⁵, L. Raffield⁶; ¹Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ²Univ of Washington, Seattle, WA, ³Stanford Univ. Sch. of Med., Stanford, CA, ⁴Fred Hutchinson Cancer Res. Ctr., Seattle, WA, ⁵Univ North Carolina, Chapel Hill, NC, ⁶UNC - Chapel Hill, Chapel Hill, NC

Abstract Body:

Studying the genetically regulated component of protein abundance can improve our knowledge of molecular mechanisms underlying cardiovascular diseases (CVD). In Proteome-Wide Association Studies (PWAS), genetic variants near the protein-coding gene (± 1 Mb), also known as *cis* or local SNPs, are used to predict the protein level. These predicted protein levels are then associated with phenotypes of interest. However, proteins can be regulated through variants outside of the local region. Thus, we propose an intermediate GWAS step to identify protein quantitative trait loci (pQTL) throughout the genome and select variants for consideration in PWAS model training. This allows for the inclusion of *trans* or distal SNPs in protein level prediction models. We here assess prediction of 552 proteins measured using targeted Olink panels in 1,002 individuals from the Women's Health Initiative (WHI) with complete proteomics data across all panels and whole genome sequencing data from TOPMed freeze 9. We split the individuals equally into a GWAS set (for identification of pQTL), elastic net training set using all nominal pQTLs genome-wide ($p < 0.0001$), and a testing set, for comparison of predicted and inferred protein levels. We compared the correlation between inferred and predicted protein level in held out testing set data (testing r^2) using this proposed approach, which incorporates both local and distal SNPs, to the testing r^2 using only local SNPs in elastic net model training (in two thirds of the total sample size, instead of one third, since there is no need for an initial GWAS to detect pQTLs genome-wide). The two approaches resulted in similar testing r^2 's for most proteins (despite the increased sample size for model training using local SNPs only), but some proteins showed a noticeable increase in testing r^2 with our method compared to using local-only SNPs. For example, for Cartilage Acidic Protein 1 (CRTAC1) the testing r^2 increased from 0.101 to 0.351 due to several large effect distal pQTLs. In total, 160 proteins have a testing $r^2 > 0.05$ in this relatively modest sample size, with a max testing r^2 of 0.71. In the future, we aim to assess associations of predicted proteins with CVD outcomes and quantitative phenotypes in WHI participants with sequencing data only, assess whether predicted protein associations are driven by local and distal SNPs or just local SNPs, and work to replicate our results using other datasets, such as UK Biobank.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3566. Publicly Available Privacy-preserving Benchmarks for Polygenic Prediction

Authors:

M. Witteveen, E. Pedersen, F. Privé, D. Speed, B. Vilhjalmsón; Aarhus Univ., Aarhus, Denmark

Abstract Body:

In recent years, many different approaches for creating polygenic scores have been proposed and this trend shows no sign of abating. Most of these new polygenic score methods claim to outperform previously proposed methods, making it confusing for users to choose an appropriate PRS approach. This apparent performance paradox is due to challenges in properly determining what approaches are superior, because different datasets, quality control, and preprocessing steps are used to determine performance. As shown previously, these steps can have a significant impact on the overall prediction accuracy for polygenic score methods (Privé et al., bioRxiv 2022), and thus lead to an incomplete comparison of PRS methods.

In the related field of Machine Learning this problem has been addressed by the introduction of publicly available benchmark datasets, which allow for fair comparison between approaches, which has been vital to the advancement of the field. However the creation of such publicly available benchmark datasets in genetics has, until now, proved challenging because of privacy concerns.

As a solution, we present a privacy-preserving and publicly available benchmark for polygenic prediction, which allows researchers to both train and test polygenic prediction methods using only linkage disequilibrium (LD) information and summary statistics from genome-wide association studies, thus preserving privacy.

Using UK biobank data and a diverse set of 8 external summary statistics, we show that with our approach we estimate the squared correlation prediction accuracy almost perfectly. The method only requires linkage disequilibrium (LD) information and genome-wide association summary statistics, and no individual level data. We further used the benchmark data to compare a variety of polygenic scoring methods including PRS-CS, LDpred2, and SBayesR and observe a remarkable concordance, with perfect recovery of model rankings and almost perfect correlation of model performance measures.

This new benchmark for PRS model development and comparison will be made publicly available for researchers to download. We believe this benchmark can be used as a clear and unbiased standard for future polygenic score methods to compare against.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3567. PUMICE-S: Prediction Using Models Informed by Chromatin conformations, Epigenomics and Summary Statistics

Authors:

L. Wang; Penn State Coll. of Med., Hershey, PA

Abstract Body:

Transcriptome-wide association studies (TWAS) has gained popularity in many complex trait studies. TWAS first create gene expression models from an eQTL dataset that measures both gene expression and genetic variants. It then predicts gene expression in a GWAS dataset and correlates predicted gene expression with phenotypes of interest. The power of TWAS critically depends on training sample size for deriving gene expression prediction models. The largest available eQTL datasets are only available as summary statistics, such as eQTLGen. It is important to extend current TWAS approaches and use summary statistics to predict gene expression levels. Here, we present a new method PUMICE-S (Prediction Using Models Informed by Chromatin conformations, Epigenomics, and Summary statistics). PUMICE-S can take advantage of summary-statistics from a much larger dataset. It also integrates 3D genomic data to properly define cis-regulatory regions and uses epigenetic annotation to prioritize causal variants. To illustrate the advantage of PUMICE-S for analyzing large eQTL summary statistics, we applied PUMICE-S to eQTLGen summary statistics with $N=31,684$ and compare it with classical methods, including PrediXcan, TIGAR and UTMOST, trained on GTEx individual-level data. Because of the large sample size, we observed PUMICE-S results in an average 76.8%, 55.6%, and 41.8% increase of in the number of significant prediction models, where the predicted and measured gene expressions are significantly correlated. It will also lead to 99.7%, 81.7%, and 41.3% improvement in prediction accuracy. To illustrate the advantage of incorporating 3D genome and epigenetic data, we compared with two polygenic risk score methods based on summary statistics, including lassosum and LDpred2-inf. PUMICE-S resulted in an average of 47.6% and 76.2% improvement in imputation accuracy and an average gain of 20.6% and 46.4% more significant imputed models. To further identify significant gene-level associations, we applied our methodology to 37 multi-ancestry genome-wide association studies in immune-related traits and showed that PUMICE-S identified 179%, 92%, 30%, 163%, 111% more significant gene-trait pairs compare to that of PrediXcan, TIGAR (trained on DGN), UTMOST (trained on GTEx multi tissues), Lassosum, LDpred2-inf (trained on eQTLGen), respectively. Based on that, we performed a comprehensive downstream analysis, including cell type enrichment analysis and computational drug repurposing, and identified significant enrichment for cell types and bioactive small molecules capable of reversing the expression profile of clinically relevant trait-associated genes.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3568. Quantifying portable genetic effects and improving cross-ancestry genetic prediction with GWAS summary statistics

Authors:

J. Miao¹, H. GUO², G. Song¹, Z. Zhao¹, L. Hou², Q. Lu¹; ¹Univ. of Wisconsin-Madison, Madison, WI, ²Tsinghua Univ., Beijing, China

Abstract Body:

Polygenic risk scores (PRS) calculated from genome-wide association studies (GWAS) of Europeans are known to have substantially reduced predictive accuracy in non-European populations, limiting its clinical utility and raising concerns about health disparities across ancestral populations. However, existing methods to address this problem have several critical limitations (e.g., most methods require some individual-level data that are independent from input GWAS; these data rarely exist in practice), and the improvement in predictive performance has been incremental. Here, we present a novel statistical framework named X-Wing with three main innovations to improve PRS performance in ancestrally diverse populations. First, we introduce and estimate cross-population local genetic correlation which directly quantifies correlated (portable) genetic effects between multiple ancestral populations. Second, we introduce a novel Bayesian method to incorporate local genetic correlation annotation into multi-population PRS modeling, where annotation-dependent statistical shrinkage amplifies the effects of annotated variants (*i.e.*, variants with correlated effects between populations). Finally, we propose an innovative strategy to combine multiple PRS trained in various populations into an omnibus score with improved prediction accuracy using GWAS summary data alone as input. The entire X-Wing procedure only requires GWAS summary data, which is a major advance compared to existing approaches. Through extensive benchmarking using numerous GWAS datasets including UK Biobank, Biobank Japan, and Population Architecture using Genomics and Epidemiology Consortium study, we demonstrate that X-Wing pinpoints portable genetic effects and substantially improves PRS performance in non-European populations. Applied to 31 traits in East Asians, cross-population genetic correlations using X-Wing-annotated SNPs are substantially higher than the genome-wide genetic correlation estimates, while correlations in the remaining genome are consistently lower. X-Wing PRS achieves 31.0%-96.1% gain in predictive R^2 compared to state-of-the-art methods based on GWAS summary statistics. Followed up on 13 traits in admixed Americans, X-Wing also shows a substantial improvement in prediction accuracy, with the R^2 increase ranging from 18.7%-122.1%. Overall, X-Wing introduces several methodological innovations that will likely have broad and impactful applications and provides an accurate and privacy-preserving method to construct PRS in ancestrally diverse populations.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3569. Quantifying the causal impact of modifiable risk factors and biomarkers on total healthcare burden via Mendelian randomization

Authors:

J. Lee¹, S. Jukarainen², P. Dixon³, N. Davies⁴, G. Davey Smith⁴, P. Natarajan⁵, A. Ganna²; ¹Broad Inst. of MIT and Harvard, Finnish Inst. for Molecular Med., Cambridge, MA, ²Inst. for Molecular Med. Finland, Helsinki, Finland, ³Nuffield Dept. of Primary Care Hlth.Sci., Univ. of Oxford, Oxford, United Kingdom, ⁴MRC Integrative Epidemiology Unit, Univ. of Bristol, Bristol, United Kingdom, ⁵Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract Body:

Background Public health aims to reduce the burden of modifiable risk factors to improve population health and decrease healthcare burden. Previous studies have attempted to quantify the total healthcare burden (e.g. total healthcare expenditure) associated with certain modifiable risk factors such as adult body mass index (BMI), systolic blood pressure (SBP), or waist circumference (WC). However, such studies are based on observational data or single disease clinical trials. Thus, they are likely to suffer from confounding or ignore the impact of modifiable risk factors on broader healthcare burden, which is mediated by multiple disease states.

Methods Here, we linked outpatient, hospital, and medication registries to the FinnGen Study to quantify the total healthcare costs for 373,460 Finnish individuals. We performed Mendelian randomization for 15 modifiable risk factors and biomarkers to identify their causal impact on healthcare expenditure using exposures from the UK Biobank and outcomes from the FinnGen Study. Costs were standardized to costs at 2017 and converted from euros to dollars.

Results For 373,460 individuals (56% female, mean age = 61, SD age = 18), the total healthcare expenditure was \$1.08B, or a mean and median of \$2,905 and \$1,409, respectively (SD = \$11,605). We estimated that a standard deviation increase in WC, adult BMI, and SBP was increased annual total healthcare costs by 22.9% [95% CI: 18.9%-27.1%], 13.6% [10.3%-17.1%], and 13.2% [9.0%-17.7%] per person, respectively. We estimated a \$217 annual increase in healthcare costs per person for 10 cm of WC, \$191 for 5 kg/m² of adult BMI, and \$90 for 10 mmHg of SBP. This is the first comprehensive, genetically-informed quantification of modifiable risk factors.

Conclusion Taken together, we used genetic information to quantify the causal impact of modifiable risk factors and biomarkers on healthcare burden and found a substantial effect on total annual healthcare costs ranging from \$90 to \$217 per person.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3570*. Quantifying the effects of high-dimensional cross-trait assortative mating on complex trait genetic architectures

Authors:

R. Border¹, G. Athanasiadis², A. Buil³, A. Schork⁴, N. Cai⁵, A. Young⁶, T. Werge⁷, J. Flint¹, K. Kendler⁸, S. Sankararaman¹, A. Dahl⁹, A. L. Price¹⁰, N. Zaitlen¹; ¹Univ. of California Los Angeles, Los Angeles, CA, ²Univ. of Barcelona, Barcelona, Spain, ³Mental Hlth.Sct Hans Hosp., Roskilde, Denmark, ⁴Res. Inst. for Biological Psychiatry, Roskilde, Denmark, ⁵Helmholtz Zentrum München, Neuherberg, Germany, ⁶Univ. of California, Los Angeles, Los Angeles, CA, ⁷Univ. of Copenhagen, Copenhagen, Denmark, ⁸Virginia Commonwealth Univ, Richmond, VA, ⁹Univ. of Chicago, Chicago, IL, ¹⁰Harvard Sch Pub Hlth, Boston, MA

Abstract Body:

Though the potential impacts of assortative mating on genetic architecture have been understood for over a century, researchers have only recently begun to characterize its impact on commonly used methods in statistical genetics, nearly all of which assume mating is random. The results of these investigations have been striking, particularly with respect to cross-trait assortative mating (xAM): xAM inflates genetic correlation estimates, even in the absence of pleiotropy, and breaks the assumptions required for Mendelian randomization to produce valid causal inferences. Still, investigations of xAM so far have considered at most two traits at a time, a simplification undercut by empirical evidence demonstrating widespread, phenome-wide cross-mate similarity. Here, we present the first analysis of high-dimensional xAM and its implications for polygenic methods.

First, we solve the problem of simulating complex mating patterns involving an arbitrary number of traits by establishing its equivalence to a fundamental combinatorial optimization problem in operations research, enabling forward-time simulations that accurately reflect the complexity of realistic human mating patterns. Then, utilizing empirical estimates of cross-mate correlations among a broad array of traits ($n \leq 500,000$), we characterize the impact of high-dimensional xAM on multiple widely-used methods for interrogating genetic architecture. We find that the large number of phenotypes involved in xAM leads to substantial upward bias in genetic correlation estimates and polygenic prediction metrics. For example, in the complete absence of pleiotropy, five generations of bivariate xAM across ADHD and Schizophrenia would yield LDSC genetic correlation estimates 74.6% as large as published estimates; simultaneously modeling cross-mate correlation across five psychiatric phenotypes increases this to 101.1%. These findings imply that not only do xAM-induced biases likely distort our understanding of pleiotropy, but that only considering two traits at a time is insufficient for disentangling causal biology from structure in complex populations.

Finally, to address the consequences of high-dimensional xAM, we introduce random-regressor-random-slope (RRRS) regression, a novel statistical framework for decomposing genetic covariation into components reflecting pleiotropy versus xAM-induced structure by explicitly modeling the joint distribution of causal variants and their effects. We present simulations demonstrating that RRRS delivers unbiased genetic correlation estimates under xAM, and describe current progress toward implementation at biobank-scale.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3571. Rapid identification of clusters of multi-level colocalized traits within hundreds of candidate traits and in the presence of multiple causal variants.

Authors:

Z. Kuncheva¹, C. Foley^{1,2}, H. Runz³, B. Sun^{3,4}; ¹Data Sci. and Engineering, Res. and Dev., Optima Partners, Edinburgh, United Kingdom, ²Sch. of Clinical Med., Univ. of Cambridge, Cambridge, United Kingdom, ³Biogen, Cambridge, MA, ⁴BHF Cardiovascular Epidemiology Unit, Dept. of Publ. Hlth. and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom

Abstract Body:

Thousands of genomic regions affecting complex diseases have now been identified. However, elucidating the causal genes and mechanisms involved remains an outstanding biological and computational challenge. By assessing shared genetic aetiology across multiple related traits (e.g., molecular traits, metabolic pathways, complex diseases and conditions), statistical colocalization attempts to unravel causal pathways and pinpoint risk variants. Insights from current approaches are limited owing to a computational bottleneck: screening for complex patterns of colocalization across hundreds of traits simultaneously is prohibitive and costly, rendering analyses to at most around ten traits. Consequently, biological information shared between potentially hundreds of related traits is unexplored.

Here we present Coloc-clust, a novel and efficient clustering algorithm, tool and online portal. Coloc-clust ingests GWAS summary statistics to detect clusters of colocalized traits within a vast set of candidate traits simultaneously, e.g., 100 traits can be jointly analysed in seconds. Moreover, Coloc-clust does this in the presence of an arbitrary number of causal variants. We demonstrate the accuracy and impact of Coloc-clust by performing a genome-wide multi-trait, multi-causal variant colocalization analysis of cardiovascular disease (CVD) and many related traits: revealing complex and rich patterns of clustered colocalized traits across CVD related traits. For ease of interpretation and use, we summarise findings in knowledge-graphs, with helpful prioritisation statistics, for target validation, pathway identification and safety assessment.

Coloc-clust provides a user-friendly and significant contribution to the post-GWAS in-silico toolkit, helping to extend the utility of GWAS outcomes toward functional insights and to generate new biological hypotheses.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3572. Rare variation analyses of over 6,900 ALS cases in a multi-ethnic population.

Authors:

T. Pottinger, J. Motelow, G. Povysil, Z. Ren, The New York Genome Center ALS Sequencing Consortium, GTAC Investigators, D. Goldstein, M. Harms; Columbia Univ. Irving Med. Ctr., New York, NY

Abstract Body:

Hypothesis: Rare genetic variation will impact the diagnosis of amyotrophic lateral sclerosis (ALS) found through the analysis of whole genome sequencing data. Gene region analyses will further help to uncover variants that play a significant role in disease manifestation for disorders, such as ALS, determined by single gene mutations. **Background:** ALS is a neurodegenerative disease affecting approximately 5 persons per 100,000 in the general US population of adults. It is characterized by the progressive decline of the nervous system that leads to the weakening of muscles and impacts physical function. While environmental factors such as smoking and environmental toxin exposure is known to contribute to ALS, 5 to 10% of persons with this disease have genetic factors that contribute to the progression of disease. Because of the known contribution of inheritance to ALS, it is vital to examine ALS manifestations in ethnically diverse and genetically defined populations. **Approach:** For these analyses, we included 6,970 individuals who were diagnosed with ALS and 22,524 controls. We utilized regional collapsing scores that have been described previously in conjunction with gene based collapsing approaches to conduct collapsing analyses to identify rare variants that associate with ALS. **Results:** A gene-based protein truncating variants collapsing model showed a significant association with *NEK1* and *TBK1*, genes that have previously been shown to be associated with ALS. Both gene and domain based functional variants collapsing models showed significant associations with *SOD1*, *TARDBP*, and *TBK1*. However, the domain-based model removed variants found in controls in the *TARDBP* gene increasing observed associations in comparison to the gene-based model. These genes have been previously associated with ALS. **Conclusions:** In one of the largest ALS rare variation studies to date, we identified rare variants in genes that have been previously associated with ALS utilizing gene and domain-based collapsing techniques.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3573. Rare-variant aggregation analysis reveals new genes involved with IgG glycosylation.

Authors:

L. Klaric¹, A. Landini¹, P. R. H. Timmers^{1,2}, A. Frkatovic³, I. Trbojevic-Akmacic³, F. Vuckovic³, G. Tzoneva⁴, Regeneron Genetics Center, O. Polasek^{5,6}, C. Hayward¹, A. R. Shuldiner⁷, G. Lauc³, J. F. Wilson^{1,2}; ¹MRC Human Genetics Unit, Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, ²Ctr. for Global Hlth.Res., Usher Inst., Univ. of Edinburgh, Edinburgh, United Kingdom, ³Genos GlycoSci. Res. Lab., Zagreb, Croatia, ⁴Regeneron Pharmaceuticals, Tarrytown, NY, ⁵Algebra Univ. Coll., Zagreb, Croatia, ⁶Dept. of Publ. Hlth., Sch. of Med., Univ. of Split, Split, Croatia, ⁷Regeneron Genetics Ctr., tarrytown, NY

Abstract Body:

Genome-wide association studies (GWAS) have identified thousands of loci associated with human complex traits and diseases. However, most of these variants are common with small effect size and found in noncoding regions, posing a challenge for uncovering their functional impact on the phenotype. On the other hand, variants having a disruptive effect on the protein, such as predicted loss of function (pLOF) and missense variants, may help identify genes with a clear biological link to the phenotype of interest. Using multiple gene-based aggregation tests, we investigated rare (MAF < 5%) pLOF and missense variants from whole exome sequencing that associate with the N-glycome of serum transferrin (N=1907) and IgG (N=4912). Overall, we identified 18 significant gene-based associations for transferrin (p-value < 8.06x10⁻⁸) and 41 for IgG glycan traits (p-value < 1.19x10⁻⁷), of which 3 and 33, respectively, were independent from sentinel SNPs identified in GWAS using imputed genotypes. Glycome-associated rare pLOF and missense variants are located in genes already known to have a biological link to protein glycosylation (*FUT6*, *FUT8* for transferrin; *FUT8*, *MGAT3* and *ST6GAL1* for IgG) but also in genes which have not been previously reported (eg, *ARHGAP45*, *PDZRN3* and *RFXAP* for IgG). While the former show that common and rare variants act in concert to regulate protein glycosylation pathways, the latter expand our knowledge about the wide network of genes involved in protein glycosylation. In summary, variant aggregation analyses discovered both known and novel genes involved in the regulation of protein glycosylation.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3574. Rare-variant association studies: When are aggregation tests more powerful than single-variant tests?

Authors:

D. Bose¹, C. Fuchsberger^{2,1}, M. Boehnke¹; ¹Dept. of Biostatistics and Ctr. for Statistical Genetics, Univ. of Michigan, Ann Arbor, MI, ²Inst. for Biomedicine, Eurac Res., Bolzano, Italy

Abstract Body:

Although single-variant tests have been remarkably successful in identifying association in common variant GWAS, their power is substantially less for rare variants. This led to the development of aggregation tests, which have been shown in some situations to have more power for rare variants. Still, in most studies to date that have employed both types of tests, single-variant tests have yielded more associations, even for rare variants. The continued use of these methods warrants a better-informed investigation of the range of genetic models and sample sizes for which aggregation tests can be expected to be more powerful than single-variant tests.

We consider a normally distributed continuous trait in n independent study participants following an additive genetic model with c causal out of v total rare variants in an autosomal gene/region having heritability h^2 . First, we calculate analytically the power of single-variant (SV), burden, and SKAT tests to detect associations under the assumption of independent variants. With stricter assumptions of all variants having the same minor allele frequency (MAF), and causal variants having the same effect size and direction, we show that power for all three tests depends on nh^2 , c , and v . Second, to account for linkage disequilibrium, we perform simulations based on 378,215 unrelated exome-sequenced White British UK Biobank participants. Aggregation tests require us to choose which rare variants in the gene to include in the gene mask for the test with the aim to include causal variants and exclude non-causal ones. We set effect sizes of causal variants proportional to $\log(\text{MAF})$ which assigns larger effect sizes to rarer variants and Beta(1,25) weights for burden and SKAT tests to upweight rarer variants in the mask. We focus on the mask, which includes only rare protein-truncating (PTV) and missense variants in a gene, for all 1,154 genes in Chromosome 2 having at least two rare PTVs and/or missense variants. Type-I error is well controlled by all tests. When heritability=0.1% is concentrated in only a small proportion of variants in a gene (e.g., <10%) and $n=100,000$, average power for the SV test (>76%) is much higher than that of aggregation tests (<30%) for all the 1,154 genes considered. In contrast, when heritability=0.1% is split among a larger proportion of variants in a gene (e.g., >50%), average power for burden (>86%) and SKAT (>64%) tests beats that of SV test (<48%). We will extend these results to other variant masks, sample sizes, and heritabilities to more specifically answer the question of when SV tests are more powerful.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3575*. Reaching 95.0 years of age: The genetic association in the elderly Croatian population

Authors:

Ž. Celinščak, M. Setinc, A. Stojanovic Markovic, T. Skaric-Juric; Inst. for Anthropological Res., Zagreb, Croatia

Abstract Body:

Human longevity is influenced both by genetic and non-genetic factors, and genetic variability accounts for approximately 25% of variation in life expectancy. Genes that can positively influence lifespan and the ageing process are known as longevity genes. We aimed to elucidate single nucleotide polymorphisms (SNPs) that are significantly related to longevity as defined by the cut-off age of 95.0 (the threshold age for extreme longevity) in a sample of elderly persons of European origin. The study sample comprised of 314 unrelated individuals from Croatia who were above 85.0 years of age when the data were collected. As the sample of long-lived individuals was gathered in the period between 2007 and 2009, the age at death for each individual has been determined from the national mortality register 10 years after the initial sampling. This enabled us to differentiate between truly long-lived individuals that had survived beyond 95 years of age, and those that died before reaching 95 years, thus elucidating which SNPs are a key component for extreme longevity in the Croatian population. Genotype data were obtained for 42 SNPs from 28 putative longevity genes, which were selected due to strong and/or replicated association to human longevity and their role in different signalling pathways of cellular ageing and senescence. Univariate and multivariate logistic regression were performed with genotypic data coded as: 2 = longevity allele homozygotes; 1 = heterozygotes; 0 = non-longevity allele homozygotes. Out of the initial 42 SNPs, 10 SNPs that reached the inclusion criteria of having a p-value of $p < 0.2$ in univariate logistic regression entered the series of multivariate logistic regressions. The best model, explaining 9.3% of the variance for the survival to the age of 95.0, consisted of five SNPs. Three SNPs that were significantly (at $p < 0.05$ level) associated with reaching 95.0 years of age are PTPN1 rs6067484, PAPPB rs4837525, and TP53 rs1042522. Two remaining SNPs were marginally significant (APOE rs429358; $p = 0.053$) or not significant (IRF4 rs12203592) but they both contribute to the strength of the model. Although the best model explains a considerable proportion of variance for surviving up to the 95-years-of-age phenotype, the modest associations of particular SNPs warrant replication in more powered studies. (CSF grant IP-01-2018-2497: HECUBA)

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3576*. REGENIE v3: more efficient analysis of rare genetic variation with an extended set of gene-based tests.

Authors:

J. Mbatchou¹, A. Ziyatdinov², J. Marchini³, Regeneron Genetics Center; ¹Regeneron Genetics Ctr., TARRYTOWN, NY, ²RGC, Regeneron Pharmaceuticals, White Plains, NY, ³Regeneron Genetics Ctr., Tarrytown, NY

Abstract Body:

Large-scale biobanks worldwide have gathered vast amount of genetic and phenotypic data derived through electronic health care records and self-reported information. This has provided researchers unique opportunities to find new targets as well as uncover new indications for existing therapies. As analyzing such large data sets can incur a high computational burden, we had proposed REGENIE has a tool to address this limitation. REGENIE (1) uses an efficient whole genome regression framework to capture population structure, relatedness and polygenicity; (2) incurs low memory costs by reading and storing the genetic data in chunks; (3) can efficiently analyze multiple phenotypes in parallel; (4) can process both quantitative and binary traits, including highly unbalanced binary traits; (5) includes gene-based burden tests to study rare variation using functional annotation information. We have extended the tool, in REGENIE v3, to be more suitable for analyses involving rare variants, such as with whole exome sequencing. We have added a wider range of gene-based tests extending beyond the burden test, including the variance component tests SKAT and SKAT-O, which are less prone to power loss from single variants having different direction of effects, the Cauchy combination tests ACAT-V and ACAT-O, as well as a novel Non-Negative Least Squares test which combines variable selection with significance testing imposing a same-effect direction constraint. These can be applied to both quantitative as well as binary traits including those with high case-control imbalance. Furthermore, REGENIE v3 enables to perform conditional analyses, such as to determine whether a rare variant signal is driven by a nearby common variant signal. Through simulations studies as well as real data applications in UK Biobank with sample sizes up to 408K, we demonstrate the computational efficiency of REGENIE v3 compared to existing approaches including with compute time and memory usage, as well as illustrate the performance of the different gene-based tests on several quantitative and binary traits with various case-control ratios.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3577. Reliability of genotyping arrays for detection of rare variants: design & performance of a bespoke array focusing on obesity & diabetes

Authors:

S. Almansoori^{1,2,3}, H. Amin², S. I. Alsters⁴, A. M Yiorkas², D. Handley², N. Adli Nor Hashim^{5,6}, N. Hanis Ramzi⁷, T. Dovey⁸, H. Chahal^{9,10}, S. Purkayastha¹¹, R. Walters^{12,13}, F. Drenos^{14,15}, A. Blakemore^{1,2}; ¹Dept. of Metabolism, Digestion and Reproduction, Imperial Coll. London, London, United Kingdom, ²Dept. of Life Sci., Coll. of Hlth., Med. and Life Sci., Brunel Univ. London, London, United Kingdom, ³Intl. Ctr. for Forensic Sci., Gen. Dept. of Forensic Sci. and Criminology, Dubai Police, Dubai, United Arab Emirates, ⁴Dept. of Clinical Genetics, Amsterdam UMC, Vrije Univ.it Amsterdam, Amsterdam, Netherlands, ⁵Inst. of Biological Sci., Faculty of Sci., Universiti Malaya, Kuala Lumpur, Malaysia, ⁶Ctr. for Drug Res. in Systems Biology, Structural Bioinformatics and Human Digital Imaging (CRYSTAL), Universiti Malaya, Kuala Lumpur, Malaysia, ⁷Inst. For Res., Dev. & Innovation, Intl. Med. Univ., Kuala Lumpur, Malaysia, ⁸Coll. of Hlth., Med. and Life Sci., Brunel Univ. London, London, United Kingdom, ⁹Imperial Weight Ctr., Imperial Coll. Hlth.care NHS Trust, St Mary's Hosp., Praed Street, London, United Kingdom, ¹⁰Section of Investigative Med., Div. of Diabetes, Endocrinology and Metabolism, Imperial Coll. London, Hammersmith Campus, Hammersmith Hosp., 6th Floor, London, United Kingdom, ¹¹Dept. of Surgery and Cancer, Imperial Coll. London, London, United Kingdom, ¹²Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Dept. of Clinical Med., Univ. of Oxford, Oxford, United Kingdom, ¹³Med. Res. Council Population Hlth.Res. Unit, Nuffield Dept. of Clinical Med., Univ. of Oxford, Oxford, United Kingdom, ¹⁴Brunel Univ. London, London, United Kingdom, ¹⁵Inst. of Cardiovascular Sci., Faculty of Population Hlth., Univ. Coll. London, London, United Kingdom

Abstract Body:

Genotyping arrays provide an efficient, economical way to address specific traits or diseases, focusing on specific variants or genomic regions. However, rare variants present particular challenges for reliability, largely due to difficulties with algorithmic interpretation of clustering. This has led to problems of false-positive genotyping results for rare variants using the previous UK Biobank Axiom and UK BiLEVE microarrays in the UK Biobank. Here we present the design and performance of a custom genotyping array focusing on obesity and diabetes to determine the overall efficiency and performance of the Axiom myDesign genotyping array by ThermoFisher for analysing rare variants. Genotyping performance was assessed through the inclusion of duplicate samples and previously exome sequenced samples along with the standard Axiom genotyping array quality control samples. A total of 2,112 samples were genotyped and the overall quality of both samples and markers was high in standard QC. As expected, the quality of genotyping for common variants was high. We used version 5.1 of the Axiom Analysis Suite Software, which executes a Rare Het Adjusted (RHA) algorithm, to improve on the previous software version (AxiomGT1). We found that the updated version provides reliable data for rare variants: the average concordance rate between rare variants in sequenced and genotyped samples was 96% and the average concordance rate in duplicated samples was 99%. Our data suggest that an array-based approach could be efficient and cost-effective for rare variant analysis and increase confidence that the results are reliable.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3578. sCLC: a Novel Statistical Method for Association Studies of Multiple Phenotypes and Genetic Variants based on GWAS Summary Statistics

Authors:

M. Wang, X. Cao, S. Zhang, Q. Sha; Michigan Technological Univ., Houghton, MI

Abstract Body:

There is strong evidence showing that joint analysis of multiple phenotypes in genome-wide association studies (GWAS) can increase statistical power when detecting the association between genetic variants and human complex traits. We previously developed the Clustering Linear Combination (CLC) method and a computationally efficient CLC (ceCLC) method to test the association between multiple phenotypes and a genetic variant. Although CLC and ceCLC perform very well, especially for phenotypes that have natural grouping, both CLC and ceCLC require individual-level genotypes and phenotypes that are often not easily accessible due to privacy concerns and some logistical considerations. Recently, a vast majority of GWAS summary statistics obtained from single-trait tests are publicly available. In this research, we develop a novel statistical method for association studies of multiple phenotypes and genetic variants based on GWAS summary statistics. This new method is called sCLC, the CLC method based on GWAS summary statistics. We use the LD score regression to estimate the phenotypic correlation matrix among phenotypes, then use this correlation matrix to cluster multiple phenotypes into different number of clusters and test the association between phenotypes in each cluster and a genetic variant. Finally, we consider all possible numbers of clusters among the phenotypes and apply the Cauchy combination to integrate p-values for each cluster. The overall test statistic of sCLC has an approximate Cauchy distribution whose p-value can be obtained from the cumulative density function. We perform a variety of simulation studies and compare sCLC with other commonly used methods. Simulation results show that sCLC can control the Type I error rates well and has the highest power in most scenarios. Moreover, we apply the newly proposed method to the UK Biobank GWAS summary statistics from the XIII category with 70 related musculoskeletal system and connective tissue phenotypes. The results demonstrate that sCLC can detect the most number of significant SNPs, and most of these identified SNPs can be matched to genes that have been reported to be associated with those phenotypes. Furthermore, sCLC also identifies some novel signals that were missed by standard GWAS, which provide new insight into the potential genetic factors of the musculoskeletal system and connective tissue phenotypes.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3579. SDPRX: A statistical method for cross-population prediction of complex traits

Authors:

G. Zhou¹, C. Tianqi¹, H. Zhao²; ¹Yale Univ., New Haven, CT, ²Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Abstract Body:

Polygenic risk score (PRS) has demonstrated its great utility in biomedical research through identifying high risk individuals for different diseases from their genotypes. However, the broader application of PRS to the general population is hindered by the limited transferability of PRS developed in Europeans to non-European populations. To improve PRS prediction accuracy in non-European populations, we develop a statistical method called SDPRX that can effectively integrate genome wide association study summary statistics from different populations. SDPRX automatically adjusts for linkage disequilibrium differences between populations, and characterizes the joint distribution of the effect sizes of a variant in two populations to be both null, population specific or shared with correlation. The prior assumption made by SDPRX is more general than current methods such as XPASS and PRS-CSx. Unlike SDPRX, XPASS assumes that the genetic architecture is polygenic and all SNPs have non-zero effect sizes. SDPRX differs with PRS-CSx in two aspects. First, it explicitly allows SNPs to have both population specific and shared effect sizes whereas PRS-CSx assumes all SNPs are shared. Second, SDPRX directly incorporates the cross-population genetic correlation into the model for better estimation of shared effect sizes.

We first evaluated the prediction performance of SDPRX, PRS-CSx, LDpred2 and XPASS via simulations across different genetic architectures and training sample sizes. We found that jointly modeling EUR and non-EUR GWAS can improve the prediction accuracy in non-EUR populations if non-EUR GWAS alone was not well powered and SDPRX outperformed the other methods in most cases.

We next compared the performance of SDPRX with PRS-CSx, LDpred2 and XPASS in predicting 15 quantitative traits and 2 binary traits for EAS individuals in UK Biobank using external summary statistics. For each method, one can derive a linear combination of the estimated effect sizes of two populations to further optimize the performance in the target population. The average improvement of SDPRX over PRS-CSx, LDpred2 and XPASS was 14%, 49% and 43% without the linear combination, and 15%, 26% and 34% after the linear combination. For AFR individuals, the performance was compared with 6 quantitative traits and 1 binary trait due to the limited number of publicly available summary statistics. The average improvement of SDPRX over PRS-CSx, LDpred2 and XPASS was 15%, 13% and 39% without the linear combination, and 14%, 31% and 44% after the linear combination.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3580. Sensitive detection of within-species contamination from low-pass whole genome sequencing data

Authors:

A. Liu¹, M. Gibson¹, L. Flagel¹, J. Pickrell², J. Li²; ¹Gencove Inc., Long Island City, NY, ²Gencove, New York, NY

Abstract Body:

The accurate detection of DNA sample contamination is an essential aspect in the quality control process of whole genome sequencing. Even moderate levels of sample contamination can lead to a substantial increase in erroneous genotype calls, and may lead to false positive associations when conducting GWAS. Currently, existing methods primarily rely on high coverage sequence data in order to accurately infer deviations from expected allele frequencies, which might be indicative of contaminating reads from another individual.

For applications using low-coverage sequencing (such as in the context of ancient DNA studies and cell-free DNA), there are few existing methods in the literature, most of which are appropriate only for very particular applications.

Here, we introduce a method that leverages known haplotype frequencies from a reference panel such as the 1000 Genomes in order to make accurate estimates of contamination in samples which are sequenced to extremely low (<1x) coverages. Our method uses a maximum likelihood framework which compares the expected and observed linkage disequilibrium between pairs of markers that overlap with distinct sequence reads. Under this framework, an observed deviation from expected haplotype frequencies indicates the presence of contaminating reads from another individual. To accurately represent the individual's haplotype frequencies, we adjust population haplotype frequencies to reflect the inferred ancestry of the target sample.

We show that our low-coverage contamination method estimates contamination more accurately at low coverages (<1x) than existing methods tuned for high-coverage samples, particularly for low levels of contamination (1%- 5%). Furthermore, we show that our method can resolve biased contamination estimates that may arise from admixed individuals or mis-specified haplotype frequencies from the reference panel. Our work emphasizes the importance of quality control metrics tailored to low-pass sequencing as applications in this area become increasingly popular.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3582. Sex-stratified vs. sex-combined analysis in the presence of genetic effect heterogeneity.

Authors:

B. Lin, L. Sun; Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

The effect of a genetic variant on a complex trait may differ between male and female, e.g. genetic effects may be sex-specific for testosterone levels. In the presence of genetic effect heterogeneity between female and male, sex-stratified analysis is often used, which provides easy-to-interpret sex-specific effect size estimates. However, from power of association testing perspective, sex-stratified analysis may not be the best approach. As sex-specific genetic effect implies SNP-sex interaction effect, jointly testing SNP main and SNP-sex interaction effects may be more powerful than sex-stratified analysis or the standard main-effect testing approach. When individual data are not available, it is then of interest to study if the interaction analysis can be derived from sex-stratified summary statistics. We considered several different sex-combined methods and evaluated them through extensive simulation studies. We observed that a) the joint SNP main and SNP-sex interaction analysis is most robust to a wide range of genetic models, and b) this joint interaction testing result can be obtained by quadratically combining sex-stratified summary statistics (i.e. squared sex-stratified summary statistics). We then provide theoretical justification for the equivalence between these joint interaction test and quadratically combined omnibus test. Finally, we provide additional supporting evidence by applying the methods to GWAS of testosterone levels in the UK Biobank data.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3583. Shared genetic etiology and causality between COVID-19 and venous thromboembolism: evidence from genome-wide cross trait analysis and bi-directional Mendelian randomization study

Authors:

Z. Liu^{1,2}; ¹Univ. of Hong Kong, Hong Kong, Hong Kong, ²Columbia Univ., New York, NY

Abstract Body:

Venous thromboembolism (VTE) occurs in up to one third patients with COVID-19. VTE and COVID-19 may share a common genetic architecture, which has not been clarified yet. To fill this gap, we leveraged summary-level genetic data from the latest COVID-19 host genetics consortium and UK Biobank and examined their shared genetic etiology and causality. The cross-trait analysis identified 8, 11, and 7 shared loci between VTE and severe COVID-19, COVID-19 hospitalization, SARS-CoV-2 infection respectively, in 13 genes involved in coagulation and immune function and enriched in the lung. Co-localization analysis identified eight shared loci in ABO, ADAMTS13 and FUT2 genes. Bi-direction Mendelian randomization suggested that VTE was associated with higher risks of all COVID-19 related traits, and SARS-CoV-2 infection was associated with higher risk of VTE. Our study provided timely evidence and novel insights into the genetic etiology between COVID-19 and VTE.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3584. Shared genetic etiology between Alzheimer's disease and stroke: A large scale genome-wide cross-trait analysis

Authors:

Z. Wang, X. Han, J. Li, IGAP Consortium, MEGASTROKE Consortium, L. Liang; Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract Body:

Alzheimer's disease (AD) and stroke are the leading causes of death worldwide and common comorbidities in the elderly population. There are many shared risk factors between the diseases, which could possibly be driven by shared genetic factors. With GWAS summary statistics obtained from IGAP and MEGASTROKE consortiums, we performed a cross-trait analysis between AD and stroke to investigate the genetic correlations, common genetic factors, shared GWAS risk loci, and causal effects between the two diseases. Results showed significant genetic correlations between AD and all strokes (AS), as well as cardioembolic stroke (CES). Based on the Mendelian Randomization (MR) methods, we evaluated the diseases' causal effects and concluded that the observed genetic correlations are not causal between the diseases. Meta-analysis revealed shared genes and risk loci, including *CELF1* and *MTCH2*. We then performed causal inference between 338 CSF metabolites and the diseases to investigate the biological mechanisms behind the genetic correlations we observed. Based on MR, we identified the shared metabolites and pathways including cholesterol, ascorbate, 2-hydroxybutyrate, which may contribute to blood-brain barrier dysfunction, oxidative stress, and other risk factors. Lastly, to investigate the similarity between AD and SVS metabolite profiles, we conducted genome-wide association analyses for 6 MRI markers using UK Biobank data, which is currently the largest in scale. With the GWAS summary statistics, we examined the genetic correlations and causal effects between MRI markers and complex diseases. Results show a strong causal effect from fractional anisotropy (FA) to CES, and causal effects from AD to FA, mean diffusivity (MD), and total volume of white matter hyperintensities (TVWMH). The results suggest the shared pathology between AD and stroke is mainly related to the white matter. In conclusion, the observed genetic correlations between AD and stroke are likely due to the non-causal effects, which are explainable using shared risk loci, shared metabolites & pathways, and MRI markers.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3585. Shared Genetics Drive Mate Selection.

Authors:

K. Burghardt¹, E. Arpawong², J-L. Ambite¹; ¹USC Information Sci. Inst., Marina del Rey, CA, ²Univ. of Southern California, Los Angeles, CA

Abstract Body:

People often use the expression "they were born to be together" when referring to their relationships, but how do the factors from birth---namely genetics---play a role in mate selection? Prior work in this area has focused on using polygenic scores to assess genetic correlations between one's polygenic scores and spouse phenotypes, focused on specific areas of similarity in the genome, or on European ancestry samples only. We expand this work by studying a racially/ethnically diverse subsample of older adults to analyze data across the genome from 2,424 participants who form 1,214 marriage pairs, which is our proxy of mate selection. In preliminary analysis, we find that principal components, extracted using principal component analysis (PCA) across all participants who shared their genetic sequences, are the dimensions that explain the most data variance, and reflect genetic ancestry. We then develop methods to determine genetic associations with mate selection throughout the genome. First, we use a model to find the genetic segments that associate with marriage, by comparing pairwise-shared segments between married participants with segments shared by random pairs of subjects. Then, we find the phenotypic covariates that mediate these associations. We find that a small set of genetic slices are significantly associated with marriage, but these slices are mediated by preferences for partners with similar ethnicity, religion, and birthplace, and (to a lesser extent) education. After controlling for these covariates, we find that some genetic slices are significantly associated with marriages. Next, we show an overall quantitative approach of scoring slices together. We calculate pairwise genetic risk scores, which we call Slice Polygenic Risk Scores (SliPRS) and find significant correlations between high SliPRS and the probability of marriage between participant pairs. These results offer new insight into the genetics and mediating factors driving mate selection.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3586. Short tandem repeat expansions are present in up to 25% of sporadic amyotrophic lateral sclerosis and frontotemporal dementia patients

Authors:

L. Henden¹, L. G. Fearnley^{2,3}, N. Grima¹, E. P. McCann¹, C. Dobson-Stone^{4,5}, L. Fitzpatrick^{4,5}, K. Friend⁶, L. Hobson⁶, S. Chan Moi Fat¹, D. B. Rowe^{1,7}, S. D'Silva^{1,7}, J. B. Kwok^{4,5}, G. Halliday^{4,5}, M. C. Kiernan^{4,8}, S. Mazumder⁴, H. C. Timmins⁴, M. Zoing⁴, R. Pamphlett^{9,5,10}, L. Adams¹, M. Bahlo^{2,3}, I. P. Blair¹, K. L. Williams¹; ¹Macquarie Univ. Ctr. for Motor Neuron Disease Res., Sydney, Australia, ²Population Hlth. and Immunity Div., The Walter and Eliza Hall Inst. of Med. Res., Melbourne, Australia, ³Dept. of Med. Biology, The Univ. of Melbourne, Melbourne, Australia, ⁴Brain and Mind Ctr., The Univ. of Sydney, Sydney, Australia, ⁵Sch. of Med. Sci., Faculty of Med. and Hlth., Univ. of New South Wales, Sydney, Australia, ⁶SA Pathology, Women's and Children's Hosp., Adelaide, Australia, ⁷Dept. of Clinical Med., Faculty of Med. and Hlth. Sci., Macquarie Univ., Sydney, Australia, ⁸Dept. of Neurology, Royal Prince Alfred Hosp., Sydney, Australia, ⁹Brain and Mind Ctr., The Univ. of Sydney, Sydney, Sydney, Australia, ¹⁰Dept. of Neuropathology, Royal Prince Alfred Hosp., Sydney, Australia

Abstract Body:

Introduction: Short tandem repeats (STRs) are tracts of repetitive DNA present throughout the human genome. Pathogenic expansions of STRs cause more than 20 neurodegenerative diseases. Historically, to assess STRs in individuals, targeted laboratory-based screening of single STRs was performed. However, the development of bioinformatic tools that detect STRs in whole-genome sequencing (WGS) data enables simultaneous assessment of multiple STRs in large disease cohorts.

Objective: To systematically assess known STR expansions in sporadic amyotrophic lateral sclerosis (sALS) and sporadic frontotemporal dementia (sFTD) patients using WGS data.

Methods: 21 STR expansions known to cause motor and neurodegenerative diseases were screened in WGS data from 608 Australian sALS and 68 Australian sFTD patients, and 4,703 European controls. PCR genotyping was performed to validate expansions.

Results: We identified and validated 162 pathogenic and/or intermediate repeat expansions in 152/676 (22%) sALS and sFTD patients across 9 genes: *C9orf72* (ALS/FTD), *ATXN1* (spinal cerebellar ataxia type 1, SCA1), *ATXN2* (SCA2), *ATXN8* (SCA8), *TBP* (SCA17), *HTT* (Huntington's disease, HD), *DMPK* (myotonic dystrophy 1, DM1), *CNBP* (DM2) and *FMRI* (Fragile-X site A, FRAXA).

Pathogenic expansions were identified in *C9orf72* (n=41), *ATXN8* (n=6), *CNBP* (n=3), *ATXN1* (n=1), *DMPK* (n=1) and *FMRI* (n=1).

Intermediate expansions were identified in *ATXN1* (n=89), *ATXN2* (n=12), *ATXN8* (n=2), *HTT* (n=2), *TBP* (n=1) and *FMRI* (n=3). Expansions in *C9orf72*, *ATXN2* and *HTT* were 37, 3 and 2.5 times more prevalent in patients than controls, respectively.

Conclusion: STR expansions that are causal for neurodegenerative diseases are present in up to 25% of sporadic ALS and FTD patients. This further supports the concept of a common genetic spectrum of neurodegenerative diseases.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3587. Simple phenotype transformations help elucidate genetic architecture in the tails of biomarker distributions.

Authors:

N. Baya^{1,2,3}, D. Palmer¹, S. Myers³, C. Lindgren¹; ¹Big Data Inst., Univ. of Oxford, Oxford, United Kingdom, ²Wellcome Ctr. for Human Genetics, Oxford, United Kingdom, ³Dept. of Statistics, Univ. of Oxford, Oxford, United Kingdom

Abstract Body:

We can better understand the genetic underpinning of complex disease by studying the genetic architecture of related biomarkers. A genetic predisposition to being in the tail of a biomarker distribution may be relevant to related diseases. We sought to discover variants associated with being in the tails of a biomarker distribution (tail status) that are overlooked in standard biomarker genome-wide association studies (GWAS).

Tail status for 33 biomarkers was defined in 185,506 European ancestry samples from the 200k UK Biobank whole exome sequencing data release. Samples residing in either tail of a biomarker distribution were considered cases in a GWAS.

We were particularly interested to determine whether any variants are more significantly associated with tail status than the biomarker viewed as a continuous trait. We hypothesized that there would be a higher proportion of rare variant associations in the tail status GWAS than the continuous trait GWAS because we expect samples in the tails to have a greater accumulation of rare, high effect variants.

We found 458 exome-wide significant ($p < 3.45e-7$, Bonferroni-corrected using number of variants tested, $n=144,561$) variant associations across 19 traits for the 'either-tail' GWAS. Among these associations, the rare variant ($MAF < 0.01$) proportion ($60/458=13\%$) was larger than the proportion of rare variants with significant associations in the continuous trait GWAS ($480/11807=4\%$). Within the either-tail GWAS significant associations, we found 66 variants that were more significant in the either-tail GWAS than the continuous trait GWAS. Of these, 40 were significant only in the either-tail GWAS. Evidence of non-monotonic trends in average allele frequency across the continuous biomarker distribution was found in 10 of 66 variants, suggesting a complex non-linear architecture underlying the effects.

To assess whether the 458 significant either-tail associations are driven by lower-tail or upper-tail associations, we performed two additional lower-/upper-tail GWAS per biomarker, either defining cases as individuals residing in the lower tail or residing in the upper tail. Among the variants significantly associated with either-tail status, 52% (236/458) had upper-tail GWAS p-values that were more significant than the lower-tail GWAS p-values.

Our results show that by using simple phenotype binarization we can uncover variant effects that are significantly associated with tail status that would not be found in a standard GWAS on the continuous trait. We expect these associations to provide clues to aid in our understanding of the genetic architecture of biomarkers and related diseases.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3588. Single cell RNA-seq data analysis of the Huntington's disease cerebellum using the interactive webserver, ICARUS.

Authors:

A. Jiang¹, E. Mears¹, R. Handley¹, V. Hawkins¹, C. McLaughlan², S. Bawden², S. Rudiger², J. Kelly², P. Verma², R. Faull¹, H. Waldvogel¹, L. You³, K. Lehnert¹, R. Snell¹; ¹The Univ. of Auckland, Auckland, New Zealand, ²South Australian Res. and Dev. Inst., Adelaide, Australia, ³Fudan Univ., Shanghai, China

Abstract Body:

Huntington's disease (HD) is a debilitating neurodegenerative genetic disorder caused by an expanded polyglutamine (CAG) trinucleotide repeat in the Huntingtin (HTT) gene resulting in a triad of behavioural, cognitive, and motor defects. Current knowledge of disease pathogenesis remains unclear and no disease modifying interventions have been discovered. We present the results of our cerebellar single cell RNA-seq data performed on 5 human HD cases and 5 controls. We will also present the analysis of striatal samples taken from 6 year-old HD sheep (*OVT73*) and 6 controls. The datasets reveal further supporting evidence for neuroinflammation, excitotoxicity and mitochondrial dysfunction in the HD brain.

The dataset was analysed with ICARUS, a new web server tool designed and implemented to enable users without experience in R to undertake single cell RNA-seq analysis. The focal point of ICARUS is its intuitive tutorial-style user interface, designed to guide logical navigation through the multitude of pre-processing, analysis and visualization steps (Available here, <https://launch.icarus-scrnaseq.cloud.edu.au/>). In summary, we will present our latest findings incorporating both the sheep model and the human results.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3589. Single-cell transcriptomic landscape in Alzheimer's disease

Authors:

V. Swarup, E. Miyoshi, S. Morabito, E. Head; Univ. of California Irvine, Irvine, CA

Abstract Body:

The gene-regulatory landscape of the brain is highly dynamic in health and disease, coordinating a menagerie of biological processes across distinct cell-types. Understanding these regulatory programs requires a holistic experimental and analytical approach. Here, we present a single-cell study of 380,000 nuclei in late-stage Alzheimer's Disease (AD) using parse biosciences whole transcriptome kit, profiling gene-expression in thousands of genes and uncovering vast neuronal and glial heterogeneity in late-stage AD. We introduce a co-expression network analysis strategy for single-cell data (hd-WGCNA) to perform a systems-level meta-analysis of AD transcriptomics to uncover underlying regulatory architecture. hd-WGCNA is based on a meta-analytical approach to jointly form co-expression networks in metacells constructed from the integrated snRNA-seq dataset as well as bulk-tissue RNA-seq data of the human cortex, where each edge in a co-expressed module is supported by both bulk RNA-seq data and snRNA-seq data. Finally, this work, provides an unbiased single-cell atlas of transcriptomic regulation of AD in distinct brain regions.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3590. Smoking-informed methylation QTLs in human nucleus accumbens

Authors:

M. Carnes¹, B. Quach², L. Zhou¹, R. Tao³, A. Deep-Soboslay³, L. Bierut⁴, B. Maher⁵, A. Jaffe³, T. Hyde³, J. Kleinman³, E. O. Johnson⁶, D. Hancock⁷; ¹RTI Intl., Durham, NC, ²RTI Intl., Research Triangle Park, NC, ³Lieber Inst. for Brain Dev., Baltimore, MD, ⁴Washington Univ Sch Med, St Louis, MO, ⁵Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ⁶RTI Intl., Duluth, MN, ⁷RTI Interntional, Durham, NC

Abstract Body:

Smoking is a leading cause of preventable morbidity and mortality. Smoking is heritable, and hundreds of genome-wide significant variants are associated with smoking behaviors (initiation, age of initiation, cigarettes per day, and cessation), as identified by the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN). Most are noncoding variants with unknown neurobiological effects. To evaluate their regulatory potential, we used genome-wide genotypes and DNA methylation (DNAm) data in postmortem human nucleus accumbens (NAc) from the LIBD Human Brain Repository to identify cis-methylation quantitative trait loci (mQTL) and investigate epistatic interactions by cigarette smoking. Active smokers (European ancestry N= 26; African ancestry N= 26; total N=52) and nonsmokers (European ancestry N= 75; African ancestry N= 93; total N=168) were defined based on cotinine and next-of-kin reporting. We used the joint 2df method to simultaneously test variant and smoking-by-variant interaction effects on DNAm, adjusting for biological and technical covariates. We accounted for multiple testing using a two-stage approach based on eigenMT and Bonferroni corrections. We found >2 million significant mQTLs (padj<0.05) representing 41,695 unique CpGs after filtering (MAF>0.05 and missingness <0.10 in both ancestries). Most of the unique CpGs (67%) lie within a gene body or in a known gene regulatory region. Results were largely driven by main effects, as 95% of significant CpGs were also identified in baseline mQTL analyses that did not account for an interaction. We assessed mQTLs among 371 overlapping GSCAN-identified variants. Of these, 229 variants (62%) were significant mQTLs, all driven by main effects of the variant on DNAm with no evidence of interaction with smoking. Of the 229, the top mQTLs overlapped with GSCAN-identified variants that were annotated to *HLA-G*, *BRWD1*, *ZNF207*, and *SLC25A20* (all p<1e-50). The most significant of which, rs3115418-cg04567952 (punadj=1.9e-101), was identified by GTEx as a NAc eQTL for the *TYW5* gene. We present the first genome-wide mQTL map in the human NAc and report enriched overlap with genetic variants of smoking behaviors, suggesting that regulation of DNAm levels in the brain may help explain the neurobiology underlying smoking GWAS loci.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3591. Specific HLA risk alleles associated with the highly variable region of T cell receptors (TCRs) in Leukemia, Melanoma and COVID-19 patients.

Authors:

J. Han; Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Abstract Body:

[Background] Human leukocyte antigen (HLA) genes are highly polymorphic and strongly affect diverse disease risks including immune-related diseases. The peripheral hypothesis is that HLA risk alleles are encrypted to proteins that may increase presenting critical autoantigens and lead to autoimmunity. In this hypothesis, the event arises in the peripheral tissues and lymph nodes. There is another non-mutually exclusive hypothesis, the central hypothesis. In the central hypothesis, HLA risk alleles influence primary lymphoid organ selection and ultimately affect the increasing number of T cell receptors (TCRs) having autoreactivity. In this hypothesis, the key event arises in the primary lymphoid organ, thymus and bone marrow selection. Nowadays, human autoimmunity researchers assert some evidence supporting the central hypothesis rather than the peripheral hypothesis. [Study design] We wanted to study the interaction between TCRs and modified patients' cells from the perspective of the central hypothesis in non-autoimmune diseases. We investigated several non-autoimmune disease categories that may be related to the HLA-TCR interaction, which are cancer diseases by the mutagen (leukemia and melanoma) and infectious diseases by the virus (COVID-19). Here we investigated the effect of HLA alleles on TCRs' highly diverse regions (e.g. complementarity determining region 3, CDR3) which have an important role in antigen recognition. We analyzed which HLA risk alleles affect TCRs' region characteristics in the patients and compared the association pattern to the healthy control, from the perspective of the central hypothesis. [Results] We found a number of associations between HLA alleles and TCR region in non-autoimmune diseases. We observed that some HLA alleles and amino acid positions affect TCR in multiple diseases. In addition, the TCR's highly diverse region involved by the HLA risk alleles were different in patients from in healthy people. In summary, these investigations show that HLA risk alleles may lead to the bias of TCR's highly diverse regions on an individual level, which can differ by an individual's disease status.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3592. ssCTPR: summary statistic based cross-trait penalized regression

Authors:

M. Weidmann, J. Rossen, L. Liang; Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract Body:

The arrival of large-scale genome-wide association studies (GWAS) in recent years has allowed the development of polygenic risk scores (PRS) for genetic risk prediction. However, current sample sizes are often still inadequate to accurately estimate the effects of variants with minor genetic effects. GWAS summary statistics provide access to effect estimates for individual variants on many traits without needing individual-level data. By leveraging shared causal variants from genetically-related traits, it is possible to improve the effect estimates and overall accuracy of PRS. Here, we propose summary-statistic-based cross-trait penalized regression (ssCTPR). ssCTPR uses GWAS summary statistics for the primary traits as well as from multiple genetically-related traits. The summary statistics for the secondary traits can be generated from either the same or a separate set of subjects. Our cross-trait penalty function utilizes information from secondary traits that could be useful to improve the causal effect estimation for the primary trait of interest while discarding non-useful information. Our method can take advantage of precomputed linkage disequilibrium (LD) matrices based on a large reference panel to better estimate independent causal genetic effects. We also developed an efficient search algorithm to ensure convergence and optimization of the tuning parameters. We used genotype data from 330,000 British individuals from UK Biobank to simulate a range of genetic architectures for continuous traits. Using a preliminary implementation of the method, ssCTPR attained a 6% increase in out-of-sample r^2 compared to single trait PRS based on LassoSum. We further extend the method by implementing a gradient-descent algorithm to rapidly search the 3-dimensional space for tuning parameters. We will also discuss criteria to ensure gradient descent algorithm convergence when candidate SNPs are in LD. Our method shows that adding GWAS summary statistics from related traits can boost the prediction accuracy of PRS compared to PRS generated using summary statistics of only the primary trait. We compute and make publicly available the PRS models using UK Biobank diseases and traits.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3593*. Subcontinental Admixture in Individuals with European Ancestry and Implications for Genetic Epidemiology Studies

Authors:

M. Gouveia, A. Adeyemo, C. N. Rotimi, D. Shriner; NIH, Bethesda, MD

Abstract Body:

Admixed populations are commonly thought to have resulted from recent interbreeding between continentally separated populations. European-ancestry populations are recognized as stratified but not as admixed, so genetic epidemiology studies including Europeans or European Americans do not fully control for potential confounding effects of local ancestry. We analyzed a large genome-wide data collection of over 18,000 European-ancestry individuals, including 1,216 individuals across 79 European populations and 17,684 European Americans from five genetic epidemiology cohorts in the US (ARIC, CARDIA, FHS, GENOA, and MESA). We found that the 1000 Genomes and Human Genome Diversity Projects provided incomplete coverage of European ancestries, so we generated a new reference panel to capture the breadth of European ancestral diversity. We inferred a three-way admixture profile in Europeans, with genetic variation significantly correlated ($p < 1.2 \times 10^{-13}$) with North-to-South ($\rho = 0.72$) and West-to-East ($\rho = 0.84$) geographic origin. We also identified population structure in European Americans, with this structure due to admixture rather than discrete subpopulations. We replicated the classical false-positive association between height and lactase persistence (rs4988235) when models were not fully adjusted for ancestral population structure. In a GWAS of height, we observed systematically better fits and smaller genomic inflation factors when models were adjusted for principal components (PCs) derived from projection of European Americans onto our reference panel, rather than for PCs derived from study-specific unsupervised analysis. Overall, we provide evidence of subcontinental admixture in Europeans and European Americans, which if not properly accounted for can produce false positives in genetic epidemiology studies.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3594*. Supervised modeling on CRISPR gene regulatory screening and cell-type-specific gene regulatory landscapes to prioritize GWAS Targeted Genes

Authors:

M-J. Tsai^{1,2}, S. Reppe³, T. Sato⁴, M. Wein⁴, K. Gautvik³, Y-H. Hsu^{1,5,2}; ¹Hinda and Arthur Marcus Inst. for Aging research, Boston, MA, ²Dept. of Med., Beth Israel Deaconess Med. Ctr. and Harvard Med. Sch., Boston, MA, ³Oslo Univ. Hosp., Oslo, Norway, ⁴Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, ⁵Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract Body:

Background: More than 95% of disease-linked genetic variants reside in the dark genome, and most of these non-coding variants have no function due to historical technical barriers. Significant efforts have been made to characterize human non-coding genomic regions, and advanced approaches have been developed to identify causal variants and prioritize targeted genes in non-coding GWAS loci. But it remains a significant challenge to connect causal variants to their target genes across cell types. **Methods:** To overcome such challenges, we apply supervised machine learning algorithms to develop a prediction method, EPFinder, to predict enhancer-gene mapping and prioritize the target genes in GWAS loci. We generated human cell-type-specific high-resolution chromatin-interaction (Hi-C seq), open chromatin regions (ATAC-seq), histone marks (ChIP-seq), and transcriptome (RNA-seq). These data were then used as input features with ensemble machine learning models to learn the patterns of 4,072 enhancer-gene pairs in human K562 Leukemia Cells obtained from the CRISPR-FlowFISH dataset. We apply this model to predict known disease genes under GWAS loci for type 2 diabetes (T2D), height, and osteoporosis GWAS to benchmark and validate model performance. **Results:** EPFinder has an area under the precision-recall curve (AUPRC) of 90.08% (AUPRC range from 87.6% to 90.3%) for different cell types, outperforming existing methods for predicting CRISPR datasets. To validate the performance of targeted genes prediction in GWAS loci, we selected the well-known T2D, height, and bone genes. For model prediction, we used cell-type-specific Hi-C, ATAC-seq, RNA-seq, and ChIP-seq from disease-relevant primary cell types. We estimated prediction ability by odds ratio (OR) of the positive prediction rate among those well-known targeted genes. Our model improved targeted prediction ability from OR = 21.1 ($p = 7.8 \times 10^{-13}$) to OR = 41.1 ($p = 3.3 \times 10^{-49}$) compared to identifying target genes using Hi-C and ATAC-seq along. **Conclusion:** With human primary cell-type-specific 3D genome architecture and gene regulatory landscapes as input, our model allows us to prioritize targeted genes in GWAS loci with traits relevant to the cell types of the input data. Machine learning to predict cell-type-specific enhancer-gene mapping enables us to accurately identify affected cell types, potential causal variants, and causal genes, especially in non-coding GWAS loci.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3595. Supervised multiset sparse partial least squares discriminant analysis in multi-omics data integration.

Authors:

K. Su¹, A. M. Alam¹, C. qiu¹, J. Greenbaum¹, Q. Tian¹, Z. Luo¹, L. Wu¹, L. Zhao¹, H. Shen^{2,1}, H-W. Deng^{2,1}; ¹Tulane Ctr. for BioMed. Informatics and Genomics, New Orleans, LA, ²Tulane Univ., New Orleans, LA

Abstract Body:

Multi-omics (MO) studies exploring the interrelationships of various biological factors have provided newfound opportunities to discover biomarkers and characterize biochemical consequences of genetic variation. However, the heterogeneous data types and high dimensionality of the data introduce extraordinary challenges for developing analytical methodologies to integrate MO datasets. Current methods integrated MO datasets without considering traits and therefore provide irrelevant results for disease outcomes. We propose a multiset sparse Partial Least Squares Discriminant Analysis (msPLSDA) method to simultaneously extract phenotype-relevant features from thousands of heterogeneous biological elements and predict phenotype status in a unified supervised MO framework.

The msPLSDA consists of three essential components. The first is a PLS method, which is well suited for handling multicollinearity issues; the second is an elastic net penalization technique incorporated into the PLS modeling that enables feature selection; and the third is a multi-block technique, which can incorporate prior biological knowledge, such as functional gene modules or pathway information. Numerical simulation experiments were conducted to demonstrate the performance in feature selection and discrimination for the proposed method. As a real data application, msPLSDA was applied to the transcriptome, methylome, and metabolome from Louisiana Osteoporosis Study for bone-related biomarker identification.

The implementation of the proposed method for feature selection indicated that high selection accuracy is achieved by increasing sample size (suggested at least 80 samples for three categorical outcomes). The high differences of features among outcome groups and lower covariances among features also contribute to the accuracy of feature selections. In the LOS dataset, the trained msPLSDA model reached a high accuracy discrimination rate (95% confidence interval: 0.96 ~ 1) and identified more significant biomarkers in terms of bone-associated pathways, such as osteoclast differentiation (KEGG, p-value: 0.0037 and odd ratio: 6.77).

Overall, msPLSDA revealed phenotype-related features from each omics type. The generated results will assist in providing interpretable outputs, which can then be used for further downstream bioinformatic analyses, such as pathway analysis to identify molecular functionalities, network analysis to estimate molecular interactions and causal inference. This method will help pave the way toward a better understanding of disease pathogenesis in translational medicine.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3596. Systematic Integration of Multi-omics Data for the Study of Coronary Artery Disease and Subclinical Atherosclerosis

Authors:

C. Yang¹, F. Aguet², K. Ardlie², R. Gerszten³, H. Wheeler⁴, K. Taylor⁵, P. Durda⁶, C. Johnson⁷, X. Guo⁵, Y. Liu⁸, R. Tracy⁹, T. Blackwell¹⁰, G. Papanicolaou¹¹, J. Rotter⁵, S. Rich¹, A. Manichaikul¹; ¹Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, ²Broad Inst. of Harvard and MIT, Cambridge, MA, ³Div. of Cardiovascular Med., Beth Israel Deaconess Med. Ctr., Boston, MA, ⁴Dept. of Biology, Loyola Univ. Chicago, Chicago, IL, ⁵Dept. of Pediatrics, The Inst. for Translational Genomics and Population Sci., The Lundquist Inst. for BioMed. Innovation at Harbor-UCLA Med. Ctr., Torrance, CA, ⁶Dept. of Pathology and Lab. Med., The Robert Larner, M.D. Coll. of Med. at the Univ. of Vermont, Burlington, VT, ⁷Univ. of Washington, Seattle, WA, ⁸Duke Univ., Chapel Hill, NC, ⁹Dept.s of Pathology & Lab. Med., and Biochemistry, Larner Coll. of Med., Univ. of Vermont, Burlington, VT, ¹⁰Univ. of Michigan, Ann Arbor, MI, ¹¹NIH, Bethesda, MD

Abstract Body:

Coronary artery disease (CAD) is a leading cause of death and disability worldwide and represents a common complex disease with genetic and environmental determinants. GWAS of CAD from international consortia, including the CARDIoGRAMplusC4D Consortia, and analysis of UK Biobank data have identified over 150 independent variants associated with the risk of CAD. In this study, we leverage multiple sources of molecular 'omics data from the NHLBI Trans-Omics for Precision Medicine (TOPMed) program to examine prior findings with (a) overlap of GWAS with eQTL/pQTL, and (b) network structure. Our study leveraged transcriptomic data obtained from peripheral blood mononuclear cells (PBMCs) and plasma proteomics from participants in the Multi-Ethnic Study of Atherosclerosis (MESA). We performed Bayesian colocalization using R/coloc to identify the molecular targets implicated by GWAS. Support for hypothesis 4 (both the QTL and GWAS are associated with the region and share a single causal variant) used a posterior probability threshold of 0.8 for Bayesian colocalization. To investigate the relationship of the colocalized genes and proteins with subclinical atherosclerosis, we conducted Bayesian colocalization using the results of GWAS for coronary artery calcium (CAC) and carotid artery intima-thickness (cIMT), focusing on those genes identified in our initial analyses of CAD. We performed weighted gene co-expression network analyses (WGCNA) to identify the modules of highly correlated genes and calculated the module membership (MM threshold of 0.8) to identify the hub genes for each module. We examined the association of the hub genes with CAC in MESA using linear regression with adjustment for age, sex, and study site to identify the CAC-related hub genes for each module. Based on our Bayesian colocalization, we identified 224 putative genes and 78 putative proteins by integrating eQTL and pQTL from TOPMed MESA with summary statistics from the prior GWAS of CAD (van der Harst and Verweij, 2018). Among these, we observed additional colocalization with subclinical atherosclerosis traits for *ILRUN* (cIMT), *GGCX* (CAC), and *RNF114* (CAC). Using WGCNA on the TOPMed RNA-seq data from MESA PBMCs, we constructed 30 modules of highly correlated genes and further identified the CAC-related hub genes for each module. We found *AHCYL1* to be a CAC-related hub gene from the "dark green" module, also located within the region of the lead variant on chromosome 1 from the GWAS of CAD. Our study identified multiple candidate genes at GWAS loci that are supported by molecular QTL and further showed the value of network analysis to highlight the core role of some genes implicated by GWAS.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3597. The effect of CA and D vitamins on the course of COVID 19 from the plasma osmolality perspective in a Turkish Cohort.

Authors:

A. Ulgen¹, H. Sivgin², S. Cetin³, W. Li⁴; ¹Dept of Biostatistics, Faculty of Med., Girne American Univ., Karmi, Cyprus, ²Tokat State Hosp., Dept. of Internal Med., Tokat, Turkey, ³Dept. of Biostatistics, Faculty of Med., Tokat Gaziosmanpasa Univ., Tokat, Turkey, ⁴The Robert S. Boas Ctr. for Genomics and Human Genetics, The Feinstein Inst. for Med. Res., Manhasset, NY

Abstract Body:

Osmolality, concentration of solute particles, can be estimated from the measurement of three other blood test variables (sodium, glucose, and urea), and was rarely used for prognosis for COVID-19. As a result of the analysis of the data obtained from COVID-19 patients, we found osmolality to be an excellent prognostic biomarker for both mortality and hospitalization. On the other hand, both hypocalcemia and vitamin-D deficiency are also significantly associated with a higher mortality rate in our data, while only calcium and not vitamin D is associated with a higher hospitalization rate. Different types of tests, including logistic regression, t-test, Wilcoxon test, lead to the same conclusion. After conditioning on osmolality in a multiple logistic regression, calcium level remains to be significantly associated with both mortality and hospitalization (p-val <0.001). However, vitamin D loses its association with mortality when conditioning on osmolality (significant only at 0.05 level).

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3598. The effect of mental diseases on chronic kidney disease: a bidirectional Mendelian randomization study

Authors:

S. Yu; Sichuan Univ., Chengdu, China

Abstract Body:

Background:

It has been released that the association between chronic kidney disease (CKD) and mental diseases such as bipolar disease (BIP), major depressive disorder (MDD), anxiety disorder (AD), attention deficit/hyperactivity disorder (ADHD), and schizophrenia (SCZ). the prevalence of depression in CKD patients varies from 22.8 to 39.3%. Therefore, psychiatric patients have 1.5-3 times more hospitalization compared to patients having only CKD. However, there was insufficient evidence on whether these psychiatric traits causally lead to CKD and vice versa, owing to possible residual confounding and reverse causation bias in observational research. Methods:we performed various MR studies including inverse-variance weighted method and MR-Egger, **latent** causal variable (LCV) model, Causal Analysis Using Summary Effect Estimates (CAUSE), using genome-wide significant single-nucleotide polymorphisms (SNPs) up to 5,890,484 participants, retrieved from large-scale genome-wide association studies (GWAS), as instrumental variables (IVS) to assess the associations of the risk of renal function and mental disease with five renal variables, including BUN (Blood Urea Nitrogen), EGFR (Estimated Glomerular Filtration Rate) of creatine, EGFR of cysteine, UACR (Urine Albumin-Creatinine Ratio), and urate.

Result: For forward MR,the LCV and CAUSE shows that MDD was associated with EGFR of creatine[GCP: -0.268, 0.417, p = 0.001], BIP was associated with UACR [GCP: 0.024, 0.476, p = 0.04].For reversed MR, Two-sample MR provided evidence that BIP was associated with higher odds of EGFR of cysteine [beta:0.001, 95%CI: 0.001, 0.002, p = 0.027]. While MDD was associated with lower BUN [beta: -0.003, 95%CI: -0.005, -0.001, p = 0.011] and higher EGFR of creatine [beta:0.001, 95%CI: 7.776, 0.002, p = 0.037] Conclusion:Our study provides further evidence that MDD is causally associated with higher odds of EGFR of creatine and lower BUN. Our results showed evidence of non-linear causal association between MDD and CKD; BIP and CKD

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3599. The effects of demographic-based selection bias on GWAS results in the UK Biobank

Authors:

T. Galama¹, S. van Alten², B. Domingue³, A. Marees², J. Faul⁴; ¹Univ. of Southern California, Los Angeles, CA, ²Vrije Univ.it Amsterdam, Amsterdam, Netherlands, ³Stanford, Stanford, CA, ⁴Univ. of Michigan, Ann Arbor, MI

Abstract Body:

The implications of selection bias due to volunteering (*volunteer bias*) for studies of genetic associations are poorly understood. Because of its large sample size and breadth of phenotypes, the UKB is included in almost all large GWASs as one of the largest cohorts, yet is known to be highly selected. In this paper, we develop inverse probability weighted GWAS (IPWGWAS) to estimate GWAS summary statistics in the UKB that are robust against volunteer bias. IPWGWAS results in effect sizes that are more predictive and reduces the odds of false positive results. Using IPWGWAS, we find (1) substantial increases in heritabilities for behavioral phenotypes (e.g., educational attainment, BMI, and age at first birth), (2) differences in genetic correlations between various behavioral phenotypes, and (3) show that some biological annotations estimated by GWAS are false positives that are driven by volunteer bias. For example, we show that the finding that SNPs associated with age at first birth are mainly expressed in the brain cerebellar hemisphere and the brain cerebellum, is entirely the result of volunteer bias. Further, we will assess whether volunteer bias can explain the surprisingly low within-sibship heritabilities of educational attainment and household income that are common in the literature. Our findings are useful for understanding the extent to which a particular phenotype is prone to volunteer bias in GWAS, and correction using IPWGWAS provides an alternative when population-representative cohorts are not available for GWAS analyses.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3600. The hidden factor: the importance of accounting for covariate effects in power and sample size computation when analyzing a binary trait.

Authors:

Z. Zhang, L. Sun; Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Accurate power and sample size estimation are crucial to the design and analysis of genome-wide association studies (GWAS). It is well known that replication studies with underestimated sample sizes can result in false negatives, missing truly associated SNPs. In GWAS of a binary trait via logistic regression, important covariates such as age and sex are typically included in the model. However, their effects are rarely properly considered, for example, in power or sample size computation for a successful replication study. Unlike when analyzing a continuous trait, the power of association analysis of a SNP with a binary trait also depends on covariate effects, even under the assumption of gene-environment (G-E) independence. Earlier methodological work recognizes this phenomenon but implemented methods are not flexible. We thus propose and implement a generalized method to honestly calculate the power of an association study, and correctly estimate the sample size necessary for a successful association study of a binary trait. The proposed method a) accommodates different types of non-genetic covariate E, b) deals with different types of G-E relationships, and c) is computationally efficient thus applicable to GWAS. Extensive simulation studies show that the proposed method is accurate and computationally efficient for both prospective and retrospective sampling designs with various covariate structures and G-E dependency. A proof-of-principle application to the UK Biobank data focused on the understudied $n = 3,460$ self-reported African participants; $n = 2,512$ after population principal component analysis. The application showed that ignoring covariate age effect led to significantly overestimated power (hence underestimated replication sample size) when analyzing the binary hypertension trait. In contrast, the computation for the continuous blood pressure trait is invariant to the covariate effect, as expected based on the analytical and simulation results. The R package *SPCompute* that implements the proposed method can be found at <https://cran.r-project.org/web/packages/SPCompute>.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3601*. The impact of 22q11.2 copy number variants on human traits in the general population.

Authors:

M. Zamarioli^{1,2}, C. Auwerx^{2,3,4,5}, M. Sadler^{4,2,3}, A. van der Graaf², K. Lepik², M. Moyses-Oliveira¹, A. Dantas¹, M. Melaragno¹, Z. Kutalik^{2,3,4}; ¹Univ.e Federal de São Paulo, São Paulo, Brazil, ²Univ. of Lausanne, Lausanne, Switzerland, ³Swiss Inst. of Bioinformatics, Lausanne, Switzerland, ⁴Univ. Ctr. for Primary Care and Publ. Hlth., Lausanne, Switzerland, ⁵Ctr. for Integrative Genomics, Univ. of Lausanne, Lausanne, Switzerland

Abstract Body:

Phenotypic consequences of 22q11.2 copy number variants (CNVs) have been demonstrated in clinical cohorts but have remained understudied in the general population. To address this gap, we performed a phenome-wide association scan in 405,324 unrelated UKBB participants using CNV calls from the genotyping array. To restrict the investigated traits, we mapped 236 Human Phenotype Ontology terms linked to any of the 90 genes encompassed by the region to 170 UKBB traits using cross-ontology mapping and web-scraping followed by manual curation. We then assessed the association between the copy-number state of 504 SNP-array probes in the region and 152 binary and 18 continuous traits using logistic and linear regression, respectively. Four association models were investigated: deletion-only, duplication-only, mirror (i.e., duplications and deletions have opposing effects) and U-shape (i.e., duplications and deletions have the same effect direction). We assessed the causal effect of the expression level of 39 testable (i.e. instrumentable) 22q11.2 genes on associated traits with transcriptome-wide mendelian randomization (TWMR). We also performed multivariable Mendelian randomization (MVMR) to assess the relationship among associated traits and infer if the pleiotropic effect of CNVs are vertical (indirect) or horizontal (genuine). Among the investigated participants, 1127 were duplication carriers and 694 had deletions within 22q11.2 low copy repeats A-D, with variable sizes. We found eight continuous and nine binary traits associated under different models revealing distinct impacts. Eight traits associated under the U-shape model, e.g. body-mass index (BMI; $\beta = 0.33 \text{ kg/m}^2$; $P = 4.9 \times 10^{-10}$), while three did so under the mirror model, e.g. mean platelet volume (MPV; $\beta = -0.54 \text{ femtolitres}$, $P = 1.3 \times 10^{-18}$). TWMR showed that increased expression of *ARVCF* increases BMI ($P = 10^{-4}$), concordantly with its encompassing CNV. Similarly, increased *DGCR6* levels causally reduced MPV ($P = 0.001$), in line with the corresponding CNV effect. Finally, cross-trait MVMR suggested a predominant role of horizontal pleiotropy. We have further showed that CNV probes in genes linked to a given HPO term are 15 times more likely ($P < 6 \times 10^{-9}$) to show significant association with the corresponding UKBB continuous trait. Our findings show that within the general population, 22q11.2 CNVs are associated with traits previously implicated by genes in the region, with duplications and deletions acting similarly or in opposite directions upon different traits. We also showed that gain or loss of distinct segments within 22q11.2 may impact a trait following different association models.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3602. The impact of genetic variation on hearing acuity variability

Authors:

J. Duran, L. Jorde; Univ. of Utah, Salt Lake City, UT

Abstract Body:

Although considerable research has been done on the genetics of deafness, relatively little is known about the genetics of normal variability in hearing acuity. We have analyzed genetic data from multigenerational families to assess the impact of genetic variation on hearing acuity variability at seven frequencies (250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz, 6000Hz, 8000Hz). Air conduction hearing acuities are measured as the lowest threshold (volume) at which an individual can hear a tone. We used a family-based methodology to estimate heritability and identify genetic variants that may be responsible for hearing acuity variability. The cohort consists of thirty-four large three- and four-generation Utah CEPH families who have been whole-genome sequenced (WGS) and screened for over 100 quantitative phenotypes. Using the SOLAR software package, we show that hearing acuity heritability ranges from 30% for 250hz and steadily decreases to 18% for both 6000Hz and 8000Hz. Hearing acuity is known to be impacted by environmental factors and may be the reason for decreased heritability at higher frequencies. Most twin studies estimate the heritability of hearing loss as 35% to 60%. A twin study measuring hearing acuity of individuals at least 65 years old estimated average hearing acuity heritability at 40%. Twin studies tend to overestimate heritability, so our family-based estimates are likely to be more accurate. Utilizing the WGS of the Utah CEPH cohort, we employed quantitative trait loci (QTL) mapping, which identifies loci predicted to be associated with changes in phenotypic variability. Our analysis uses single nucleotide polymorphisms (SNP)s as the variants of interest and accounts for covariates to predict QTLs. We identified 27 significant (Bonferroni-corrected $p < 0.05$) SNPs predicted to impact hearing acuity. In addition, chromosomes two, nine, and twelve are enriched for significant QTLs across the lower frequencies (250Hz, 500Hz, 1000Hz, and 2000Hz). When analyzing significant QTLs, we identified two SNP alleles segregating through two generations in at least one family. Further assessment of the identified QTLs is required to identify families that carry these alleles and whether they affect hearing acuity and segregate through families. We have identified approximately 15,000 individuals with hearing acuity assessment in the UK Biobank as a replication cohort to validate our findings.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3603. The impact of rare regulatory variation from epigenetics to protein

Authors:

T. Li¹, N. Ferraro², M. Cirnigliaro³, S. Arteaga³, L. Pérez-Cano⁴, J. Lowe³, B. Strober⁵, B. Ni⁵, NHLBI TOPMed Consortium, D. Geschwind³, S. Montgomery², A. Battle⁵; ¹Johns Hopkins Univ. Sch. of Med., Baltimore, MD, ²Stanford Univ., Stanford, CA, ³Univ. of California, Los Angeles, Los Angeles, CA, ⁴Stalicia DDS, Barcelona, Spain, ⁵Johns Hopkins Univ., Baltimore, MD

Abstract Body:

Any individual genome has thousands of rare genetic variants (RVs) and determining the subset that influence traits remains a key challenge due to the lack of a known regulatory code. We previously demonstrated that incorporating transcriptomic data can prioritize functional RVs as individuals with outlier gene expression are more likely to carry a nearby candidate RV. However, the transcriptome captures only a part of the regulatory cascade from genotype to phenotype, and it is unknown how RVs underlying other omic signals such as epigenetic changes or proteomics are related to disease risk. Here, we analyzed transcriptomic, methylomic, proteomic, and whole genome sequencing (WGS) data from 1,319 individuals in the TOPMed Multi-Ethnic Study of Atherosclerosis (MESA). We observed strong enrichment of RV burden nearby outlier signals in each molecular phenotype, where multimodal outliers show even stronger enrichments. We extended Watershed, a Bayesian hierarchical framework incorporating genomic annotations and observed outlier status for RNA expression and splicing to include methylation and protein expression levels to prioritize RVs. The multiomic Watershed model learned features corresponding to known variant biology and outperformed genomic annotation models in predicting regulatory status of pairs of individuals who share the same RVs nearby the same gene. When trained on European samples and evaluated on other populations, prediction accuracy remained similar (area under precision-recall curve for Europeans between [0.05, 0.14] and others [0.03, 0.16]), suggesting cross-population portability. Each omics signal prioritized a distinct set of variants which showed significantly larger effects on complex traits such as height, rheumatoid arthritis, Alzheimer's disease, and schizophrenia.

To further demonstrate the capability of Watershed to inform trait associations, we analyzed WGS data from 192 brain samples with paired transcriptomic data from the PsychENCODE UCLA Autism Spectrum Disorder (ASD) dataset. Watershed posteriors for splicing and protein signals significantly distinguishes individuals with idiopathic ASD from controls for SFARI Category I genes, suggesting that the multiomic Watershed model can be transferred to other tissues and disease contexts. Lastly, we observed that individuals with ASD harbor a higher number of brain-specific RVs prioritized by Watershed, demonstrating its ability to identify causal rare variants and mechanisms in ASD. Overall, our work presents a framework to integrate multiomic data and improve gene-trait prioritizations and genetic risk prediction through rare variants.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3604*. The largest deep-coverage whole genome seq meta-analysis on osteoporosis identified novel GWAS loci

Authors:

F. Yu¹, M-J. Tsai¹, H. Xu², M. Montasser³, C. Crandall⁴, C. Delaney⁵, M. Fu⁶, A. Justice⁷, D. Karasik¹, J. O'Connell³, C. Farber⁸, C. Watson⁵, J. Perry⁹, S. Salimi³, E. Streeten¹⁰, X. Zhang², R. Wallace¹¹, D. Evans¹², J. Tobias¹³, B. Richards¹⁴, J. Blangero¹⁵, B. Psaty⁵, K. Wiggins⁵, R. Kwon¹⁶, R. Jackson¹⁷, B. Mitchell¹⁸, D. Kiel¹⁹, C-T. Liu²⁰, TOPMed Bone Mineralization Working Group, Y-H. Hsu²¹; ¹Hebrew SeniorLife Inst. for Aging Res., Harvard Med. Sch., Boston, MA, ²Boston Univ., Boston, MA, ³Univ. of Maryland, Baltimore, MD, ⁴UCLA David Geffen Sch. of Med., Los Angeles, CA, ⁵Univ. of Washington, Seattle, WA, ⁶Univ Maryland, Baltimore, MD, ⁷Geisinger, Danville, PA, ⁸Univ. of Virginia, Charlottesville, VA, ⁹Univ. of Maryland Sch. of Med., Lewes, DE, ¹⁰Univ of Md Sch. of Med., Cockeysville, MD, ¹¹The Univ. of Iowa, Iowa, IA, ¹²Univ Queensland, Brisbane, Australia, ¹³Bristol Med. Sch., Bristol, United Kingdom, ¹⁴McGill Univ., Montreal, QC, Canada, ¹⁵Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX, ¹⁶Univ of Washington, Seattle, WA, ¹⁷Ohio State Coll. of Med., Columbus, OH, ¹⁸Univ Maryland, Baltimore, Baltimore, MD, ¹⁹Hebrew SeniorLife, Boston, MA, ²⁰Boston Univ. SPH, Boston, MA, ²¹Hebrew SeniorLife Inst. for Aging Res., Ha, Boston, MA

Abstract Body:

Osteoporosis is common in the elderly characterized by systematically reduced bone mineral density (BMD) and poor bone quality, which increases the risks of osteoporotic fracture, especially at hip and spine. Osteoporosis is diagnosed by BMD measured from DXA machines. Previously, we reported the largest common-variant GWAS analyses on estimated BMD from heel ultrasound (N~400K). However, the identified loci only explained ~20% variance of estimated BMD, comparing to the 85% heritability of DXA-derived BMD in twins. Thus, the current common-variant-based GWAS approach only detects a small fraction of genetic heritability. To explain the missing heritability, WGS has been incorporated to provide comprehensive enumeration of sequence variation. To identify less common, rare and structural variants that are associated with DXA BMD, we utilized the deep-coverage WGS (30X coverage) in 50k Caucasians (~40K from the UK Biobank study and ~10K from the NHLBI TOPMed Program) with clinical defined BMD derived from DXA at lumbar spine (LS) and femoral neck (FN) skeletal sites. From single variant association analysis, we identified 60 GWAS loci associated with FN BMD and 68 GWAS loci associated with LS BMD ($p < 5E-8$). Among them, the lead SNPs on 20 GWAS loci are missense variants. A subset of these variants was also associated with osteoporotic fracture in the same study samples. The most significantly associated locus with FN BMD was for SNP rs6886306 located in the intron of MEF2C (MAF=0.51, $\beta = -0.085$, $p = 9.26E-39$). The most significantly associated locus with LS BMD was for an Indel variant, SNP rs60521551, located in the intergenic region on chr13q14.1 locus (MAF=0.53, $\beta = 0.125$, $p = 3.94E-76$). For rare variant gene-based association analyses (MAF< 0.5%), we performed SKAT-O gene-based analysis to test the joint effect of the predicted to be functional variants annotated to protein-coding genes with the following variant filtering criteria: high-confidence loss-of-function variants, or missense variants with predicted pathogenicity. 18 genes were found to be genome-widely associated with either LS BMD or FN BMD ($p < 5E-6$), among which BCAS3, LRP5, and TMEM41B have been found to be associated with bone-related phenotypes; and the other 13 genes are novel association findings. These novel associated genes/loci are underway with CRISPR gene-editing experiments in the zebrafish model. The WGS provides fine resolution of sequence variation to further reveal the underlying genetic mechanism of less common/rare, structural, and missense variants involved in bone loss pathophysiological mechanisms and provided important clinical insight into osteoporosis treatment.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3605. The nature of nurture is not in the parents' genes: estimating indirect genetic effects from parents using imputed grandparental genotypes

Authors:

A. Young^{1,2}, R. Cheesman³, M. Lehtovirta⁴, G. Hatem⁵, J. Guan², S. Moeen Nehzati², H. Jayashankar⁶, L. Hakaaste⁴, T. Tuomi^{4,5}, A. Okbay⁷, D. Cesarini^{6,8,9}, D. Benjamin^{1,2,6}, R. Prasad^{4,5}, E. Ystrom^{3,10}; ¹Human Genetics Dept., UCLA David Geffen Sch. of Med., Los Angeles, CA, ²UCLA Anderson Sch. of Management, Los Angeles, CA, ³PROMENTA Res. Ctr., Dept. of Psychology, Univ. of Oslo, Oslo, Norway, ⁴Inst. for Molecular Med. Finland (FIMM), Univ. of Helsinki, Helsinki, Finland, ⁵Lund Univ. Diabetes Ctr. (LUDC), Malmö, Sweden, ⁶Natl. Bureau of Economic Res., Cambridge, MA, ⁷Dept. of Economics, Sch. of Business and Economics, Vrije Univ.it Amsterdam, Amsterdam, Netherlands, ⁸Dept. of Economics, New York Univ., New York, NY, ⁹Res. Inst. of Industrial Economics (IFN), Stockholm, Sweden, ¹⁰Dept. of Mental Disorders, Norwegian Inst. of Publ. Hlth., Oslo, Norway

Abstract Body:

Genetic associations are often interpreted as the result of direct genetic effects, i.e. effects of alleles in an individual on that individual. However, indirect genetic effects (IGEs) from relatives can also lead to genetic association: for example, if alleles in parents affect offspring through the environment, a phenomenon termed 'genetic nurture'. Previous work has suggested that IGEs from parents explain around 1/3rd of the association between educational attainment (EA) and polygenic predictors (PGSs) of EA. These studies have relied on fitting models that include the proband (phenotyped individual) and parents' PGSs. Because the variation in the proband PGS that is independent of the parents' PGSs is due to random segregations of genetic material in the parents, these models can isolate the component of the association between phenotype and PGS that is due to direct genetic effects. However, these studies are not able to isolate the component of the association due to IGEs from parents: the estimates are confounded due to the parental PGSs being correlated with other environmental and genetic factors due to population structure and assortative mating. While some studies have attempted to adjust for assortative mating, these adjustments rely upon strong assumptions that may not be valid. In this study, we extend the two generational model (proband and parents) to a three generational model including grandparents. Just as the two generation model can isolate the component of association due to direct effects, the three generation model can also isolate the component due to IGEs from parents. To overcome the lack of probands with genotyped parents and grandparents, we extend recently developed imputation methods to impute PGSs of ungenotyped grandparents from the observed grandparental genotypes and the genotypes of parents and their sibling(s). We apply our method to an educational attainment PGS derived from a 3 million person genetic association study, isolating the component of association due to IGEs from parents using data from the Norwegian Mother, Father and Child Cohort Study (MoBa) and Generational Scotland (GS). The standardized indirect genetic effect (i.e. partial correlation coefficient) estimate from MoBa is 0.01 (S.E.=0.037) and from GS is -0.02 (S.E.=0.068), compared to a standardized direct genetic effects of 0.23 (S.E.=0.010) in MoBa and 0.259 (S.E.=0.040) in GS. Our results show that IGEs likely explain only a small fraction of the association between EA and EA PGSs, implying that assortative mating and population stratification explain most of the apparent indirect genetic effect previously found.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3606. The relationship between kinship and heritability estimation accuracy

Authors:

D. Chen, J. Storey; Lewis-Sigler Inst. for Integrative Genomics, Princeton Univ., Princeton, NJ

Abstract Body:

Heritability, a quantity that captures the proportion of phenotypic variation explained by genetics, is fundamental to the study of the inheritance of human traits. Genome-wide association studies involving densely genotyped SNPs have made it possible to estimate heritability in population-based studies using linear mixed models (LMMs). Typically, heritability estimation fits a LMM utilizing a covariance based on a matrix of estimated kinship values, making the estimation of kinship values critical in this process. There are several approaches to estimating kinship. The accuracies of previous kinship estimates depend on the underlying structure and relatedness of the population, while the recently developed Ochoa-Storey estimator is suitable for arbitrary population structures. Here, we develop a framework to characterize the accuracy of genetic variance components and heritability estimation based on the accuracy of the kinship estimate. From this framework, we are able to characterize heritability estimation accuracy utilizing the standard kinship estimate as well as the recently developed Ochoa-Storey kinship estimate. We specifically show that (1) the standard kinship estimate results in downwardly biased heritability estimation, determined by the average of the true kinship coefficients and (2) the Ochoa-Storey kinship estimate results in nearly unbiased heritability estimation. We demonstrate our results on simulated and real data, including admixed populations.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3607. The role of sleep in human brain and heart health: an investigation using 40,000 brain and cardiac magnetic resonance images from the UK Biobank

Authors:

Z. Fan¹, Y. Li¹, X. Yang¹, J. Lin², Q. Wang³, J. Xie¹, P. Paschou¹, P. Drineas¹, T. Li⁴, H. Zhu⁴, B. Zhao¹; ¹Purdue Univ., West Lafayette, IN, ²Yale Univ., New Haven, CT, ³Peking Univ. Sixth Hosp., Beijing, China, ⁴Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract Body:

Increasing evidence indicates that poor sleep is a modifiable risk factor for various cardiovascular diseases and brain disorders. However, sleep's association with cardiac and brain structure and function, as well as their genetic overlapping, is not well understood. Here we present a systematic investigation of sleep-brain/heart connections using seven sleep traits and multi-modality cardiac and brain imaging data from over 40,000 subjects in the UK Biobank. Sleep conditions were phenotypically and genetically associated with heart structure, heart function, brain grey matter, brain white matter, and brain functional architectures. We prioritized important brain regions and functional networks associated with sleep and identified sleep-associated genomic loci that were colocalized with imaging measures of cardiovascular and brain systems. Mendelian randomization shows potential causal links between sleep traits and brain functional connectivity. In conclusion, biobank-scale imaging genetic data provides insights on the role of sleep in human health from a multi-organ perspective. An interactive web browser was developed to facilitate exploring our results.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3608*. The unusual suspects: genome-wide significant sex-difference in minor allele frequency in gnomAD v3.1.2 whole genome sequence via a novel population-aware retrospective regression

Authors:

L. Sun¹, Z. Wang², A. D. Paterson³; ¹Univ. of Toronto, Toronto, ON, Canada, ²Natl. Univ. of Singapore, Singapore, Singapore, ³Hosp. for Sick Children, Toronto, ON, Canada

Abstract Body:

Recently, a non-negligible proportion (~2%) of X chromosomal SNPs have been reported to have genome-wide significant sex-difference in minor allele frequency (sdMAF), through the analyses of high-coverage whole genome sequence in multiple populations and datasets, including the non-Finnish Europeans and African/African Americans from gnomAD v3.1.2 (Zhong et al., 2022).

The existing sdMAF test is intuitive, contrasting the MAF between female and male, but it is conservative when applied to multiple populations. Additionally, not all gnomAD populations nor the whole genome were analyzed.

We have developed a novel regression-based sdMAF testing framework, applicable to multiple populations, and to both the X chromosome and autosomes. Unusually, the proposed regression model is retrospective, regressing genotype on sex, population and sex-population interaction. Unusually, the regression model is linear despite the discrete nature of genotype data. Unusually, the variance of the regression is covariate-dependent (i.e. sex- and population-specific) to account for the potential MAF differences between sexes and populations.

When there is only one population, testing the regression coefficient of sex results in a Wald test that coincides with the existing sdMAF test. In the presence of multiple populations, different tests can be derived from the proposed multivariate retrospective regression model, depending on the scientific question. For example, to test for the presence of sdMAF in any population, one can jointly test the regression coefficients of the sex and sex-population interaction. To test for the difference in sdMAF between populations, one can test the coefficient of sex-population interaction. Additionally, the proposed test incorporates population-aware Hardy-Weinberg disequilibrium correction factors, making the test not conservative. Finally, if other factors (e.g. age and fine-scale population structure) are deemed important in sdMAF analysis, such factors can be readily incorporated in the proposed regression model as covariates.

We apply the proposed method to gnomAD v3.1.2, focusing on n~75,000 from seven populations with population-specific n>1000. We perform both population-stratified and -combined analyses, across the whole genome. We demonstrate a) genome-wide significant sdMAF predominantly is a X chromosomal phenomenon, b) there is a striking consistency across the populations among many SNPs with significant sdMAF, and finally c) there are some SNPs with notable differences in sdMAF between populations. These sdMAF findings potentially impact down-stream association analyses.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3609. The use of population-based common controls in GWAS of infectious diseases can result in biased association signals

Authors:

D. Duchon¹, C. Vergara¹, C. Thio², P. Kundu¹, N. Chatterjee¹, D. Thomas², G. Wojcik¹, P. Duggal¹; ¹Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, ²Johns Hopkins Sch. of Med., Baltimore, MD

Abstract Body:

Genome-wide association studies (GWAS) involving population-based common controls from biobanks, national cohorts, and large consortiums have identified genes and variants for a range of complex diseases. Utilizing population-based common controls can drastically increase sample size and statistical power with available genetic and phenotypic data. However, the use of these population-based common controls can also result in non-differential misclassification of the outcome attenuating true effect sizes. There has been increased attention to genetic studies of infectious diseases, like COVID-19, and the use of convenient population-based global controls. However, there has been concern about a lack of phenotypic or pathogen exposure information on these controls and what effect that may have on the results. Through simulation, we show that infectious disease outcomes that use population-based common controls with unknown pathogen exposure can result in biased effect estimates and spurious signals of genome-wide significance. We demonstrate that the magnitude of this bias depends upon the strength of the association between a locus and pathogen exposure and also the prevalence of the pathogen in the community. For example, in simulated comparisons between 20,000 cases and 20,000 pathogen exposed controls or common controls, no spurious associations are observed for a locus strongly associated with pathogen exposure when cases are compared to well-characterized controls regardless of pathogen prevalence or common controls when pathogen exposure is universal. However, when prevalence is rare (5%) or common (50%), >99% and >20% of simulated replicates, respectively, result in spurious associations ($P < 5 \times 10^{-8}$) between a locus strongly associated with pathogen exposure and the outcome. We also compare the results of an empirical GWAS of hepatitis C Virus (HCV) clearance using well-characterized persistently infected controls to population-based controls from the UK Biobank. We find biased effect estimates for known HCV clearance-associated loci and potentially spurious associations. These findings suggest that the use of population-based common controls may be inappropriate for outcomes that are conditional upon environmental exposure, including infectious diseases, and that results from infectious disease GWAS involving common controls be interpreted with caution.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3610*. Tissue-specific impacts of aging and genetics on gene expression patterns in humans.

Authors:

R. Chung¹, R. Yamamoto², J. Vazquez¹, H. Sheng¹, P. Steinberg¹, N. Ioannidis¹, P. Sudmant¹; ¹Univ. of California, Berkeley, Berkeley, CA, ²Univ. of California, Los Angeles, Los Angeles, CA

Abstract Body:

Age is the primary risk factor for many common human diseases including heart disease, Alzheimer's dementias, cancers, and diabetes. Determining how and why tissues age differently is key to understanding the onset and progression of such pathologies. Here, we set out to quantify the relative contributions of genetics and aging to gene expression patterns from data collected across 27 tissues from 948 humans. We show that age impacts the predictive power of expression quantitative trait loci across several tissues. Jointly modelling the contributions of age and genetics to transcript level variation we find that the heritability (h^2) of gene expression is largely consistent among tissues. In contrast, the average contribution of aging to gene expression variance varied by more than 20-fold among tissues with 5 tissues having a larger proportion of age-related gene expression than heritable gene expression. We find that the coordinated decline of mitochondrial and translation factors is a widespread signature of aging across tissues. Finally, we show that while in general the force of purifying selection is stronger on genes expressed early in life compared to late in life as predicted by Medawar's hypothesis, a handful of highly proliferative tissues exhibit the opposite pattern. These non-Medawarian tissues exhibit high rates of cancer and age-of-expression associated somatic mutations in cancer. In contrast, gene expression variation that is under genetic control is strongly enriched for genes under relaxed constraint. Together we present a novel framework for predicting gene expression phenotypes from genetics and age and provide insights into the tissue-specific relative contributions of genes and the environment to phenotypes of aging.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3611. Tools for efficient manipulation and meta-analysis of summary statistics from multi-exposure, genome-wide, gene-environment interaction studies

Authors:

K. Westerman; Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

Summary statistics from genome-wide association studies enable valuable downstream analyses such as heritability estimation, enrichment testing, and meta-analysis. Genome-wide interaction studies (GWIS), are increasingly being performed to generate similar summary statistics describing environmental modification of genetic effects. Our recently published software program, GEM, enables GWIS models with multiple interaction exposures (for increased power), and adjustment for gene-covariate interactions (to avoid hidden confounding). However, the software infrastructure does not yet exist to harmonize and meta-analyze GWIS summary statistics in the context of these more complex analytical designs. Here, we introduce methods and associated tools for efficient manipulation of multi-exposure summary statistics based on a single GWIS run: REGEM enables the assessment of alternative model configurations (partitioning exposures into interactions, interaction covariates, and standard covariates) using only summary statistics, and METAGEM performs meta-analysis allowing for multiple exposures. We deployed these tools to conduct a multi-ancestry analysis of genetic interactions influencing waist-hip ratio (WHR), incorporating multiple exposures (both sex and body mass index [BMI]). In six ancestry groups from the UK Biobank ($N = 381,089$), we conducted both ancestry-stratified and pooled GWIS with individual-level data for common, autosomal genetic variants using GEM. Next, using REGEM and METAGEM with these multi-exposure summary statistics, we derived equivalent results corresponding to single-exposure tests (sex or BMI only), and conducted multi-exposure, multi-ancestry meta-analysis. First, we found that GxAcestry interaction covariates were necessary to control genomic inflation of interaction estimates due to population stratification in the pooled analysis. Second, the pooled analysis found the same set of loci as meta-analysis; this comparison was enabled by the multi-exposure meta-analysis capability of METAGEM and indicates the potential for more inclusive and powerful pooled GWIS with multiple ancestries. Third, the derivation of corresponding single-exposure summary statistics (e.g., sex-only GEI estimates) for comparison to the multi-exposure results required only 0.65 CPU-hrs rather than the 350 CPU-hrs required to re-run the equivalent individual-level GWIS. Our suite of tools, GEM, REGEM and METAGEM, provides key software infrastructure for maximizing the utility of summary statistics from ancestrally diverse and analytically complex GWIS studies.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3613*. Tradeoff between prediction accuracy and transferability in the design of polygenic risk scores.

Authors:

A. Dominguez¹, Y. Zhang¹, C. Pato², S. McCarroll³, M. Boehnke¹, L. Scott¹, S. Zollner¹; ¹Univ. of Michigan, Ann Arbor, MI, ²Rutgers Univ., Newark, NJ, ³Harvard Med. Sch., Boston, MA

Abstract Body:

Polygenic risk scores (PRS) are commonly used to estimate individual risk for a given complex trait. PRS include each marker's marginal association with the trait by using summary statistics from genome-wide association studies (GWAS). However, most GWAS participants are of European (EUR) descent, which biases risk estimates for non-EUR populations. The bias depends on the construction of the PRS, but how parameters such as number of markers affect this bias is not well understood. In this study, we investigate the influence of markers on predictive accuracy and transferability of PRS of bipolar disorder (BPD).

Using clumping & p-value thresholding and summary statistics from the Psychiatric Genomics Consortium (PGC), we calculate PRS for 3,334 African Americans from InPSYght (1,224 cases/2,110 controls) and 5,670 EUR Americans from Prechter Bipolar Study (909 cases) & Michigan Genome Initiative (1,186 cases/3,575 controls) (PMGI). We combine our sample's genomic data with 1,000 Genomes Project (1KGP) and perform principal component analysis to obtain continuous measures of ancestry. To assess predictive performance of PRS, we use logistic regression models to calculate Nagelkerke's R-Squared (R^2) in InPSYght and PMGI separately while adjusting for sex and first five principal components (PCs). We observe increasing predictive accuracy with more markers and maxima at threshold $p < 0.9$ ($R^2 = 2.4\%$) for InPSYght and $p < 10^{-3}$ for PMGI ($R^2 = 4.1\%$).

To assess the impact of ancestry PCs on PRS, we use linear regression models to calculate R^2 while adjusting for sex and affection status. For thresholds near $p < 10^{-6}$, PRS are broadly independent of PCs in both samples ($R^2 < 2\%$). For thresholds between $p < 0.05$ and $p < 0.9$, PCs explain much more variance in PRS for InPSYght ($R^2 > 3\%$) and PMGI ($R^2 > 16\%$). Furthermore, the mean PRS for InPSYght is consistently larger than PMGI and the difference increases with more markers.

To evaluate whether our findings are exclusive to BPD, we calculate PRS for our sample and 1KGP for three additional traits: height (GIANT), schizophrenia (PGC), and type 2 diabetes (DIAGRAM/GERA/UKB). We observe substantial differences in mean PRS across all traits which are exacerbated with larger thresholds, indicating systematic differences in allele frequencies between populations for the underlying variants. Lastly, variance explained by PCs increases as PRS are constructed with less stringent thresholds.

Our findings establish a tension between having complex, informative PRS with more markers and their susceptibility to population structure. This results in more accurate PRS for individuals with EUR ancestry but biased PRS for more diverse ancestries.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3614. Transcript-aware gene burden analyses in the UK Biobank reveal isoform-specific compositions and functions

Authors:

D. Jakubosky, G. McInnes, A. Leverentz, Z. Romer, B. Gonzalez, N. Karra, B. Cajés, S. Bruse, O. Gottesman; Empirico Inc., San Diego, CA

Abstract Body:

A commonly utilized method for rare variant association analysis involves aggregating carriers of predicted-deleterious coding variants in a given gene. So-called gene burden tests have become increasingly necessary with the availability of large-scale biobanks comprising whole exome and whole genome sequencing data linked to rich phenotypic information, such as the UK Biobank, where protein-altering sequence variants are often too rare to be tested individually. Previous studies have generally focused on burdens that aggregate variants across all protein-coding transcripts of a gene in a single test. However, alternative splicing of genes is an essential part of gene regulation and transcript isoforms are expressed in a tissue-specific and context-specific manner, a process that has a critical role in both development and disease. Sequence variants may therefore only be present in, or have an impact on, one (or a subset) of protein-coding transcripts; and in cases where transcripts have distinct biological functions, the performance of a pan-transcript burden test may suffer. We utilized whole exome sequencing data from the UK Biobank to generate transcript-specific deleterious variant burdens, representing 19,037 protein-coding genes, derived from a set of 3,761,661 variants (MAF <0.05) observed in 421,226 individuals. Among the represented genes, more than 70% had multiple protein-coding transcripts with unique sets of variants in their transcript-specific burdens. We further evaluated the similarity of all pairs of transcript-specific burdens within genes. Depending on the stringency of deleteriousness predictions for variants included, approximately 50% of tested genes had at least one highly-divergent pair of transcript-specific burdens, characterized by a transcript having at least 5 unique variants that account for at least 50% of variants linked to that transcript. Transcript-specific burdens were tested for association with thousands of binary and quantitative phenotypes, revealing a substantial number of associations that would not have been detected using traditional pan-transcript gene burden tests and suggesting a high prevalence of transcript isoforms with distinct functions. This study provides important insight into the relevance of both composition and function of gene isoforms when conducting gene burden analyses, and highlights the value of deriving transcript-aware gene burdens to maximize the utility of rare variant association analysis in large-scale biobanks. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3615. Transcriptome and GWAS identified disease-specific eosinophilia-pathways and their upstream regulators in primary biliary cholangitis.

Authors:

K. Ueno¹, Y. Hitomi¹, Y. AIBA², O. Gervais³, Y. Kawai¹, S-S. Khor⁴, M. Kawashima⁵, N. Nishida⁶, K. Kojima⁷, M. Nagasaki⁸, K. Tokunaga¹, M. NAKAMURA⁹; ¹Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, ²Nagasaki Med. Ctr., Omura, Japan, ³Kyoto Univ., Mishima, Japan, ⁴Natl. Ctr. for Global Hlth.and Med., Tokyo, Tokyo, Japan, ⁵Japan Sci. and Technology Agency, Chiyoda-ku, Japan, ⁶Natl. Ctr. for Global Hlth.and Med., Ichikawa, Chiba, Japan, ⁷Tohoku Univ., Sendai, Japan, ⁸Kyoto Univ., Kyoto-City, Japan, ⁹NHO Nagasaki Med. Ctr., Omura, Nagasaki, Japan

Abstract Body:

Primary biliary cholangitis (PBC) is a chronic cholestatic liver disease presumably by autoimmune mechanism. Our recent transcriptome analysis has identified that mRNA expression associated with eosinophilia is increased in PBC as compared to chronic hepatitis C (CHC). The aim of this study is to investigate the molecular mechanisms of eosinophilia in PBC. Firstly, we identified 256 genes that are significantly increased in PBC as compared to CHC by mRNA-microarray of liver samples. Downstream analysis with Ingenuity Pathway Analysis (IPA) of these 256 genes revealed that top 3 pathways are associated with eosinophilia and 11 genes (*CLC*, *IFNG*, etc) were categorized in eosinophilia. The analysis of upstream regulators by IPA of these 11 genes revealed that *IFNG* was the top regulator. The mRNA expression of *IFNG* was significantly correlated with the mRNA expression of six genes such as *BCL2A1* indicating that *IFNG* plays a central role in eosinophil infiltration in PBC as compared to CHC. Secondly, we investigated the cause of eosinophilia using two GWAS datasets from European and Japanese populations. We identified one locus *CLC* which was significantly associated with eosinophil counts in the peripheral blood in GWAS. *CLC* is abundantly expressed in eosinophils and in silico analysis identified *rs2238678* in *CLC* locus as a causal variant and its effector gene was also identified as *CLC* in GTEx. Subsequent analysis indicated that *rs2238678* is a functional SNP. G allele of this SNP was associated with increased number of eosinophils in the two GWAS. These results indicated that genetic polymorphism of eosinophil-constituents has some role in eosinophil differentiation. Thirdly, transcription factors were predicted from the three gene datasets (a transcriptome data and two GWAS data). As a result, androgen receptor was predicted as the most significant common upstream regulator. We also utilized the expression data from peripheral blood of PBC and extracted 102 genes whose expression was significantly correlated with *CLC* expression. Three genes (*IL4*, *IL5*, and *GATA2*) which are well known major factors in differentiation of eosinophil were predicted as the top regulators, indicating that differentiation and activation of eosinophil might occur in peripheral blood of PBC. In conclusion, the mechanisms of eosinophilia in PBC are a complex phenomenon which involve activation of *IFNG* pathway in the liver, activation of *IL4*, *IL5*, *GATA2* pathways in the peripheral blood, the genetic polymorphism of eosinophil-constituents such as *CLC*, and regulation by sex hormone such as androgen. Further studies are needed to clarify the role of eosinophilia in the pathogenesis of PBC.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3616*. Transcriptomics-based drug-repurposing screen discovers potential novel therapies for primary open-angle glaucoma.

Authors:

K. Duchinski^{1,2,3}, J. Rouhana^{2,4,3}, S. Huggett^{5,6}, A. Segre^{2,4,3}; ¹Dept. of BioMed. Informatics, Harvard Med. Sch., Boston, MA, ²Ocular Genomics Inst., Dept. of Ophthalmology, Massachusetts Eye and Ear Infirmary, Boston, MA, ³Broad Inst. of Harvard and MIT, Cambridge, MA, ⁴Dept. of Ophthalmology, Harvard Med. Sch., Boston, MA, ⁵Behavioral Genetics of Addiction Lab., Emory Univ., Atlanta, GA, ⁶5AM Ventures, Cambridge, MA

Abstract Body:

Primary open-angle glaucoma (POAG), a heritable neurodegenerative eye disease, is the second leading cause of irreversible blindness worldwide. In POAG, gradual degeneration of the optic nerve that transmits visual stimuli from the retina to the brain, results in vision field loss. Increased intraocular pressure (IOP) is the main risk factor for POAG. Current treatments attempt to delay POAG progression via IOP-lowering medications. However, up to 40% of POAG patients do not have elevated IOP. We devised a drug-repurposing approach that searches for drugs whose perturbational gene expression profiles anti-correlate with genetically-determined expression profiles associated with disease risk. Such compounds may exert protective or therapeutic effects on disease. We used the LINCS L1000 project that measured the effects of 1,796 FDA-approved and experimental small molecules on the expression of 12,328 genes in >80 human cell lines. We statistically inferred expression changes associated with POAG risk by applying S-PrediXcan, a transcriptome-wide association study (TWAS) method, to the summary statistics of a trans-ancestry POAG genome-wide association study (GWAS) meta-analysis of 34,179 cases and 349,321 controls (International Glaucoma Genetics Consortium), using a transcriptome prediction model trained on 403 peripheral retina RNA-seq and imputed genotype data (Ratnapriya et al., 2019). We next computed the heritability-weighted Pearson correlation between POAG TWAS z-scores and 38,479 independent LINCS drug perturbation signatures ($R^2 < 0.8$) across 910 compounds for 184 genes with significant TWAS p-values ($P < 0.01$). We applied a multi-level meta-regression model of the correlation coefficients as a function of drugs, adding cell line as a random effect. We identified 38 small molecules, including anti-inflammatory and anti-cancer drugs, whose perturbation signatures were significantly anti-correlated with gene expression associations with POAG ($P < 1E-06$). Literature review of the top small molecules' gene targets revealed promising links to POAG. Future work will entail building POAG expression profiles based on relevant GTEx tissues. To validate our approach, we analyzed a GTEx aorta TWAS of a cross-ancestry systolic blood pressure (SBP) GWAS (Million Veterans Program). The leading result for SBP was the widely-used antihypertensive drug amlodipine. This work suggests novel POAG treatments for experimental follow-up and presents a promising genetically and transcriptionally-driven approach for identifying new treatments, which can be applied to any complex disease with available GWAS summary statistics.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3617. Transferability of polygenic scores for anthropometric traits to the Native Hawaiian population

Authors:

Y-C. Lo¹, S. Jeon¹, L. Wilkens², L. Le Marchand², C. Haiman¹, C. Chiang^{1,3}; ¹Ctr. for Genetic Epidemiology, Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, ²Epidemiology Program, Univ. of Hawai'i Cancer Ctr., Univ. of Hawai'i, Honolulu, HI, ³Dept. of Quantitative and Computational Biology, Univ. of Southern California, Los Angeles, CA

Abstract Body:

Polygenic scores (PS) are promising in predicting or stratifying individuals based on the genetic susceptibility to common diseases or complex traits. However, as large-scale genome-wide association studies (GWAS) are heavily biased towards European-ancestry individuals, there are concerns that PS models trained largely in European-ancestry populations would not transfer well to populations of different ancestries. This poor transferability has been demonstrated for a number of ethnic minority populations, but have not been evaluated for Native Hawaiians, who makes up 0.5% of the U.S. population and is the second-fastest-growing ethnic minority in the U.S. Native Hawaiians are largely admixed, having ancestry components from both European and East Asian ancestry, among others. It is therefore also unclear whether the admixture alleviates some of the transferability issues of PS observed in other ethnic populations. Using height and BMI as examples of highly polygenic human traits, we evaluated the transferability of PS to Native Hawaiian populations. We obtained the summary statistics from currently largest GWAS for each anthropometric traits from Europe-ancestry cohorts and East Asian-ancestry cohorts. We trained a genome-wide PS model for each trait using the pruning-and-thresholding approach and evaluated each model in an out-of-sample cohort of Japanese, White, and Native Hawaiian individuals from the Multiethnic Cohort (MEC). For PS models trained with GWAS summary statistics from Europe-ancestry cohorts (the GIANT consortium and UK Biobank), the models can predict better in out-of-sample MEC non-Latino white individuals (partial $r^2 = 0.31$ and 0.097 for height and BMI, respectively) compared to Native Hawaiians (partial $r^2 = 0.18$ and 0.081) or Japanese (partial $r^2 = 0.15$ and 0.064). Moreover, PS model trained with GWAS summary statistics from East Asian-ancestry cohorts (Biobank Japan) can predict BMI approximately similarly in MEC Japanese individuals (partial $r^2 = 0.025$) compared to Native Hawaiians (partial $r^2 = 0.036$). Our results thus confirm a degree of loss of transferability of PS models to Native Hawaiians for highly polygenic traits, although admixture from European and East Asian-ancestry components in Native Hawaiians may also have helped with prediction based on PS trained in GWAS datasets from other ancestries. Nevertheless, increasing representation of Native Hawaiians in genetic studies, better modeling of the differences in ancestry or linkage disequilibrium patterns, or further incorporating of non-genetic risk factors will be helpful to improve risk predictions.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3618. Transformer generative adversarial networks and variational autoencoder for causal analysis of genetic variation in the presence of unobserved confounding

Authors:

J. Zhao¹, T. Xu¹, M. Xiong²; ¹Univ Florida, Gainesville, FL, ²Univ Texas Sch. of Publ. Hlth., Houston, TX

Abstract Body:

Despite significant progress in genetic association studies (GWAS), the signals identified by association analysis may not have specific pathological relevance to diseases so that a large fraction of disease-causing genetic variants is still hidden. Association signals provide limited information on the causal mechanism of diseases. The use of association analysis as a major analytical platform for genetic studies of complex diseases is a key issue that may hamper discovery of disease mechanisms, calling into the questions the ability of GWAS to identify loci-underlying diseases. It is time to move beyond association analysis toward techniques, which enables the discovery of the underlying causal genetic structures of complex diseases. To achieve this, we develop transformer generative adversarial networks and variational autoencoder as a general framework for uncovering causal clinical relevance of genetic variation in the presence of unobserved confounding. Proxy variables such as omics or imaging data can be used to approximate the unobserved confounding. Deep latent variable models and representation learning with proxy variables will be developed to uncover the unobserved confounding. Deep learning-based classifier two sample test will be developed to rigorously test causation of the genetic variant with the disease. Large simulations will be performed to calculate type I error rates and power of the proposed method for testing causation. The proposed method will be applied to genetic studies of Alzheimer's disease. We anticipate that our analysis will stimulate serious discussion of paradigm shift from association to causation.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3619. TWAS of the X Chromosome in Neurodegeneration

Authors:

P. Evans; Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract Body:

16.00 Normal 0 false false false EN-US X-NONE X-NONE /* Style Definitions */
table.MsoNormalTable {mso-style-name:"Table Normal"; mso-tstyle-rowband-size:0; mso-tstyle-colband-size:0; mso-style-noshow:yes; mso-style-priority:99; mso-style-parent:""; mso-padding-alt:0in 5.4pt 0in 5.4pt; mso-para-margin-top:0in; mso-para-margin-right:0in; mso-para-margin-bottom:8.0pt; mso-para-margin-left:0in; line-height:107%; mso-pagination:widow-orphan; font-size:11.0pt; font-family:"Calibri",sans-serif; mso-ascii-font-family:Calibri; mso-ascii-theme-font:minor-latin; mso-hansi-font-family:Calibri; mso-hansi-theme-font:minor-latin; mso-bidi-font-family:"Times New Roman"; mso-bidi-theme-font:minor-bidi;} Neurodegeneration is a key phenotype in aging and in diseases such as Parkinson's and Alzheimer's disease. The X chromosome is enriched for genes functioning in the brain and has many links to brain physiology and development. However, the X chromosome is also an often overlooked chromosome in analysis of the genetic etiology of diseases. We performed analysis of the X chromosome in the Vanderbilt University Medical Center biobank samples, called BioVU, across neurodegeneration phenotypes. Genome-wide association analysis (GWAS) was performed for each neurodegeneration phenotype within BioVU. GWAS results were then used in a combined polygenic risk score analysis within BioVU for neurodegenerative traits using existing brain phenotype GWAS as training sets. Finally, predicted expression of each gene on the X chromosome was calculated from models using elastic net from the Genotype-Tissue Expression (GTEx) consortium data. The models were then applied to all samples in BioVU and transcriptome-wide association analysis (TWAS) was performed on all neurodegenerative phenotypes. Results of these analyses will be presented and the function of implicated genes will be discussed.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3620. Uncovering ancestry-specific differences in genetically driven transcriptomic dysregulation in schizophrenia and bipolar disorder.

Authors:

D. Mathur¹, S. Venkatesh¹, T. Bigdeli², K. Therrien¹, D. Burstein¹, A. Fanous³, P. Harvey⁴, G. Voloudakis¹, P. Roussos¹; ¹Icahn Sch. of Med. at Mount Sinai, New York, NY, ²SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, ³Washington VA Med. Ctr., Washington, DC, ⁴Miami VA Hlth.care System, Miami, FL

Abstract Body:

Genome-wide association analyses (GWASs) identify genetic variants associated with disease; however, translating genetic findings to downstream applications is challenging since most variants reside in non-coding regions. Transcriptome-wide association studies (TWASs) leverage tissue-specific molecular profiling data (for the construction of transcriptomic imputation models) to bridge this gap by converting genetic risk variation into transcript-level dysregulation. However, the current status quo as it pertains to the use of TWAS for gene discovery is limited to the usage of GWAS and transcriptomic imputation models of European (EUR) ancestry. Here, we explore the shared and unique components of genetically driven transcriptomic dysregulation between EUR and African (AFR) ancestry for schizophrenia and bipolar disorder.

We leverage molecular profiling data (genotypes and gene expression) from CommonMind Consortium dorsolateral prefrontal cortex brain samples comprising 925 EUR and 276 AFR individuals to train ancestry-specific transcriptomic imputation models. To compare power, an additional EUR imputation model is generated using 276 individuals matched in terms of sample size, age, sex, and disease status to the individuals making up the AFR model. We assess the trans-ancestry predictive performance of the EUR transcriptomic imputation model to predict observed AFR expression and vice versa. We then perform cis-ancestry summary-level schizophrenia and bipolar disorder TWAS for EUR and AFR. Finally, using the cis-ancestry TWASs as ground truth, we (1) explore shared and unique gene and pathway dysregulation in these disorders and (2) perform trans-ancestry TWASs to evaluate the loss of power when using models of different ancestry.

We find greater convergence of gene expression dysregulation at the gene and pathway level in schizophrenia than in bipolar disorder between ancestries. Trans-ancestry TWASs exhibit significant loss of power but overall maintain moderate correlation in top genes.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3621. Unpicking the Gordian knot: Mendelian randomisation identifies priority groups for prophylactic EBV vaccination.

Authors:

M. Muckian¹, J. F. Wilson¹, G. S. Taylor², H. R. Stagg¹, N. Pirastu³; ¹Univ. of Edinburgh, Edinburgh, United Kingdom, ²Univ. of Birmingham, Birmingham, United Kingdom, ³Human Technopole, Milan, Italy

Abstract Body:

Epstein Barr virus (EBV) infects ~95% of the population worldwide and is known to cause adverse health outcomes such as Hodgkin's, non-Hodgkin's lymphomas, and multiple sclerosis. There is substantial interest and investment in developing infection-preventing vaccines for EBV. To effectively deploy such vaccines, it is vital that we understand the risk factors for infection. The current literature, describes complex, often conflicting webs of intersecting factors- sociodemographic, clinical, genetic, environmental- rendering causality difficult to decipher. We aimed to use Mendelian Randomization (MR) to overcome the issues posed by confounding and reverse causality to determine the risk factors for the acquisition of EBV. We mapped the complex evidence from the literature on factors associated with EBV serostatus (as a proxy for infection) into a causal diagram to determine putative risk factors for use in our study. Using data from the UK Biobank of 8,422 individuals genomically deemed to be of white British ancestry between the ages of 40 and 69 years at recruitment between the years 2006 and 2010, we performed a genome wide association study (GWAS) of EBV serostatus, followed by a Two Sample MR to determine which putative risk factors were causal. Our GWAS identified two novel loci associated with EBV serostatus (rs71449058, rs1210063). In MR analyses, we confirmed shorter time in education, an increase in number of sexual partners, and a lower age of smoking commencement, to be causal risk factors for EBV serostatus. Given the current interest and likelihood of a future EBV vaccine, our findings on causal factors for EBV acquisition are informative for future vaccine development and deployment. Knowing these risk factors aids the determination of who to prioritise when a vaccine is introduced. We also highlight the power and value of MR for untangling infectious disease risk factors.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3622. Unravelling the genetic contributions to gestational duration in Autism Spectrum Disorder individuals: results from the SPARK sample.

Authors:

C. Chatzigeorgiou, M. Janecka; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract Body:

Autism spectrum disorder (ASD) is a highly heritable neurodevelopmental disorder with complex etiology. While previous epidemiological studies have identified prematurity as a risk factor for the disorder, the mechanisms underlying potential causal effects of this exposure remain unknown. Given that the duration of pregnancy in the general population is influenced by fetal and maternal genetic and non-genetic factors, and our earlier work demonstrating phenotypic differences between ASD cases born term and pre-term, we investigated systematic genetic differences between ASD cases born term and pre-term. Utilizing a sample of ASD cases with and without the exposure allowed us to (1) uncover variants which potentially exert their effects through interacting with other genetic/non-genetic factors enriched in ASD, and (2) elucidate potential molecular differences in ASD associated with this early-life exposure. We first conducted a genome wide association study (GWAS) using 993 ASD cases born pre-term (cases) and 7,464 unrelated cases of ASD born term (controls) identified in the Simons Foundation Powering Autism Research for Knowledge (SPARK) dataset. Polygenic risk scores (PRSs) for pre-term birth were calculated based on Liu *et al*, summary statistics using PRSice-2. We observed an association between pre-term birth in ASD cases. To advance biological insight among the set of significant variants and map them to a smaller set of disease-relevant genes and pathways, functional genomic resources were used to functionally annotate variants and map them to genes (<https://fuma.ctglab.nl/>). These analyses revealed 76 possible causal genes identified by positional mapping and expression quantitative trait locus mapping, including top hits *PTPRD* ($p= 1.09 \times 10^{-13}$, previously shown to interact with maternal stress to increase the risk of pre-term birth), *PCDH15* ($p=2.97 \times 10^{-12}$ previously associated with neuropsychiatric disorders) and *DLGAP1* ($p=1.95 \times 10^{-12}$ previously linked to schizophrenia and other brain disorders). MAGMA tissue expression analysis identified enrichment across brain tissues and blood vessels. In conclusion, our findings revealed possible causal genes underlying the phenotypic presentation in the pre-term ASD cases. As the top hit identified in our analyses *PTPRD* has been associated with pre-term when interacting with maternal stress, the analyses are ongoing to validate potential interactive nature of this gene, and other ones identified in our analyses, and their relevance to ASD etiology.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3623. Unsupervised representation learning significantly improves genomic discovery for lung function and respiratory disease prediction.

Authors:

T. Yun¹, B. Behsaz¹, J. Cosentino², A. Carroll³, C. Y. McLean¹, F. Hormozdiari¹; ¹Google Res., Cambridge, MA, ²Google Res., San Francisco, CA, ³Google Res., Palo Alto, CA

Abstract Body:

High dimensional clinical data (e.g. time series, images) provide a unique opportunity in genetic studies of complex traits; however, we lack statistical methods to fully utilize them in association studies. We introduce a novel method to discover associations between genotypes and high dim phenotypes by learning low dimensional, nonlinear representations with unsupervised deep learning. Using variational autoencoders (VAE) incentivized to learn encodings that have uncorrelated coordinates, separable biological factors can be better captured in each coordinate. We perform GWAS on each coordinate to study the genetics of the underlying phenotype, agnostic to manually defined labels.

We demonstrate this approach on spirometry tests for lung function, and generate synthetic phenotypes we call spirogram encodings (SPINCs) and residual spirogram encodings (RSPINCs). SPINCs use a standard convolutional VAE, while RSPINCs use a novel VAE modified to admit 5 widely used spirometry features (FEV1, FVC, FEV1/FVC, PEF, FEF25-75%) in the encodings, encouraging remaining coordinates to capture residual signals not represented by those 5 manual features (MFs).

From GWAS on just 2 RSPINCs in UK Biobank (n=325K), which carry substantial heritability ($h^2 = 0.161 \pm 0.011, 0.045 \pm 0.003$), we discover 74 additional genome-wide significant (GWS) loci, in addition to 724 GWS loci from GWAS on 5 MFs. With SPINCs (dim=5) we discover 634 GWS loci, 141 of which are novel with respect to GWAS on 5 MFs. We also compare SPINCs (dim=5) to 5 principal components (PCs) of raw spirogram values. GWAS on SPINCs generate 22% more GWS loci than PCs (634 vs 521) and 3.28 \times more novel loci (141 vs 43).

We also construct a set of polygenic scores (PGS) on (R)SPINCs as general genetic predictors of lung function. To study their predictive capacity, we consider two lung diseases, asthma and COPD, and evaluate the accuracy of a linear model on PGS in a held-out set in 3 metrics: prevalence in the top PGS decile, AUC, AUPRC. We observed a 9% increase in asthma prevalence in top PGS decile by using 5 SPINCs PGS instead of 5 MF PGS, and a 6% increase by adding 2 RSPINCs PGS to 5 MF PGS ($p < 0.05$ for the differences). In fact, 5 SPINCs PGS outperforms 5 MF PGS in all metrics for both diseases ($p < 0.05$ for asthma). With RSPINCs, adding 2 RSPINCs PGS to 5 MF PGS again improves upon using just the 5 MF PGS in all metrics for both diseases ($p < 0.05$ for all but COPD prevalence). This general set of PGS can be used for any trait associated with spirometry, especially when direct GWAS is underpowered due to insufficient labels. We believe our approach can become a standard method for using high dim phenotypes for GWAS and prediction.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3624. Using external reference panel and meta-analysis summary statistics for rare-variant aggregation tests.

Authors:

B. Ryan¹, R. Welch¹, M. Boehnke¹, C. Fuchsberger²; ¹Univ. of Michigan, Ann Arbor, MI, ²Eurac Res., Bolzano, Italy

Abstract Body:

Genome-wide association studies (GWAS) have identified hundreds of thousands of associations between common genetic variants and a wide range of human diseases and traits. These studies are often of low power to identify associations with rare genetic variants, which are thought to contribute to the heritability of many common diseases and traits. Aggregation tests pool the genetic signal across multiple variants in a gene or other region of the genome to test the cumulative effect of these variants on a disease or trait. These aggregation tests can increase the power to detect rare variant genetic association. To further increase power, meta-analysis is employed to pool information across studies via summary statistics such as effect sizes and p-values. To perform proper aggregation test meta-analysis, accurate estimates of the covariances for the single-variant test statistics also are needed. Covariance files are often too large to be shared and estimation of the covariances requires access to individual level data for each of the participating studies. Unfortunately, individual-level genetic data often cannot be shared due to privacy concerns. In this study, we apply a previously proposed method of estimating single-variant test statistic covariance from an external reference panel to perform aggregation tests on a variety of traits from the UK Biobank. The method uses correlation between variants in the reference panel and allele frequencies from the meta-analysis studies to estimate the test statistic covariance. We propose a two-stage approach by first filtering genes using a null covariance where covariance between all variants is set to zero to perform aggregation tests, and in stage two testing only those genes passing a p-value threshold. Using the null covariance generally overestimates significance of genes, allowing us to confidently exclude non-significant genes in the first stage. Using this approach we are able to reproduce known associations such as for *APOB* with LDL cholesterol and *LPL* with HDL cholesterol. When p-values on the log scale produced using individual-level data and our method are compared, Pearson correlation coefficients are close to one across various traits. This approach allows for substantial memory cost improvements and will be a useful tool for collaboration between studies.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3625. Using genetic associations from shared clustered genetic predictors to infer causation

Authors:

F. Batool¹, A. Patel¹, D. Gill², S. Burgess¹; ¹Univ. of Cambridge, Cambridge, United Kingdom, ²Univ. of London, London, United Kingdom

Abstract Body:

Genetic variations are driver of phenotypic diversity and provide evidence on nature of causal relationship through associations among complex traits and diseases. Genome-wide association studies have shown a large number of genetic variants association with complex diseases, that governs the mechanism of these complex diseases with each variant having a very small effect size. For the gene region that contains a cluster of genes that controls the gene expression resulting in gene product such as risk factor, it is difficult to determine which proteins are causal risk factors. Often, selection of genetic variants from whole genome is made through strict pruning to a near independence threshold. When genetic variants are highly correlated, inclusion of too many highly correlated variants lead to ill-conditioning, and strict pruning on variants lead to imprecise estimates. To disentangle complex cis-gene regions, causal inference method is developed for complex traits that perform statistical dimensionality reduction as an alternative to traditional pruning methods to extract genetic proxies (instrumental variables) to avoid problems of collinearity. This method is illustrated using circulating levels of cytokines, namely, MCP-1, MCP-3, eotaxin for the variants from gene in CCL2 region with stroke which is the third leading cause of death worldwide. CCL2 that encodes these cytokines is located in gene region that also includes genes CCL7 and CCL11. Previously these cytokines were shown to draw associations on single exposure trait in a uni-variable MR setting. To discover the most likely casual risk factor, data for genetic associations of strokes (ischemic and cardioembolic) from MEGASTROKE consortium, genetic association of cytokines from GWAS of eQTL, for the variants from gene in CCL2 region. The correlations were taken from UK Biobank for European ancestries.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3627. Validation of large-scale Hispanic S-PrediXcan findings by independent Hispanic RNA-Seq data analysis confirms novel lipid-associated genes.

Authors:

L. Petty^{1,2}, H-H. Chen², W. Zhu¹, C. G. Downie³, M. Graff³, M. Lee⁴, K. E. North³, J. B. McCormick⁴, S. P. Fisher-Hoch⁴, J. M. Mercader⁵, H. M. Highland³, J. E. Below², Hispanic Lipids Consortium; ¹Vanderbilt Univ., Nashville, TN, ²Vanderbilt Univ. Med. Ctr., Nashville, TN, ³Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, ⁴Univ. of Texas Hlth.Sci. Ctr. at Houston, Brownsville Regional Campus, Brownsville, TX, ⁵Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract Body:

Background: Most genetically regulated gene expression imputation models have been derived from reference data that is primarily of European ancestry, and the utility of these models for disease gene discovery in non-European ancestry populations remains largely unexplored. Additionally, Hispanic/Latino populations have both higher prevalence of cardiometabolic disease (CMD) and are drastically underrepresented in genetic studies of CMD traits relative to non-Hispanic whites. To help address both knowledge gaps, we applied GTEx project-derived S-PrediXcan models to our Hispanic Lipids Consortium meta-genome wide association study (GWAS) results. We compared top findings to RNA-Seq association results in an independent Hispanic cohort.

Methods: We first completed a meta-GWAS of blood lipid/lipoprotein traits (high-density lipoprotein cholesterol [HDL-C], low-density lipoprotein cholesterol [LDL-C], total cholesterol [TC], and triglycerides [TG]), comprising up to 63,185 Hispanic/Latino individuals. The meta-analysis was performed using MR-MEGA and included two study-level principal components to capture ancestry differences. Then, we used MR-JTI GTEx v8 S-PrediXcan models to test the association of genetically regulated expression with each trait and performed Mendelian randomization to assess causality of all identified associations. Next, we tested association of whole blood expression between each of our S-PrediXcan-identified genes and the trait of interest in 880 Mexican American individuals from the Cameron County Hispanic Cohort (CCHC). Finally, we tested if genes associated with each CMD trait using gene expression were enriched in those genes associated with each CMD trait using S-PrediXcan.

Results: We identified strong enrichment of genes associated with measured gene expression in the S-PrediXcan findings (Kolmogorov-Smirnov $p_{HDL-C} < 2.2 \times 10^{-16}$, $p_{LDL-C} = 1.39 \times 10^{-3}$, $p_{TC} = 7.13 \times 10^{-6}$, $p_{TG} = 2.51 \times 10^{-9}$). In addition to the overall enrichment, our measured gene expression association analyses validated two of the novel genes discovered in our S-PrediXcan analyses, *NEDD9* for HDL-C and *GOLGA8A* for LDL-C.

Conclusion: Inclusion of diverse populations in RNA-Seq reference data to is necessary to realize the full potential of genetically regulated expression analyses in all populations. Despite low representation, here, we demonstrate that GTEx S-PrediXcan models have utility for identification of dysregulated genes associated with CMD in Hispanics/Latinos. In addition, we are constructing Hispanic expression prediction models in our RNA-Seq and genotype data, and will make these models publicly available.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3628. Variants in *CRY2* and *MTNR1B* influence the diurnal variability of blood glucose levels

Authors:

S. Jones¹, N. Sinnott-Armstrong^{1,2,3,4}, J. K. Pritchard^{2,5}, H. M. Ollila^{1,2,3,6}; ¹Inst. for Molecular Med. Finland, HiLIFE, Univ. of Helsinki, Helsinki, Finland, ²Dept. of Genetics, Stanford Univ., Stanford, CA, ³Broad Inst. of Harvard and MIT and Ctr. of Genomic Med., Massachusetts Gen. Hosp., Boston, MA, ⁴Fred Hutchinson Cancer Res. Ctr., Seattle, WA, ⁵Dept. of Biology, Stanford Univ., Stanford, CA, ⁶Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA

Abstract Body:

Introduction Circadian rhythms coordinate not only the timing of sleep and wakefulness, but also a wide range of homeostatic processes in the body. Environmental cues including light, temperature, feeding and exercise contribute to diurnal variation in gene homeostasis and expression. In particular, diurnal rhythms play a role in regulating the timing of glucose metabolism. Conversely, a disrupted metabolism manifesting, for example, as high glucose levels or even type-2 diabetes mellitus (T2DM) is associated with insomnia, later chronotype and short sleep in both epidemiological and genetic studies.

Early genetic associations with glucose levels and T2DM were discovered at the *MTNR1B* (a melatonin receptor) and *CRY2* loci that are also known canonical regulators of circadian rhythms. Yet the biological mechanisms behind the associations with glucose levels are not understood. We hypothesized that the circadian coordination of glucose levels may be under genetic control.

Methods In data from ~335,000 unrelated European-ancestry UK Biobank participants, we assessed whether genetic variants were associated with glucose timing, as well as glucose levels. We performed a "diurnal variability" GWAS of blood glucose levels. In the analysis, we assumed a cosine glucose profile, including terms relating to genotype interaction with the sine and cosine of blood-draw time, in addition to the standard additive genotype and covariate terms. For variants showing genome-wide significant interaction with the sine and cosine terms, we validated the results by applying a generalized additive model that does not assume a specific diurnal glucose profile.

Results We identified significant associations with both amplitude and phase of glucose levels at *MTNR1B* (rs10830963, $P=4.8 \times 10^{-8}$, $\text{beta}[\text{se}]=0.0076[0.0014]$) and *CRY2* variants (rs11605924, $P=1.7 \times 10^{-8}$, $\text{beta}[\text{se}]=0.0070[0.0012]$). This effect was notable since the levels are antiphase in the morning vs. the evening for the *MTNR1B* variant. Furthermore, the *CRY2* variant carriers showed a phase shift in glucose levels (of ~1 hour) compared to non-risk allele carriers. These effects were independent of diurnal preference (morningness/eveningness) with the variants showing stronger effects than an individual's diurnal preference.

Conclusion Our findings suggest that there are variants involved in coordinating the timing of blood glucose levels over the course of the day, and implicate *MTNR1B* and *CRY2* in the regulation of both overall glucose levels and glucose timing in humans. We are continuing to develop the methodology to allow more general application to other phenotypes and analysis with mixed-model approaches.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3629. Whole-exome sequencing in 545,835 individuals implicates 27 genes in eosinophil biology

Authors:

C. Sidore, L. Gurski, M. Suci, K. Siminovitch, Regeneron Genetics Center, GHS-RGC DiscovEHR Collaboration, G. Abecasis, M. Ferreira; Regeneron Genetics Ctr., Tarrytown, NY

Abstract Body:

Background. Eosinophils play a key role in immune responses to parasites, bacteria and allergens. Previous GWAS of peripheral blood eosinophil counts have identified over 500 loci with common variant associations, but the genes which underlie the observed associations are largely unknown. Analysis of rare coding variation provides an independent approach to identify genes involved in the regulation of eosinophil numbers in peripheral blood.

Methods. We sequenced the exomes of 545,835 individuals from the UK Biobank and Geisinger Health System (GHS) cohorts, and tested the association between eosinophil counts and rare (MAF<1%) non-synonymous variants in 19,411 genes, individually and in aggregate through gene burden and SKAT-like tests. We also performed a GWAS using imputed TOPMed data in the same individuals and tested if the rare variant associations were independent of common variant signals.

Results. We analyzed the association between eosinophil counts and 8M rare variants from exome sequencing, performing single variant and gene-based tests. We found significant ($P < 5 \times 10^{-8}$) associations with rare coding variants in 38 genes located in 28 loci. Of these, the association with 27 genes remained significant after accounting for 950 common variants (average effect size 0.023 SD, range 0.008-0.14) identified in the TOPMed GWAS. These include 25 genes for which the strongest association was with a burden of predicted loss-of-function (pLOF) variants (7 genes; average effect 0.255 SD, range 0.085-0.442) or deleterious missense variants (18 genes; average effect 0.192 SD, range 0.055-0.615). Twelve genes were the closest gene to a GWAS signal, and therefore are the likely effector genes of the common variant signals. For 5 genes (SLC27A3, ARHGAP25, RASGRP2, SRSF2, RASAL3) we found evidence for association with rare variants from exome sequencing but not with common variants from GWAS; the association with SRSF2 was explained by rare somatic mutations. Overall, 5 genes (SLC27A3, CISH, IL33, HABP2, ALOX15) had suggestive associations ($P < 5 \times 10^{-4}$) with allergic disease (159,544 cases vs. 321,114 controls), while 2 genes (ARHGAP25, RASGRP2) have not been previously implicated in eosinophil biology. We also tested for associations with a burden of rare non-coding variation in gene enhancers using imputed TOPMed data, but found no convincing novel associations.

Conclusions. We demonstrate that analysis of rare coding variation from exome sequencing can help pinpoint likely effector genes underlying common variant associations with eosinophil counts, identify new gene associations missed by GWAS and illuminate novel associations for allergic disease.

Statistical Genetics and Genetic Epidemiology Posters - Wednesday

PB3630*. Winner's curse in rare variants association analysis: bias depends on effect direction and the pooled methods used

Authors:

D. Soave^{1,2}, S. Kokic¹, L. Sun³; ¹Wilfrid Laurier Univ., Waterloo, ON, Canada, ²Ontario Inst. for Cancer Res., Toronto, ON, Canada, ³Univ. of Toronto, Toronto, ON, Canada

Abstract Body:

Rare variants (RVs) may explain a large portion of genetic heritability, however, studies involving RVs inherently have low power to detect associations with SNPs individually. Gene and set-based tests aggregate across multiple SNPs to improve power by jointly testing for association between a trait and a group of SNPs. A common approach to effect estimation for the group of SNPs is the average genetic effect (AGE), where the parameter corresponds to the effect of a single collapsed genotype variable. The inclusion of non-causal SNPs or SNPs with effects in opposite direction may lead to a downward bias of the effect size estimate. On the other hand, hypothesis testing and effect estimation performed using the same sample will lead to an upward bias, due to the phenomenon known as selective inference or winner's curse. A modified bootstrap correction to AGE estimation for the rare variants context may provide good biased-reduced AGE estimates, however, it is not clear (1) if the individual estimates for each of the causal variants suffers equally from the winner's curse (i.e. with the same amount of upward bias) regardless of true effect size or effect direction, and (2) if the bias estimation/correction depends on the type of tests used (e.g. burden/linear- versus SKAT/quadratic-type tests). To investigate these questions, we conducted a simulation study using the Genetic Analysis Workshop 17 data (real human sequence data from the 1000 Genomes Project). We examined the implications of the competing winner's curse and heterogeneity of individual effects under both the linear and quadratic testing classes. We observed that both the modified bootstrap method and likelihood method effectively reduce the winner's curse in the AGE estimate under both the unidirectional and bidirectional scenarios. Under the unidirectional model, each individual naïve effect estimate has upward bias. However, the amount of bias differs between rare variants, depending on the MAF and its true effect. Furthermore, under the bidirectional model when not all effects have the same direction, the individual estimates can be downward biased depending on the test used. The pooling of rare variants is clearly an issue for effect size estimation for individual variants. The situation becomes even more complicated as we consider the possibility of including causal variants with bidirectional effects. Unfortunately, the usefulness of our ability to quantify genetic effects and the missing heritability due to rare variants remains in question as a deeper understanding of the consequences of pooling/grouping variants with heterogeneous effects is warranted.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3631. Machine learning-based phenotyping significantly improves power for genetic discovery in NAFLD and NASH

Authors:

B. Ungun¹, F. Casale², **H. Sominen**³, K. Sandor¹, S. Modi¹, M. Albert⁴, S. Satapati¹, T. Karaletsos¹, T. Soare³, P. Palmedo¹, D. Koller¹; ¹insitro, South San Francisco, CA, ²Dept. of Informatics, Technical Univ. of Munich, Munich, Germany, ³insitro, Boston, MA, ⁴Precision Immunology NewCo, South San Francisco, CA

Abstract Body:

Genetically supported drug targets are at least twice as likely to be approved; however, large human cohorts are required to catalog genetic associations for complex traits. Assembling large datasets with gold-standard clinical diagnosis is difficult, particularly for diseases such as nonalcoholic fatty liver disease (NAFLD) or nonalcoholic steatohepatitis (NASH) due to the invasive nature of the liver biopsy. Here, we used a machine-learning (ML) algorithm to quantify NAFLD- or NASH advanced (F4) fibrosis-risk in the UK Biobank, and leveraged it to drive genetic discovery. We trained an ML classifier using age, sex, and 37 blood measurements and anthropometric markers to predict ground truth cases (n=1,774) and controls (n=2,209), and then applied the ML model to estimate the risk of NAFLD in 275,727 Europeans in the UK Biobank (individuals meeting our inclusion criteria). Within individuals with known or ML-imputed NAFLD, we specifically identified a high risk subset (n = 77,086), i.e. individuals who were predicted to have NAFLD with “high confidence” by the above classifier (20% false positive rate), and then trained a binary classifier using 15 biomarkers of fibrosis with consistent literature support to predict ground truth cases (n=97 with an ICD-10 code for F4 fibrosis) from simple steatosis (n=1,246), and used it to estimate the risk of NASH F4 fibrosis. ML-based risk scores of NAFLD and NASH F4 fibrosis predicted ICD-10 codes with high accuracy within the UK Biobank dataset (AUC=0.87±0.01 for NAFLD and 0.81±0.03 for NASH F4 fibrosis, on held out sets). A genome-wide association study of these ML-derived traits identified > 100 genome-wide significant loci across both traits, significantly expanding our understanding of the genetic architecture of these complex diseases (P < 5 x 10⁻⁸). Amongst these loci were well known associations, including HSD17B13, PNPLA3, TM6SF2 and GCKR, confirming established biology. In addition, we replicated published associations using proxy phenotypes by other groups (e.g. liver fat quantified from ML of abdominal MRI data or chronic elevation of alanine aminotransferase), including TRIB1, COBL1 and MAST3. Overall, these findings demonstrate the value of leveraging ML-based trait predictions to identify novel genetic associations.

Statistical Genetics and Genetic Epidemiology Posters - Thursday

PB3633. A unifying statistical framework to discover disease genes from GWAS

Authors:

J. McManus, R. Lovelett, D. Lowengrub, S. Christensen;
Kallyope, Inc., New York, NY

Abstract Body:

Genome-wide association studies (GWAS) identify genomic loci associated with complex traits, but it remains a major challenge to systematically pinpoint the specific genes underlying the association signals. In the biotechnology and pharmaceutical industries, the difficulty in ascertaining disease genes from GWAS has hampered the use of human genetics in drug discovery. We therefore developed a computational framework to rigorously discover the genes dysregulated by causal variants, without restricting ourselves to given cell types or relying on assumptions from biological networks. In particular, we extend the equations of statistical fine-mapping, to compute the probability that each gene in the human genome is effected by a causal variant, given a particular trait. Previously, statistical fine-mapping has been applied to compute the probability that particular genetic variants mediate disease risk. We show how to compute analogous probabilities of association for genes, which are the basis for drug development. Furthermore, we demonstrate that genes targeted by causal variants can be confidently identified, even when the causal variants themselves cannot be pinpointed. Our computations are enabled by several key innovations. First, we partition the genome into optimal linkage disequilibrium blocks, enabling genome-wide detection of trait-associated genes. Second, we unveil a comprehensive mapping that associates genetic variants to the target genes they affect. The combined performance of the map on high-throughput functional genomics and eQTL datasets supersedes the state of the art. Lastly, we describe an algorithm which learns, directly from GWAS data, how to incorporate prior knowledge into the statistical computations, significantly improving their accuracy. We validate each component of the statistical framework individually and in combination. Among methods to identify genes targeted by causal variants, this paradigm rediscovers an unprecedented proportion of known disease genes, without sacrificing precision. Moreover, it establishes human genetics support for many genes previously implicated only by clinical or preclinical evidence, and it discovers an abundance of novel disease genes with compelling biological rationale.

Pharmacogenomics Posters - Thursday

PB3634. Local ancestry stratified GWAS identifies pharmacogenomic variants associated with metformin glycemic response in African American individuals with type-2 diabetes

Authors:

S. Xiao¹, B. Wu¹, S. Yee², F. Xu³, S. Sridhar³, M. Yang¹, S. Hochstadt¹, W. Cabral¹, D. Lanfear¹, M. Hedderson³, K. Giacomini², K. Williams¹; ¹Henry Ford Hlth., Detroit, MI, ²Univ. of California San Francisco, San Francisco, CA, ³Div. of Res., Kaiser Permanente Northern California, Oakland, CA

Abstract Body:

Metformin is considered first-line therapy in treating type-2 diabetes (T2D), yet little is known regarding the genetic underpinnings of drug response. African Americans are disproportionately affected by T2D, and in earlier work, we observed differences in metformin treatment response by race-ethnicity. Unfortunately, African Americans have been understudied in most genomic and pharmacogenomic studies to date, and the studies that have been performed do not fully account for the effects of underlying population structure. Here we performed the first local-ancestry informed genome-wide analysis for pharmacogenomic variants associated with metformin treatment response. Our discovery population included African American participants from the Diabetes Multi-omic Investigation of Drug Response (DIAMOND) cohort, a study of health plan members from metropolitan Detroit. The replication cohort consisted of African American participants who received care at Kaiser Permanente Northern California (KPNC). Pharmacy claims were used to calculate an average daily drug exposure among individuals on metformin monotherapy. Among individuals continuously exposed to greater than or equal to 425mg of metformin daily, we evaluated for genetic variants associated with a change in blood glycated hemoglobin (HbA1c) levels. A unique aspect to our approach was that we evaluated variant effect sizes only at locations unambiguously homozygous for African ancestry. In this way, our discovery and replication analyses were not affected by an averaging of effect sizes across ancestral groups, and the effect of population group differences in linkage disequilibrium were also minimized. Our analysis revealed a genome-wide significant signal in the DIAMOND cohort at variant *rs143276236* located in an intronic region of the gene *ARFGEF3* ($P=9.4E-9$) on chromosome 6. Re-evaluation in KPNC demonstrated an association with consistent direction ($P=0.021$). The meta-analyzed results showed a $P=8.6E-9$. Functionally, this gene has been previously predicted to have role in the regulation of glucose homeostasis through its effect on insulin secretion. Our study demonstrates the potential power and importance of using an ancestry-informed association approach when seeking to identify pharmacogenomic variants in admixed populations.